The 12th International Congress of Human Genetics and the American Society of Human Genetics 61st Annual Meeting October 11-15, 2011 Montreal, Canada

PLATFORM ABSTRACTS

1

Identification of a mosaic activating mutation as the molecular basis of Proteus syndrome using massively parallel sequencing of affected tissues. *L.G. Biesecker[1], M.J. Lindhurst[1], J.C. Sapp[1], J.K. Teer[1], J.J. Johnston[1], K. Peters[1], J. Turner[1], D. Bick[2], L. Blakemore[3], K. Brockman[4], P. Calder[5], M. Deardorff[6], D.B. Everman[7], R.M. Greenstein[8], B.M. Kato[9], K.M. Keppler-Noreuil[10], R.T. Miyamoto[11], K. Newman[12], S. Rothenberg[13], D.J. Schwartzentruber[14], V. Singhal[15], J. Upton[16], S. Wientroub[17], E.H. Zackai[7], P.G. Robey[18], P.L. Schwartzberg[1], T.N. Darling[19], L.L. Tosi[3], J.C. Mullikin[20].* 1) NHGRI, NIH, Bethesda, MD, USA; 2) Division of Pediatric Genetics, Medical College of Wisconsin, Milwaukee, WI, USA; 3) Division of Orthopaedics and Sports Medicine, Children's Hospital and Medical Center, Washington, DC, USA; 4) Department of Pediatrics and Pediatric Neurology, Georg August University, Göttingen, Germany; 5) The Royal National Orthopaedic Hospital, Brockley Hill, Stanmore, London, UK; 6) Division of Human Genetics, Children's Hospital of Pennsylvania, Philadelphia, PA, USA; 7) Greenwood Genetics Center, Greenwood, SC, USA; 8) Department of Genetics and Developmental Biology, University of Connecticut, West Hartford, CT, USA; 9) The Ear Institute, Palm Desert, CA, USA; 10) University of Iowa, Department of Pediatrics, Division of Genetics, Iowa City, IA, USA; 11) Department of Otolaryngology, Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA; 12) Department of Surgery, Children's National Medical Center, Washington, DC, USA; 13) Department of Pediatric Surgery, Rocky Mountain Hospital for Children, Denver, CO, USA; 14) Surgery Branch, National Cancer Institute, NIH, Bethesda, MD, USA; 15) Department of Pediatric Plastic and Craniofacial Surgery, Children's Mercy Hospitals and Clinics; Kansas city, MO, USA; 16) Department of Surgery, Children's Hospital Boston, Boston, MA, USA; 17) Department of Pediatric Orthopaedics, Dana Children's Hospital, Tel Aviv Medical Center, Tel Aviv University, Tel Aviv, Israel; 18) NIDCR, NIH, Bethesda, MD, USA; 19) Department of Dermatology, Uniformed Services University of the Health Sciences, Bethesda, MD, USA; 20) NIH Intramural Sequencing Center, Rockville,MD, USA.

Proteus syndrome manifests mosaic overgrowth in skin, connective tissue, brain, and other tissues. It was hypothesized to be caused by a mosaic mutation, lethal in the non-mosaic state. We tested this hypothesis using massively parallel sequencing of DNA from Proteus syndrome tissues, comparing affected to unaffected tissues. Seventeen samples from 11 patients were sequenced using targeted exome capture. These samples included four affected-normal sample pairs, two affected samples from a fifth patient, one pair of samples from discordant monozygotic twins, and five unaffected parents. We generated ~1.6 Gb of sequence data, with >87% coverage of exons with a high quality genotype. A novel variant was initially identified in a single affected sample using an affected-unaffected filter. Upon manual examination, we found this variant in three additional affecteds but not in the parents of affecteds or the unaffected identical twin, 1000 genomes (n= 634), or ClinSeq (n=401). The sequence variant data were confirmed and extended by a custom restriction enzyme assay of >150 samples from 29 patients. Most affected samples were specimens removed from clinically abnormal areas at surgery. Twenty-seven of 29 patients with Proteus syndrome were found to have the identical mutation in this gene, but it was not present in >20 cell lines and tissues from persons with unrelated disorders. Tissues and cell lines from patients with Proteus syndrome harbored admixtures of mutant alleles that varied from 1% to ~50%. Patient-derived cell lines with the mutation showed evidence of activation of this gene product using anti-phosphoprotein antibodies and western blot analyses. We also show that a pair of single cell clones established from the same starting culture and differing only by their mutation status had differential activation of this protein. We conclude that a somatic mosaic activating mutation in this gene causes Proteus syndrome, validating the Happle mosaicism hypothesis. That this mutation is the same in all patients, mosaic, and not in a CpG dinucleotide explains the rarity of this disorder. These data show the power of massively parallel sequencing to identify causative genes in disorders that are not amenable to positional cloning. Further, these results provide a therapeutic target for this severe and progressive disorder. (The author confirms that the gene and mutation will be disclosed at the ICHG meeting, should the abstract be selected).

2

Germline deletion of the miR-17~92 cluster causes developmental defects in human. *J. Amiel[1], E. Yao[2], P. Callier[3], L. Faivre[3], V. Drouin[4], S. Cariou[1], A. Van Haeringen[5], D. Genevieve[6], A. Goldenberg[4], M. Oufadem[1], S. Manouvrier[7], A. Munnich[1], J. Alves Vidigal[2], S. Lyonnet[1], A. Henrion-Caude[1], A. Ventura[2], L. de Pontual[1].* 1) Dept Genetics and INSERM U781, Hosp Necker, Univ Paris 5, Paris, France; 2) Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA; 3) Service de Génétique, Hôpital d'enfants, Dijon, France; 4) Service de Génétique, Hôpital Charles Nicolle, Rouen, France; 5) Department of Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands; 6) Service de Génétique, Hôpital Arnaud de Villeneuve, Montpellier, France; 7) Service de Génétique Clinique, Hôpital J. de Flandre, Lille, France.

Oculodigitoesophagoduodenal syndrome (ODED, MIM164280) is an autosomal dominant syndrome with microcephaly, relative short stature and digital anomalies as core features. ODED is caused by germline loss-of-function mutations of the MYCN gene (MIM 164840) at 2p24 and genetic heterogeneity has been suspected. We performed a 244k CGH-array in 10 index cases displaying skeletal abnormalities consistent with a diagnosed as ODED but lacking any mutation at the MYCN locus. We first identified a 13q31.3 in 2 familial cases (with 2 and 3 affected individuals respectively), spanning 2.98 Mb to 165 kb with a shortest region of overlap encompassing miR-17~92 and the first exon of GPC5. No GPC5 coding sequence mutation could be identified in the remaining 8 cases. The DECIPHER Database quoted an individual with a similar phenotype and harbouring a 180kb hemizygous 13q31.3 microdeletion encompassing the entire miR-17~92 locus and the first exon of GPC5. miR-17~92 is a polycistronic miRNA cluster, also known as Oncomir-1, encoding for six distinct miRNAs essential for mammalian development. Indeed, its complete inactivation leads to perinatal lethality in mice. Both MYC and MYCN can activate the transcription of miR-17~92 and can directly bind to the miR-17~92 promoter region in human and murine cells. Mice carrying a heterozygous targeted deletion of miR-17~92 are viable and fertile. We show that they recapitulate the ODED phenotype as they are smaller, microcephalic and present a striking shortening of the mesophalanx of the fifth ray of the upper limbs. Homozygous knock-out embryos show complete absence of the mesophalanx of the fifth digit, the presence of a small mesophalanx of the second digit, and hypoplasia of the first digital ray. Neither syndactyly nor intestinal atresia was observed. Importantly, GPC5 expression is similar in miR-17~92 mutant and control mice. Our results show that various aspects of mammalian development via MYCN occur through transactivation of miR-17~92 (for which a fine tuning is required). This study provides the first evidence of a germline mutation of a miRNA gene leading to a developmental defect in humans.

3

The Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging: Developing a Novel Fully-Automated System to Assay Telomere Lengths and Preliminary Results on a Cohort of 100,000. *K. Lapham[1], J. Lin[1], L. Fang[1], M. Kvale[1], B. Dispensa[1], S. Hesselson[1], L. Walter[2], D. Ludwig[2], S. Miles[2], S. Rowell[2], W. McGuire[2], S. Connell[2], C. Zau[2], D. Smethurst[2], P. Kwok[1], N. Risch[1], C. Schaefer[2], E. Blackburn[1].* 1) UCSF Institute for Human Genetics, San Francisco, CA; 2) Kaiser Permanente Northern California Division of Research, Oakland, CA.

A growing body of literature links shorter telomere length with various age-related diseases and earlier mortality. Telomere length has also been more broadly interpreted as a marker for biological age and an indicator of overall health. This project was directed at adding telomere length data via qPCR to a genome-wide association study of 100,000 participants in the Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging. Analysis of telomere length was conducted at UCSF on DNA extracted from saliva samples. The telomere length analysis in this project required the design and construction of a novel Automated Telomere Length Analysis System (ATLAS) to handle the high throughput processing of samples. ATLAS is a fully automated DNA-to-data qPCR robot, which integrates over 20 separate instruments and currently runs and analyzes 30,720 qPCR reactions (5,760 samples) per day. The entire system is capable of running 24 hours/day with live video, data streams, and environmental variables monitored and controlled remotely from any internet-connected computer or smart phone. To complete the analysis, ATLAS ran over 1.8 million qPCR reactions. Novel sample arrangement, large scale reagent batch production methods, liquid handling techniques, control scheme, and workflow procedures were developed to both overcome the numerous challenges and take advantage of the immense scale of this project. A custom data analysis and quality control pipeline was created for extracting sample information and reporting individual telomere lengths in real time.

The analyzed telomere lengths have a log normal distribution and raw coefficient of variation among replicates of 5.5%. We are developing more sophisticated data cleaning and quality control tools to improve the reliability of the TL results. Once complete, these telomere length measurements will be combined with comprehensive longitudinal electronic medical records, demographic, behavioral, and geographic information as well as genome-wide association data covering 675,000+ SNPs for each sample. These data will make it possible to examine age-adjusted differences in telomere length by gender, race-ethnicity, and socioeconomic status, as examples. In addition, detailed data on health history and current health status of cohort members will enable examination of telomere length with many different age-related conditions and changes in health over time.

4

Natural Selection in the Cholera Endemic Ganges River Delta Region. *E. Karlsson[1,2], A. Rahman[3], I. Shlyakhter[1,2], J. Harris[3], F. Qadri[4], P. Sabeti[1,2], R. LaRocque[3].* 1) Harvard University, Cambridge, MA; 2) Broad Institute, Cambridge MA; 3) Massachusetts General Hospital, Boston MA; 4) ICDDR,B, Centre for Health and Pop. Research, Dhaka, Bangladesh.

Vibrio cholerae has likely played an important role in human evolution, especially in the Ganges River Delta, the historical and current epicenter of cholera. Observational data support this hypothesis. In particular, individuals with blood group O are at increased risk of severe cholera, and the lowest prevalence of type O in the world is found in the Ganges River Delta. In previous work, we showed that in Bangladesh, cholera aggregates within families independent of blood group, and that the gene LPLUNC1, a component of the innate immune system, is associated with cholera susceptibility. Here, we report the first genome-wide study of the non-migrating ethnic group occupying the Ganges River Delta. Using the Illumina 1M SNP array, we genotyped 36 Bengali mother-father-child trios from Dhaka, Bangladesh. Our results show that Bengalis are a homogenous population group on a distinct branch of the human genetic tree from the 11 populations of HapMap3. Using the Composite of Multiple Signals method, we identified 322 signals of natural selection in the Bengalis (average size 180kb with 2.5 genes; ~163 have 0 or 1 genes). We tested these regions for enrichment of particular gene categories using INRICH, a new pathway analysis tool designed to detect significantly enriched gene sets in genome-wide datasets. We found two especially interesting patterns. First, we repeatedly identified potassium channel genes in strongly selected regions. Second, we found exceptionally strong enrichment (p=2x10-4) for a module of co-expressed genes linked to the gene IKBKG, part of the immunity / inflammation NF-2B complex. Infectious diseases can exert strong selective pressure, and tests for natural selection are a powerful way to find genes influencing susceptibility. We show that, by leveraging massive public datasets and powerful new computational tools, we can identify multiple candidate genes using just 108 individuals. We are now evaluating our candidate genes for association with cholera susceptibility, using transmission disequilibrium testing (TDT) of parents / cholera-affected child trios.

5

Contrasting human X-linked and autosomal variation in population-scale whole genome sequencing. *A. Keinan, S. Gottipati, A. Siepel, A.G. Clark, L. Arbiza.* Cornell University, Ithaca, NY 14853.

The ratio of normalized genetic diversity between chromosome X and the autosomes (X/A) is expected, under equilibrium assumptions, to be 3/4. However, both the forces of natural selection and many types of demographic events can lead to deviations from this ratio. To characterize X/A in human populations while disentangling the effects of these two factors, we contrasted diversity on chromosome X and the autosomes from whole-genome sequences of 36 West African (YRI) and 33 European (CEU) females from the 1000 Genomes Project. Partitioning the data by genetic distance from the nearest gene, we found that diversity reduction due to selection at linked sites has been a more powerful force on chromosome X, and that as a result X/A increases with distance from genes (P<0.001 for both CEU and YRI). We next contrasted X/A between populations by studying the CEU-to-YRI ratio of diversity (*relative diversity*) and the CEU-to-YRI ratio of the X/A ratio (*relative X/A*). Interestingly, neither X-linked nor autosomal relative diversity is sensitive to distance from genes. As a consequence, the relative X/A remains constant across all distances (P=0.42 in a test of correlation), and is always consistent with the genome-wide estimate of 0.84±0.03, despite the pronounced dependence of selective effects on proximity to genes. These results reveal a reduction of X/A in non-Africans relative to West Africans that cannot be attributed to the effects of diversity-reducing selection. Rather, this relative reduction is most parsimoniously explained by sex-biased demographic events associated with the dispersal of modern humans out of Africa. More generally, this study leads us to propose a new approach that entails analyzing ratios of diversity between different populations. This approach allows focusing—with increased resolution—on events occurring after the populations have split. It is not sensitive to the effects of natural selection if these are similar on a genome-wide average across populations, and it is also robust to unannotated functional elements, to normalization by genetic divergence, and to differential ascertainment biases between chromosome X and the autosomes, all of which have confounded previous studies.

6

The geographic structure of allele sharing and rare variant diversity assessed from re-sequencing of 202 genes in 15,000 individuals. *J. Novembre[1], D. Wegmann[1], M. Zawistowski[2], A. Rakshi[1], S. Gopalakrishnan[2], D. Kessner[1], P. St. Jean[3], L. Li[3], M.G. Ehm[3], J. Li[3], Y. Li[3], G. Abecasis[2], J.C. Whittaker[3], S.L. Chissoe[3], V.E. Mooser[3], M.R. Nelson[3], S. Zöllner[2].* 1) Dept Eco & Evo Biol, Univ California, Los Angeles, Los Angeles, CA; 2) Biostatistics Department, School of Public Health, University of Michigan, Ann Arbor, MI; 3) GlaxoSmithKline, Research Triangle Park, NC.

The geographic distribution of rare variants is poorly understood because the sample sizes of most contemporary studies have been too limited to document their abundances . Here we present patterns of rare variant diversity observed amongst 15,000 subjects who have been re-sequenced to a median depth of ~27x per individual across 202 genes. This large sample provides the opportunity to document how rare variants are shared across populations as it includes >275 individuals from African-American and Indian ancestries, ~85 from East Asia, and within Europe, >500 individuals from each of 6 geographically defined populations. For both single nucleotide variants and bi-allelic indels, we quantify how allele sharing between populations decreases as a function of minor allele frequency, how the divergence between populations impacts rare variant sharing, and how populations vary in the relative abundance of rare variation. The results are relevant to on-going genotype-phenotype analyses using these samples and have broader implications for how population stratification impacts rare variant mapping methods.

7

A direct characterization of human mutation. *J.X. Sun[1,2], A. Helgason[3], G. Masson[3], S.S. Ebenesersdóttir[3], H. Li[1,4], S. Mallick[1], N. Patterson[1,4], A. Kong[3], D. Reich[1,4], K. Stefansson[3].* 1) Dept of Genetics, Harvard Medical School, Boston, MA; 2) Division of Health Sciences and Technology, MIT, Cambridge, MA; 3) deCODE Genetics, Reykjavik, Iceland; 4) Broad Institute, Cambridge, MA.

Mutation and recombination provide the raw material of evolution. This study reports the largest study of new mutations to date: 2,058 germline mutations discovered by analyzing 85,289 Icelanders at 2,477 microsatellites. We find that the paternal-to-maternal mutation rate ratio is 3.3, and that the mutation rate in fathers doubles between the ages of 15 to 45 whereas there is no association to age in mothers. Strong length constraints apply for microsatellites, with longer alleles tending to mutate more often and decrease in length, whereas shorter alleles tending to mutate less often and increase in length. Based on these direct observations of the microsatellite mutation process, we build a model to estimate key parameters of evolution without calibration to the fossil record. The sequence substitution rate per base pair is estimated to be $1.84\text{-}2.21\times10^{-8}$ per generation (95% credible interval). Human-chimpanzee speciation is estimated to be 3.92-5.91 Mya, challenging views of the Toumaï fossil as dating to >6.8 Mya and being on the hominin lineage since the final separation of humans and chimpanzees.

8

Evolution and functional impact of human coding variation from deep sequencing of 2,440 exomes. *A.W. Bigham[1], J.A Tennessen[2], T. O'Connor[2], E. Kenny[3], S. McGee[2], R. Do[4,5], X. Liu[6], G. Jun[7], H.M. Kang[7], D. Jordan[8], G. Abecasis[7], E. Boerwinkle[6], S. Sunyaev[4,8], C.D. Bustamante[3], M.J. Bamshad[1,2], J.M. Akey[1,2] on behalf of the NHLBI Exome Sequencing Project.* 1) Department of Pediatrics, University of Washington, Seattle, WA, USA; 2) Department of Genome Sciences, University of Washington, Seattle, WA, USA; 3) Department of Genetics, Stanford University, Palo Alto, CA, USA; 4) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 5) The Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 6) Department of Human Genetics, University of Texas Health Sciences Center at Houston, TX, USA; 7) Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA; 8) Division of Genetics, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA.

Deep exome resequencing provides an unbiased perspective of the spectrum of protein-coding variation and is a powerful approach for delineating patterns of genetic diversity among genes, pathways, individuals, and populations. We analyzed exome data from 2,440 individuals of European (n= 1,352) and African (n=1,088) ancestry as part of the NHLBI Exome Sequencing Project, the aim of which is to discover novel genes and mechanisms contributing to heart, lung and blood disorders through application of next-generation exome sequencing. Each exome was sequenced to a mean coverage of 116x, allowing detailed inferences into population genomic patterns of both common and rare coding variation. We identified 503,481 single nucleotide variations (SNVs), the majority of which were novel and rare (76% of variants had a minor allele frequency < 0.1%), reflecting the recent dramatic increase in human population size. The mean number of SNVs per individual was 13,595, 42% of which are predicted to be functionally important. The unprecedented magnitude of this dataset allowed us to rigorously characterize the large variation in nucleotide diversity among genes (ranging from 0 - 1.32%), the role of positive and purifying selection in shaping patterns of protein-coding variation, and the differential signatures of population structure from rare and common variation. This dataset provides a rich resource for studies of human evolutionary history and identification of variants for both Mendelian and complex traits. Finally, allele frequency data from these exomes will be available publically via dbSNP and a large fraction of the genotype data and linked phenotype information have been deposited in dbGaP.

9

Recent admixture in an Indian population of African ancestry and its potential in admixture mapping. *A. Basu[1], A. Narang[2], P. Kumar[2], V. Rawat[2], A. Mukhopadhyay[2], D. Dash[2], M. Mukherjee[2], Indian Genome Variation.* 1) National Institute of BioMedical Genomics, Kalyani, West Bengal, India; 2) Institute of Genomics and Integrative Biology, New Delhi, India.

dentification and study of genetic variation in recently admixed populations not only provides insight into historical population events but is also a powerful approach for mapping disease genes. We studied a population (OG-W-IP) of African-Indian origin residing in the western part of India for 500 years who are believed to be descendants from Bantu speaking population of Africa . We have carried out this study using a set of 18534 autosomal markers common between Indian , CEPH-HGDP and HapMap populations. Principal component analysis clearly revealed that the African-Indian population derives its ancestry from Bantu-speaking west-African as well as Indo-European speaking northern and north-west Indian population(s). STRUCTURE and ADMIXTURE analysis show that, overall the OG-W-IPs derive 58.7% of their genomic ancestry from their African past with very little inter-individual ancestry variation (8.4%). The extent of linkage-disequilibrium also reveals that the admixture event has been recent. Functional annotation of genes encompassing the ancestry informative markers which are closer in allele frequency to the Indian ancestral population revealed significant enrichment of biological processes, like ion channel activity and cadherins. We briefly examine the implications of determining the genetic diversity of this population which could provide opportunities for possible studies using admixture mapping.

10

A Sequence-Based Approach To Investigating Balancing Selection In Classical Human Leukocyte Antigen (HLA) Loci. *P.G. Bronson[1], S.J. Mack[2], H.A. Erlich[3], M. Slatkin[1].* 1) Department of Integrative Biology, University of California, Berkeley, USA; 2) Children's Hospital Oakland Research Institute, Oakland, USA; 3) Roche Molecular Systems, Inc., Pleasanton, USA.

HLA allele diversity is maintained by balancing selection but it is not clear if this is symmetrical or asymmetrical. We tested the hypothesis that HLA alleles are under asymmetrical balancing selection in single populations. We also examined population differentiation in alleles within and between geographical regions and the relationship between allele ages and frequencies in single populations. Allele frequencies for 427 populations in 11 regions and exon sequences encoding peptide-binding domains were used to generate allelic genealogies (trees) for each locus in single populations. The lengths of terminal branches in trees indicated relative ages of alleles in populations. Allele ages for each tree were quantified with R_{SD} ($L^*(1\text{-}1/n)/D$; no. alleles (n), sum of terminal branch lengths (L), tree depth (D)). Under the null model the expectation of R_{SD} is 1.25. The significance of deviation between observed and expected trees was determined through a randomization test with 10,000 replicates. Significant deviations (P<0.05) in R_{SD} indicated that some alleles in a tree were older than expected under symmetrical balancing selection. Wright's Fixation Index (F_{ST}) was estimated to examine population differentiation. Spearman's tests were used to test correlation between allele age and frequency. Results provided significant evidence of deviation from symmetrical balancing selection at all loci except *DQA1*. The strongest evidence was observed for *DRB1*, *B* and *DPA1* (>50% populations had $P\leq1\text{x}10^{-4}$; no. populations analyzed (N)=212, 119, 22, respectively). The largest R_{SD} observed was 16.7 (*B*, Chinese, P=$1\text{x}10^{-5}$). Strong results were also observed for *A* and *C* (50% populations had $P\leq1\text{x}10^{-3}$; N=105, 89, respectively), and *DQB1* and *DPB1* (85% of populations had $P\leq0.05$; N=143, 77, respectively). Observed F_{ST} values were similar to the mean genomewide F_{ST} for HapMap SNPs (0.15). Correlation tests were not significant. We observed low to moderate genetic variation within and between regions. Allele age was not correlated with frequency. In conclusion, this is the largest sequence-based study of balancing selection in HLA. Our analyses provide strong evidence suggesting that the balancing selection present in HLA alleles is asymmetrical, where some heterozygotes enjoy greater fitness than others. Because HLA polymorphism is crucial for pathogen resistance, this may manifest as a frequency-dependent selection with fluctuation over time in specific heterozygote fitness.

11

An integrated map of genetic variation in over 1000 human genomes. *G. McVean, The 1000 Genomes Project Consortium.* Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom.

By combining low coverage whole genome sequencing, targeted exome capture, and array genotype data on over 1000 individuals with a worldwide geographic distribution, we have generated an integrated map of over 40 million SNP, short indel and larger structural variants in the human genome. We show the ability of this resource to provide a comprehensive view of low-frequency genetic variation, as well as providing a catalogue of functional variants and a framework for imputation into existing studies. The new data allow us to describe the sharing of rare variants across related and admixed populations and to evaluate the benefit of different data types and experimental designs for population-scale sequencing. I will also discuss recent advances in variant calling within the project, the challenges presented by integration of different variant types and the future directions of the 1000 Genomes Project. The data are available for browsing and downloading at www.1000genomes.org.

12

Dysfunction is the normal state: analysis of gene-disrupting variants in 1,094 human genomes. *D.G. MacArthur[1,8], S. Balasubramanian[2], A. Frankish[1], N. Huang[1], L. Habegger[2], J. Morris[1], L. Jostins[1], C.A. Albers[1], J. Rosenfeld[3], E. Garrison[4], D.F. Conrad[5], M.A. DePristo[6], X. Mu[2], E. Khurana[2], K. Walter[1], R. Handsaker[6,7], S.B. Montgomery[9], J.K. Pickrell[10], Z. Zhang[11], J.C. Barrett[1], J. Harrow[1], M.E. Hurles[1], M.B. Gerstein[2], C. Tyler-Smith[1], 1000 Genomes Project Consortium.* 1) Wellcome Trust Sanger Institute, Hinxton, UK; 2) Yale University, New Haven, CT; 3) Zucker Hillside Hospital, New York, NY; 4) Boston College, Chestnut Hill, MA; 5) Washington University, St Louis, MO; 6) Broad Institute of Harvard and MIT, Cambridge, MA; 7) Harvard Medical School, Boston, MA; 8) University of Sydney, Sydney, Australia; 9) Stanford University School of Medicine, Stanford, CA; 10) University of Chicago, Chicago, IL; 11) Albert Einstein College of Medicine, Bronx, NY.

Recent genome sequencing studies have revealed a surprising number of variants predicted to disrupt the function of protein-coding transcripts, even in apparently healthy individuals. These loss-of-function (LOF) variants - such as nonsense SNPs and frameshift indels - are skewed towards low frequencies, suggesting an excess of deleterious variants, consistent with an impact on gene function and human disease risk. However, these variants are also enriched for sequencing and annotation artefacts, making careful filtering essential for interpretation.

We have developed a pipeline for annotating LOF polymorphisms from large-scale human sequence data, using 185 whole genomes sequenced by the 1000 Genomes pilot projects as well as analysis of all candidate LOF variants in a single high-quality genome. Filters were validated using independent genotyping for 1,877 variants and manual reannotation of 807 putative LOF genes. This pipeline removes >50% of the 2,688 candidate LOF variants identified in the pilot project. After filtering, we estimate that the average European carries ~140 true LOF variants, 20-30 in a homozygous state.

As expected, validated LOF variants are enriched at low frequencies, and often alter RNA expression of affected exons. LOF-tolerant genes are less conserved, less tissue-specific and have fewer protein-protein interactions than other genes, and show striking differences in properties to known recessive disease genes. These differences provide a potential approach for prioritising candidate genes identified in clinical samples. We have also performed genotype imputation in 16,994 individuals to explore the effects of these variants on complex disease risk.

Here we describe the application of our annotation pipeline to low-coverage whole genomes and high-coverage exomes of the first 1,094 individuals from the main phase of the 1000 Genomes Project. Combined annotation of SNPs, multi-nucleotide polymorphisms, small indels and large deletions in this data-set will provide the largest collection of high-confidence LOF variants yet assembled, particularly for the important class of variants with a frequency below 1%. We will present analyses of the impact of these variants on human evolution, gene function and disease risk.

13

Dual function of DNA sequences: coding exons function as enhancers of nearby genes. *N. Ahituv[1, 2], E.J. Clowney[3], M.J. Kim[1, 2], O. Agamy[4], S.L. Clarke[5], A.M. Wenger[6], J. Jeong[7], F. Gurrieri[8], D.B. Everman[9], C.E. Schwartz[9,10], J.L.R. Rubenstein[11], O.S. Birk[4], G. Bejerano[6,12], S. Lomvardas[3], R.Y. Birnbaum[1,2].* 1) Department of Bioengineering and Therapeutic Sciences, UCSF, San Francisco, CA, USA; 2) Institute for Human Genetics, UCSF, San Francisco, CA, USA; 3) Department of Anatomy, UCSF, San Francisco, CA, USA; 4) The Morris Kahn Laboratory of Human Genetics, NIBN, Ben-Gurion University, Beer-Sheva, Israel; 5) Department of Genetics,Stanford University, Stanford, CA, USA; 6) Department of Computer Science, Stanford University, Stanford, CA, USA; 7) Department of Basic Science and Craniofacial Biology, New York University College of Dentistry, New York, NY, USA; 8) Istituto di Genetica Medica, Università Cattolica S. Cuore, Rome, Italy; 9) JC Self Research Institute, Greenwood Genetic Center, Greenwood, SC, USA; 10) Department of Genetics and Biochemistry, Clemson University, Clemson, SC, USA; 11) Department of Psychiatry, UCSF, San Francisco, CA, USA; 12) Department of Developmental Biology, Stanford University, Stanford, CA, USA.

Enhancers are essential gene regulatory elements whose alteration can lead to morphological differences between species, developmental abnormalities and human disease. Current strategies to identify enhancers focus primarily on noncoding sequences. Here, using available ChIP-seq datasets, followed by zebrafish and mouse enhancer assays, we demonstrate that enhancers can also reside within coding exons and regulate their neighboring genes. Using 3C and FISH, we further show that one of these exonic enhancers, *Dync1i1* exon 15, regulates *Dlx5*/*6* limb expression 900kb away and its removal by chromosomal abnormalities in humans can cause split hand and foot malformation 1 (SHFM1). While genes make up less than 2% of our genome, we found that ~5% of functional enhancers could reside in coding exons, alluding to a potential novel mechanism of gene regulation both for the regulated gene and the gene encompassing the exonic enhancer. These results demonstrate that DNA sequences can function both as coding exons and as enhancers of other genes, suggesting that phenotypes resulting from coding mutations could be caused not only by protein alteration, but also by disrupting the regulation of another gene.

14

Genome copy number variation landscape in 68,000 humans and relevance to complex disease. *J.T. Glessner[1,2], D. Hadley[1], K. Wang[1,3], J. Bradfield[1], C. Kim[1], F. Mentch[1], H. Qiu[1], E. Frackelton[1], J. Li[1], C. Hou[1], F.G. Otieno[1], K. Thomas[1], K. Seidler[1], R. Chiavacci[1], J. Connolly[1], G. Lyon[1], L. Tian[1], B. Keating[1], P.M.A. Sleiman[1], S.F.A. Grant[1,2], M. Li[2], H. Hakonarson[1,2].* 1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA; 2) University of Pennsylvania School of Medicine, Philadelphia, PA, 19104, USA; 3) University of Southern California, Los Angeles, CA, 90007, USA.

Copy number variation (CNV) effects many dosage-sensitive genes but is also an abundant phenomenon in healthy individuals. We sought to deeply characterize rare CNV by assembling the largest population CNV map of the human genome through the study of 68,028 individuals from four populations with ancestry in Europe(52,321), Africa(12,548), Asia(2,299), and Latin America(860). We processed genotype and intensity data for CNV detection using Illumina single-nucleotide polymorphism (SNP) genotyping arrays intersection set of 520,017 SNPs. CNVs called per individual averaged 18.6. The length of the CNV calls averaged 68 Kb. The average individual CNV burden was 600 kb with rare CNV component of 200 kb. By mapping individual CNV calls into population statistics, 5,378 copy number variable regions (CNVRs) were identified, with deletions covering 2.35 gigabases (78% of the genome) and duplications covering 2.46 gigabases (82% of the genome), showing the pervasive nature of CNV. While many CNVRs were rare, 4,969 deletion, 2,633 duplication, and 263 homozygous deletion CNVRs were detected in multiple unrelated individuals. Reported loci in disease genotype associations were present in 1,409 of CNVRs identified. The median deletion CNV frequency of these CNVRs was 0.05% and the average was 0.22%. The median duplication CNV frequency of these CNVRs was 0.06% and the average was 0.21%. We found 964 deletion and 343 duplication CNVRs that were not reported in the DGV. 64% of CNVRs overlapped genes, which were functionally enriched for growth factor mediation, signal transduction and intermediate filaments with European population enriched CNVRs impacting regulation of fibrinolysis and intrinsic prothrombin activation. Distribution of CNVRs with enriched Illumina beadchip version, ethnicity, and sample source were investigated. Segmental duplications, genes, and disease associated regions were also impacted by many CNVs. Linkage disequilibrium between common SNP genotypes and rare CNVs was poor. In addition to healthy subjects, broad disease categories of cancer, autoimmune, cardiovascular, and neurological populations CNV frequency provides high statistical power for comparison. We investigate the frequency of CNVs in loci across the genome, how frequency differs between populations, the process of evolution through gene family extension enabled by CNVs, and applications to understanding of disease from the standpoint of CNVs impact on gene networks.

15

Impact of rare copy number variation (CNV) in autism spectrum disorders - evidence from 2,000 screened trio families. *D. Pinto[1,2], S.W. Scherer[1,2], Autism Genome Project Consortium (AGP).* 1) Gen & Genome Biol, Hosp Sick Children, Toronto, ON, Canada; 2) The Centre for Applied Genomics, Toronto, ON, Canada.

The autism spectrum disorders (ASDs) are a group of early-onset conditions that affect 0.6% of the general population. ASDs are characterized by impairments in reciprocal social interaction and communication, and the presence of restricted and repetitive behaviours. Individuals with an ASD vary greatly in cognitive development, which can range from above average to intellectual disability. ASDs have a substantial genetic aetiology, but the underlying genetic determinants are still largely unknown. Recent association studies provide only weak evidence for common allele risk effects. In contrast, the role of rare variants in ASDs is being increasingly recognized with recent surveys for CNVs and emerging sequencing efforts. In Phase II of the Autism Genome Project (AGP), we recently reported on the genome-wide characteristics of rare (<1% frequency) CNV in ~1,000 ASD trios using the Illumina 1M SNP platform (Pinto et al. 2010). Here we report on an additional independent collection of 1,000 ASD trios screened using a similar analysis pipeline. After stringent quality control filters, a total of 985 children and their parents were available for analyses, with a 3:1 ratio of singletons vs. cases from multiple incidence (MI) families. The de novo CNV rate was estimated to be 4.0% (73/1,854) with 5.6% of trio families (48/862) having at least one de novo CNV. Out of these 48 families with at least one de novo CNV, 68% are singletons, 6% are MI and another 31% are unknown (extended family not evaluated for ASD). Taken together, the integration of the two CNV sets from up to 2,000 ASD trios highlights an extremely heterogeneous genomic architecture of ASD. Hallmarks of these CNVs included combinations of de novo and inherited events in a given family, incomplete penetrance, and non-segregation in families, often implicating many novel ASD genes. Our results reveal additional new genetic and functional targets in ASD that point towards connected pathways in brain development. We expect that integration of the two CNV sets coupled with deeper phenotyping, and additional genomic analysis such as sequencing, will further aid in establishing genotype-phenotype correlations.

16

A duplication CNV that protects against metabolic syndrome. *W. Gu[1], M. Lacaria[1], P. Saha[2], L. Potocki[1,4], W. Bi[1], J. Yan[1], S. Girirajan[5], B. Burns[5], K. Walz[7], S. Elsea[5,6], L. Chan[2], J. Lupski[1,3,4].* 1) Molecular and Human Genetics, Baylor College Med, Houston, TX; 2) Diabetes & Endocrinology Research Center,Baylor College Med, Houston, TX; 3) Department of Pediatrics,Baylor College Med, Houston, TX; 4) Texas Children's Hospital, Houston, TX; 5) Department of Molecular and Human Genetics, Virginia Commonwealth University, Richmond, VA; 6) Department of Pediatrics, Virginia Commonwealth University, Richmond, VA; 7) Department of Human Genetics, University of Miami, Miami, FL.

About 400 million people worldwide are obese and are likely to suffer from premature mortality and associated morbidities such as the metabolic syndrome (MetS). MetS includes the clustering of metabolic risk factors such as hyperglycemia, dyslipidemia, hypertension and obesity. Both obesity and MetS have genetic origins, but the specific genetic factors remain largely unknown. Single nucleotide polymorphisms (SNP) identified through linkage and GWAS explain only ~10% of variation in obesity. Smith-Magenis syndrome (SMS) is a multiple congenital anomalies syndrome associated with obesity and hypercholesterolemia. SMS usually results from a deletion CNV at chromosome 17p11.2, but in some cases due to haploinsufficiency of a single gene *RAI1*. The reciprocal duplication in 17p11.2 causes Potocki-Lupski syndrome (PTLS), for which lowered total cholesterol (TC) and LDL were reported. By chromosomal engineering, we previously constructed mouse strains with a deletion (*Df(11)17*/+) or duplication (*Dp(11)17*/+) of the mouse genomic interval syntenic to the human SMS/PTLS region, respectively. We demonstrate that the *Dp(11)17* CNV is obesity-opposing; it conveys a highly penetrant, strain independent phenotype of reduced body weight, leaner body composition, lower TC/LDL and increased insulin sensitivity that is not due to alteration in food intake or activity level. When fed with a high-fat diet, *Dp(11)17*/+ mice display much less weight gain and metabolic change than WT mice, showing the *Dp(11)17* CNV to be protective against metabolic syndrome. Reciprocally, the *Df(11)17*/+ mice with the deletion CNV have increased body weight, higher fat content, decreased HDL and reduced insulin sensitivity, manifesting a bona fide metabolic syndrome. This finding is corroborated by human data obtained from 76 SMS subjects. By utilizing *Rai1* knockout and transgenic mice, we further showed that the metabolic consequences of *Dp(11)17*/+ and *Df(11)17*/+ CNVs are not the sole result of the dosage change of *Rai1* gene, although *RAI1/Rai1* has been considered the “main” dosage-sensitive gene for SMS and likely also PTLS. This study illustrates the contribution of CNV to the physiology of body weight control and energy metabolism in both mice and humans. The high penetrance and strain independence of these contributions is unique and further highlights the potential importance of CNV in the etiology of both obesity and MetS as well as in the protection from these traits.

17

Identification and characterization of structural variation breakpoints in an individual human genome sequence. *A.W.C. Pang[1,2], J.R. MacDonald[1], O. Migita[1], L. Feuk[1,3], S.W. Scherer[1,2].* 1) The Centre for Applied Genomics, The Hospital for Sick Children, 101 College Street, Toronto, Ontario M5G 1L7, Canada; 2) Department of Molecular Genetics, University of Toronto, 1 King's College Circle, Toronto, Ontario M5S 1A8, Canada; 3) Department of Immunology, Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Uppsala 75185, Sweden.

Delineating the content of variation in an individual's DNA is fundamental to comprehensive genetic studies. Despite the advances in technology, the proportion of structural variant (SV) studies that have resolved variant breakpoints at the nucleotide level is low, limiting the understanding of the population genetics features, as well as the molecular mechanism leading to SV formation. Moreover, the number of studies focusing on detecting inversions is most notably underrepresented. In the first sequenced personal human genome (J. Craig Venter's DNA, or HuRef), over 800,000 SVs have been discovered, yielding a total nucleotide difference from the consensus reference sequence by ~1.5%. In the current study, we attempted to refine the breakpoint resolution of known HuRef SVs, to examine the allele frequency of a subset of these variants in the human population, and to deduce their likely mutational mechanism of origin. First, we used NimbleGen sequence capture followed by 454 sequencing to target the boundaries for 3,693 HuRef SVs. Using multiple computational methods, we delineated the breakpoint junction sequences for 100 insertions, 42 deletions and 12 inversions, with sizes ranging from 550 bp to 1.1 Mb. Combining these results with our previous studies, there are 415,719 gains, 389,374 losses and 97 inversions in HuRef that have been mapped at the base-pair level. In addition, we are now conducting PCR assays to evaluate a random set of inversions (>30) and gains and losses (18) on a panel of 40 HapMap human (10 European, 10 Yoruba, 10 Chinese, 10 Japanese), three chimpanzee and one orangutan samples. This genotyping effort will allow us to examine the prevalence of the variant alleles, identify variant allele frequencies in different populations, and deduce their ancestral state. Precise breakpoint information enables the examination of sequence features to infer potential mechanisms contributing to the origin of SVs. We are identifying reported sequence signatures of various mechanisms: variable number of tandem repeats (VNTR); retrotransposons; non-allelic homologous recombination (NAHR); nonhomologous end joining (NHEJ); and microhomology-mediated break-induced replication (MMBIR). Determining the relative proportion of these mechanisms will enable us to better understand the roles of various mutational forces in shaping the human genome.

18

Sequencing of Isolated Sperm Cells for Direct Haplotyping of a Human Genome. *E.F. Kirkness[1], R. Grindberg[1], J. Yee-Greenbaum[1], C.R. Marshall[2], S.W. Scherer[2], R.S. Lasken[1], J.C. Venter[1].* 1) J Craig Venter Inst, Rockville, MD; 2) The Centre for Applied Genomics, The Hospital for Sick Children, 101 College Street, Toronto, Ontario, Canada.

There is increasing evidence that the phenotypic effects of genomic sequence variants are best understood in terms of variant haplotypes rather than as isolated polymorphisms. Haplotype analysis is also critically important for uncovering population histories, and for the study of evolutionary genetics. Although the sequencing of individual human genomes to reveal personal collections of sequence variants is now well established, there has been slower progress in the phasing of these variants into pairs of haplotypes along each pair of chromosomes. To date, complete haplotyping of an individual human genome has not been achieved. Previously, we described long-range haplotyping of heterozygous variants in a sequenced genome using the physical linkage of paired sequence reads from the ends of cloned DNA fragments. Here, we have developed a distinct approach to haplotyping that does not rely on paired sequence reads, but can yield chromosome-length haplotypes that include the majority of heterozygous SNPs in an individual human genome. This approach exploits the haploid nature of sperms cells, and employs a combination of genotyping and low-coverage sequencing of isolated sperm cells on a short-read platform. The combination of beadarray-derived genotypes and sequencing-derived genotypes permitted phasing of 1.27 million heterozygous SNPs, or 65% of the known complement of an individual human genome, with an average resolution of 2.4 kb. Owing to the phased haplotyping for the majority of heterozygous SNPs, it was possible to identify genes with potentially harmful mutations on both chromosomal copies.

19

Full Exome Sequencing of Autism Cases, Families, and Controls. *B.M. Neale[1,2,9], C. Stevens[2,9], A. Sabo[3,9], E. Lim[1,2], K. Samoocha[1,2], A. Kirby[1], S. Purcell[1,2], J. Flannick[2], D. Muzny[3], I. Newsham[3], U. Nagaswamy[3], Y.Q. Wu[3], M. Wang[3], J. Reid[3], E. Boerwinkle[3], C. Boyko[2], S. Ripke[1,2], M. Rivas[2,4], M. Velankar[3], L. Wang[5], S. Gabriel[2], J. Buxbaum[6], B. Devlin[7], G. Schellenberg[5], J. Sutcliff[8], R. Gibbs[3], M.J. Daly[1,2], ARRA Autism Sequencing Project.* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Richard Simches Bldg CPZN 6818, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Baylor College of Medicine, Houston, TX; 4) Wellcome Trust Center for Human Genetics, Cambridge, UK; 5) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia AP; 6) Mount Sinai School of Medicine, New York, NY; 7) Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA; 8) Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN; 9) These Authors Contributed Equally.

We conducted whole-exome sequencing on a case control cohort of approximately 1,000 cases and 1,000 controls. For half of these samples, data were generated at Baylor College of Medicine, using Nimblegen capture and ABI Solid Sequencing and the other half were generated at the Broad Institute of MIT and Harvard using Agilent capture and Illumina GAII and HiSeq technologies. We will describe the alignment, variant calling and general quality control analyses of both of these datasets including a description of the relative merits of both capture and sequencing technologies. As part of our study design, we have sequenced a handful of samples at both centers, enabling a more comprehensive comparison of genotype calls. We will also demonstrate that these two datasets are largely comparable in terms of more common variation, indicating that even though the entire data generation pipeline is radically different between these two sites, the assessment of common variation shows striking consistency. We have conducted a range of association tests on these two datasets including variant-specific and gene-based approaches. For variant-specific analyses, we are primarily interested in the variation that is too rare to have been effectively captured using genome-wide association (i.e. frequency less than 5 percent), with a particular focus on obviously functional variation. For gene-based analysis, we have calculated burden of nonsense and missense mutations for autism as well as applied C-alpha approach. We demonstrate that the genetics for autism must be complex from the case-control analysis. In addition to the case-control sequencing, we have also generated data on 100 trios, which enables the assessment of de novo mutations for risk to autism. We identified de novo mutations at approximately the rate of 1 per trio and that these mutations follow a Poisson distribution across the families analyzed. We have assessed the extent to which the genes that carry de novo mutations are excessively connected via protein-protein interactions.

20

Exome and targeted sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *B.J. O'Roak[1], E. Karakoc[1], L. Vives[1], S. Girirajan[1], I. Stanaway[1], A. Kumar[1], E.H. Turner[1], M.J. Rieder[1], D.A. Nickerson[1], R. Bernier[2], J. Shendure[1], E.E. Eichler[1,3].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA; 3) Howard Hughes Medical Institute, Seattle, WA.

We performed sequencing of 420 exomes (140 parent-child trios), copy-number variation (CNV) analysis and targeted sequencing of candidate genes in 1,700 probands to investigate the genetic basis of sporadic autism spectrum disorders (ASD). The genetic basis of ASD has proven difficult in the face of widespread genetic heterogeneity, but several lines of evidence have suggested a major role of *de novo* mutations especially in simplex or sporadic ASD cases. Previously, we sequenced a small subset of trios (n = 20) (O'Roak *et al. Nat Genet.* 2011) and identified *de novo* disruptive mutations in several outstanding candidates genes (*FOXP1*, *GRIN2B*, *SCN1A* and *LAMC3*). In addition, by integrating the global burden of all types of genetic variation in these samples, we explained some of the variability in phenotypic outcome. We have now sequenced an additional 120 trios consistent with sporadic ASD using an expanded exome definition and performed a detailed CNV analysis focusing on *de novo* mutation events. Overall, ~1.2 *de novo* events were identified per trio. Several mutations were found in genes involved with normal neuronal/cortical development as well as genes previously implicated in syndromes such as Coffin-Lowry and CHARGE. In the full set of 140 probands, no individual locus is recurrently disrupted highlighting the significant genetic heterogeneity within ASD. However, we identified recurrent *de novo* events in the ubiquitination and Wnt signaling pathways, suggesting that common biologic themes exist. To allow for detailed analysis of candidate genes, we developed a modified molecular inversion probe based protocol that allows for the parallel, highly cost-effective capture of hundreds of exons from 50 ng of DNA and 384-plex sample sequencing on the Illumina HiSeq2000 platform. To test this method, we performed capture and are currently sequencing 1,700 probands across six candidate genes (target 24 kb). Compared with exome data, we find strong genotypic concordance, increased sensitively, and target coverage >97%. Already we have found additional disruptive *de novo* events in *LAMC3* and *GRIN2B.* Based on these results, we are performing capture on 20 additional top candidate genes. We believe that trio-based exome sequencing and CNV analysis combined with efficient screening of candidate genes in several thousand additional cases will be a powerful path to unraveling the genetic basis of ASD.

21

Whole exome sequencing identifies novel changes in *AP4M1*, *CDKAL1* and *SYNGAP1* in extended multiplex autism families. *H.N. Cukier[1], S.H. Slifer[1], J.M. Jaworski[1], P.L. Whitehead[1], J.L. Robinson[1], I. Konidari[1], W.F. Hulme[1], H.H. Wright[2], R.K. Abramson[2], J.E. Dallman[3], J.L. Haines[4], M.L. Cuccaro[1], J.R. Gilbert[1], M.A. Pericak-Vance[1].* 1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) University of South Carolina School of Medicine, Columbia, SC; 3) Department of Biology, University of Miami, Coral Gables, FL; 4) Center for Human Genetics Research, Vanderbilt University, Nashville, TN.

There is a strong genetic component to autism, but studies to date have demonstrated that the underlying genetic architecture is complex, with numerous genes potentially involved. One strategy for uncovering these genes is the use of large, extended families with multiple, distantly related, affected individuals. We performed next-generation sequencing of whole exomes to identify potential causative variants in a minimum of two pairs of affected individuals across 10 multiplex autistic families. We used the Agilent SureSelect Human All Exon kit and the Illumina HiSeq 2000 in 2x100 paired end runs, for an average depth of 58x. Variants within each family were filtered to include only novel, heterozygous alterations shared amongst all affected individuals and predicted to be detrimental, either by altering amino acids or splicing patterns. We then further refined our genes of interest by merging with identical by descent (IBD) regions delineated with high density SNP genotyping data. Preliminary analysis of the dataset identified three such changes that we validated by Sanger sequencing, each within a distinct autism family. The first is an I1115T change in *SYNGAP1* (*synaptic Ras GTPase activating protein 1*) that was identified in second degree affected cousins. *SYNGAP1*, a gene previously associated with autism and mental retardation, encodes a brain specific protein shown to play a role in spine formation and synaptic plasticity. A second alteration shared between two first cousins was identified in *AP4M1* (*adaptor-related protein complex 4, mu 1 subunit*) at R271C. This gene is thought to act in cerebellar development and has previously been found altered in patients with tetraplegic cerebral palsy and mental retardation. Finally, a R29Q change in *CDKAL1* (*CDK5 regulatory subunit associated protein 1-like 1*) was found in a family with three affected cousins in distinct family branches. *CDKAL1* has been associated with Type 2 diabetes. Functional analyses of these variants are underway. The fact that we were able to identify three alterations of interest, two which have already demonstrated neuronal functions and neurological phenotypes, supports the hypothesis that the analysis of extended autism families is a successful approach to recognizing additional novel genes that contribute to the etiology of autism.

22

Characterization of the Function and Regulation of the Autism Susceptibility 2 (*AUTS2*) Gene. *N. Oksenberg[1,2], L. Weiss[2,3], N. Ahituv[1,2].* 1) Department of Bioengineering and Therapeutic Sciences, UCSF, San Francisco, CA, USA; 2) Institute for Human Genetics, UCSF, San Francisco, CA, USA; 3) Department of Psychiatry, UCSF, San Francisco, CA, USA.

Heterozygous chromosomal aberrations in the autism susceptibility candidate 2 (*AUTS2*) gene region, some of which remove only intronic regions, have been identified in individuals with autism spectrum disorders (ASDs). In addition, *AUTS2* was shown to encompass the most significant accelerated genomic region differentiating humans from Neanderthals (Green et al., *Science* 2010, 328 : 710-722). However, the function and regulation of this gene remain largely unknown. To characterize *AUTS2* function, we used morpholinos (MOs) to knock-down *auts2* in zebrafish. Two different *auts2* MOs led to a smaller head size, decreased mobility and a loss of neurons in the midbrain. To characterize the regulation of *AUTS2*, we set out to identify tissue specific enhancers that could regulate this gene. Using comparative genomics and available ChIP-seq data sets we scanned the human-Neanderthal accelerated region and *AUTS2* intron 4, which encompasses chromosomal deletions associated with ASD, for potential enhancer sequences. Using transgenic enhancer assays in zebrafish, we found 13 of the 36 tested sequences to be enhancers in the central nervous system (CNS) and to overlap with *auts2* expression. Several selected sequences in this region were also tested and found to be positive CNS enhancers in mice. Combined, our results show that *AUTS2* is an important neurodevelopment gene. In addition, they provide a regulatory map for CNS enhancers that could regulate *AUTS2*, revealing candidate sequences where nucleotide variation could lead to ASD susceptibility and human specific traits.

23

Mutations in the *SHANK1* synaptic scaffolding gene in autism spectrum disorder and intellectual disability. *D. Sato[1], C.R. Marshall[1], A.C. Lionel[1,2], A. Prasad[1], D. Pinto[1], J.L. Howe[1], I. O'Connor[3], G.A. Rappold[4], V. Endris[4], R. Roeth[4], J.L. Michaud[5], F.F. Hamdan[5], B. Fernandez[6], W. Roberts[7,8], P. Szatmari[3], S.W. Scherer[1,2].* 1) The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Molecular Genetics, and the McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada; 3) Offord Centre for Child Studies, Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario, Canada; 4) Department of Molecular Human Genetics, Ruprecht-Karls-University, Heidelberg, Germany; 5) Centre Hospitalier Universitaire (CHU) Sainte-Justine Research Center in the center of Excellence in Neuromic, Université de Montréal, Montréal, Quebec, Canada; 6) Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St. John's, Newfoundland, Canada; 7) Autism Research Unit, The Hospital for Sick Children, Toronto, Ontario, Canada; 8) Bloorview Kids Rehabilitation, University of Toronto, Toronto, Ontario, Canada.

Recent studies have highlighted the involvement of rare copy number variations (CNVs) and point mutations in the genetic etiology of autism spectrum disorder (ASD), particularly affecting genes involved in the neuronal synaptic complex (e.g. *NRXN1*, *NLGN3*, *NLGN4*, *SHANK2* and *SHANK3*). The SHANK gene family consists of three members (comprising *SHANK1*, *SHANK2* and *SHANK3*), which encode scaffolding proteins required for the proper formation and function of neuronal synapses. While *SHANK2* and *SHANK3* mutations have been implicated in ASD and intellectual disability (ID) and *Shank1* knock-out mice have been reported to exhibit elements of ASD phenotype, the involvement of human *SHANK1* was not yet established. Due to the high degree of DNA sequence identity and similar gene expression profiles across the SHANK gene paralogues, we hypothesized that mutations at the *SHANK1* locus contribute to ASD in a subset of subjects. CNV analysis of microarray data from 996 individuals with ASD led to discovery of a deletion removing most exons of *SHANK1* in a male individual with ASD. Additional genotyping revealed that this deletion segregated with ASD phenotype in other males from the same multi-generation family. Deletions at this locus were not detected in microarray data from 10,453 control individuals. Following up on these findings, we sequenced all exons and splice sites of the *SHANK1* gene in 326 unrelated ASD cases and 340 ID cases to look for point mutations, which would provide additional evidence for the gene's involvement in ASD etiology. We detected 26 rare missense variants in the cases and assessed their population frequencies by means of TaqMan assay testing in 285 control individuals. Taken together, our data provides the first report of *SHANK1* mutations in ASD and adds to the evidence implicating neuronal synaptic genes as etiologic factors.

24

Mutation or deletion of the epigenetic regulator, *MBD5*, causes intellectual disability, epilepsy, and autism spectrum disorder. *S.V. Mullegama[1], M.E. Talkowski[2,3,4], J.A. Rosenfeld[5], B.W.M. van Bon[6], Y. Shen[4,7], E.A. Repnikova[8], J. Gastier-Foster[8,9,10], D.L. Thrush[8,10], C. Chiang[4], C. Ernst[4], A. Lindgren[11], C.C. Morton[3,11], C. Astbury[8,9], L.A. Brueton[12], K.D. Lichtenbelt[13], L.C. Ades[14], M. Fichera[15], C. Romano[16], J.W. Innis[17], C.A. Williams[18], D. Bartholomew[19], M.I. Van Allen[20], A. Parikh[21,22], L. Zhang[21,22], R.E. Pyatt[8,9], L.G. Shaffer[5], S. Schwartz[23], B.B.A de Vries[6], J.F. Gusella[2,3,4], S.H. Elsea[1,24].* 1) Department of Human and Molecular Genetics , Virginia Commonwealth University School of Medicine, Richmond, VA; 2) Departments of Genetics and Neurology, Harvard Medical School, Harvard University, Cambridge, MA; 3) Program in Medical and Population Genetics, Broad Institute of Harvard and M.I.T, Cambridge, MA; 4) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 5) Signature Genomic Laboratories, Spokane, WA; 6) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 7) Department of Laboratory Medicine, Children's Hospital of Boston and Department of Pathology, Harvard Medical School, Boston, MA; 8) Department of Pathology and Laboratory Medicine, Nationwide Children's Hospital, Columbus, Ohio; 9) Department of Pathology, The Ohio State University, Columbus, OH; 10) Department of Pediatrics, The Ohio State University, Columbus, OH; 11) Departments of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hopsital, Boston, MA; 12) Clinical Genetics Unit, Womens Hospital, Edgbaston Birmingham, UK; 13) Department of Medical Genetics, University Medical Centre Utrecht, Utrecht, The Netherlands; 14) Discipline of Pediatrics and Child Health, University of Sydney, Sydney, Australia; 15) Laboratory of Genetic Diagnosis, IRCCS Associazione Oasi Maria Santissima, Troina, Italy; 16) Pediatrics and Medical Genetics, IRCCS Associazione Oasi Maria Santissima, Troina, Italy; 17) Division of Pediatric Genetics, Department of Pediatrics, University of Michigan, Ann Arbor, MI; 18) Department of Genetics and Metabolism, Department of Pediatrics, University of Florida College of Medicine, Gainesville, FL; 19) Division of Molecular and Human Genetics, Nationwide Children's Hospital, Columbus, OH; 20) Department of Medical Genetics, University of British Columbia, Vancouver, BC; 21) Department of Genetics, Case Western Reserve University School of Medicine, Cleveland, OH; 22) Department of Pediatrics, Case Western Reserve University School of Medicine, Cleveland, OH; 23) Lab Corp, USA; 24) Department of Pediatrics, Virginia Commonwealth University School of Medicine, Richmond, VA.

Autism spectrum disorder (ASD) encompasses a wide range of neurodevelopmental disorders that include impaired communication skills and social interactions and repetitive and stereotyped behaviors. ASD has a strong but highly complex genetic etiological component wherein visible chromosomal abnormalities (~5%), CNVs (10-20%), and single-gene disorders (~5%) are contributors. We report a detailed genotype-phenotype correlation analysis of 66 2q23.1 microdeletion syndrome cases exhibiting intellectual disability, epilepsy, and autistic features with microdeletions ranging from 38 Kb to 19.3 Mb. Delineation of the critical region revealed a single gene, *MBD5*, an important protein in methylation patterning and epigenetic regulation. Deletions affecting only *MBD5* included 16 intragenic alterations and 2 translocations disrupting *MBD5* transcription. Expression profiling confirmed haploinsufficiency of *MBD5* in all cases resulting from deletion of the entire gene, as well as those restricted to the 5'-untranslated region. The size of our cohort enabled identification of a core set of phenotypic features common to all individuals with disruption of *MBD5*, including intellectual disability, developmental delays, epilepsy, language impairment, and behavioral problems, with most individuals meeting diagnostic criteria for ASD. We further surveyed coding variation in 747 ASD cases by next-generation pooled sequencing and identified missense changes at two conserved amino acid positions within critical functional domains, including a previously described mutation in the MBD binding domain observed in 6 independent cases (G79E). Taken together, these results confirm that *MBD5* is the primary causative gene for 2q23.1 microdeletion syndrome and that haploinsufficiency of *MBD5* contributes to the majority of phenotypic features present in this disorder. Furthermore, mutations of *MBD5* in ASD cases provide evidence for a significant role for *MBD5* in the etiology of ASD. These findings support a role for *MBD5* in ASD and neuronal development by mediating central nuclear functions important for regulation of gene dosage and expression. Finally, since *MBD5* alone causes a distinct syndromic disorder, it is worthy of consideration in the diagnostic evaluation of individuals with ASD complicated by epilepsy and intellectual disability.

25

Genome-wide DNA Methylation Profiling of Monozygotic Twins Discordant for Autism Spectrum Disorder. *C.C.Y. Wong[1], L.C. Schalkwyk[1], E.L. Meaburn[1,2], A. Ronald[1,2], T.S. Price[1], R. Plomin[1], J. Mill[1].* 1) MRC Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, UK; 2) Birkbeck, University of London, UK.

Autism Spectrum Disorder (ASD) is a group of common, complex neurodevelopmental disorders that share common features of social difficulties, communicative impairments and repetitive behaviours and interests. Although the etiology of ASD has a strong inherited component, monozygotic (MZ) twins are sometimes discordant for diagnosis indicating the involvement of non-shared environmental influence which could include epigenetic mechanisms. Since MZ twins share the same DNA code and a common environment, disease-discordant MZ twin pairs are an excellent model for elucidating the contribution of environmentally-mediated epigenetic factors on disease. To investigate the role of DNA methylation in ASD, we assayed peripheral blood DNA samples obtained from a sample of MZ twin pairs discordant for ASD and ASD-like traits as well as concordant controls (n= 53 twin pairs, 106 individuals) using the Illumina 27K Human Methylation array. Our methylomic profiling revealed numerous significant and substantial ASD-associated DNA methylation differences across the genome within our MZ discordant twin pairs. Significant DNA methylation differences between ASD cases and unrelated controls were also observed. Further bioinformatic analysis of the most differentially methylated genes from our discordant twin pair analyses demonstrated an enrichment of disease-relevant networks, including biological processes that have been previously implicated in autism. The methylomes of six ASD-discordant MZ twin pairs were also assessed using methylated DNA immunoprecipitation combined with ultradeep sequencing (MeDIP-seq), again uncovering widespread epigenetic changes associated with disease. Finally, we also investigated the role of differential allele-specific methylation (ASM) in MZ twin pairs discordant for ASD using an array-based method that quantitatively assesses ASM in amplicons covering 7.6% of the human genome. Significant and sizeable changes in allelic methylation patterns were identifed between the affected and the unaffected twins. Taken together, our study represents the first in-depth epigenomic analyses of MZ twins discordant for ASD and our findings confirm the imperative role of DNA methylation in the etiology of ASD.

26

Sequencing based comprehensive genome and transcriptome analyses of Velocardiofacial Syndrome. *A.E. Urban[1], Y. Zhang[2], X. Zhu[3], D.F. Levinson[3], S.M. Weissman[2].* 1) Psychiatry and Genetics Departments, Stanford University, Palo Alto, CA; 2) Genetics Department, Yale University, New Haven, CT; 3) Psychiatry Department, Stanford University, Palo Alto, CA.

Velocardiofacial Syndrome (VCFS) is a relatively common genomic disorder, typically caused by a heterozygous 3 Mbp deletion on chromosome 22q11. There are various developmental defects including neurodevelopmental disorders such as schizophrenia, autism spectrum disorders (ASD) and learning disabilites; however, the combinations and severity of symptoms vary widely between patients with no clear genetic basis for such variability. Earlier [Urban, Korbel et al., PNAS 2006], we developed very-high resolution array-CGH and used it to determine the exact breakpoints of a panel of VCFS deletions, even in regions of segmental duplication (SD). We discovered that even `typical3 Mbp' deletions can differ in their endpoints by up to several hundred thousand bp in length, affecting the copy number of many genes between patients. Then we developed a method to use 2nd-generation sequencing (i.e. “next-generation sequencing”) for the mapping of structural variation in the human genome (paired-end mapping, PEM) [Korbel, Urban, Affourtit et al., Science 2007]. We found hundreds of small to medium size sequence variations in the normal human genome. Many of those are in functionally relevant regions genome wide - and also, additional smaller sequence variants are enriched in 22q11. These findings point to possible types of modifiers of the main deletion on 22q11: difference in deletion endpoints, small sequence variants within the deletion region on the non-deleted chromosome, and sequence variation genome wide. Here we report on multi-level genomics analyses of VCFS using 2nd -generation sequencing (Illumina HiSeq 2000). We have sequenced the genomes of a panel of cell lines from VCFS patients (paired-end, 4x or deeper genomic coverage) and catalogued their genomic sequence variants from 1 bp to several hundreds of kbp in size. For the same cell lines as well as for controls we carried out very deep RNA-Seq analysis (up to 100 million mapped 2x100bp paired-end constructs per cell line). These data allow us to correlate with very high accuracy the effects on gene expression patterns of both the copy number changes caused by the main deletion and the additional, patient specific, genomic sequence variants. We observe multiple small sequence variants, in the non-deleted 22q11 regions as well as genome-wide, with some variants potentially affecting loci of functional significance. The gene expression patterns show evidence of incomplete gene dosage compensation.

27

HDAC8 mutations in Cornelia de Lange syndrome. *M.A. Deardorff[1], M. Bando[3], K. Saitoh[3], T. Itoh[3], Y. Katou[3], M. Kaur[1], L. Francey[1], J.J. Wilde[1], S. Ernst[1], D. Clark[1], K.E. Cole[2], P.M. Lombardi[2], K. Takagaki[3], T. Hirota[3], N. Nozaki[3], P.J. Willems[10], G. Gyftodimou[6], M.B. Petersen[6], N. Tyschenko[11], E. DeBaere[7], G.R. Mortier[8], G. Gillessen-Kaesbach[5], V.M. Siu[9], D.W. Christianson[2], F.J. Kaiser[5], L.G. Jackson[4], K. Shirahige[3], I.D. Krantz[1].* 1) Children's Hospital of Philadelphia, Philadelphia, PA; 2) University of Pennsylvania, Philadelphia, PA; 3) University of Tokyo, Japan; 4) Drexel University School of Medicine, Philadelphia, Pennsylvania; 5) Institut für Humangenetik, Lübeck, Germany; 6) Institute of Child Health, Athens, Greece; 7) Ghent University Hospital, Ghent, Belgium; 8) Antwerp University Hospital, Edegem, Belgium; 9) University of Western Ontario, London, Ontario, Canada; 10) GENDIA, Antwerp, Belgium; 11) Technische Universität Dresden, Germany.

Cornelia de Lange syndrome is a dominantly inherited congenital malformation disorder caused by mutations in the cohesin regulatory protein NIPBL (~60%) and the core cohesin components, SMC1A (~5%) and SMC3 (n= 1). Recent work has demonstrated that the ESCO1 acetyltransferase is required for the acetylation of SMC3 to establish cohesiveness of chromatin-loaded cohesin during S-phase. Furthermore, SMC3 is deacetylated during mitosis, suggesting S-phase specificity of SMC3 acetylation in establishing cohesion, as well as the existence of a deacetylase to regulate this activity. Recently yeast HOS1, a class I histone deacetylase, has been reported to deacylate yeast SMC3 during anaphase. To identify a vertebrate SMC3 deactylase, we developed an acetylated SMC3 (SMC3-ac)-specific antibody and performed an RNAi screen of all known vertebrate deacetylases to identify HDAC8 as the likely SMC3 deacetylase. After verification of its function in sister chromatid cohesion and cohesin recycling assays, we screened for mutations in mutation-negative CdLS patients and identified 6 probands with loss-of-function mutations in the X-linked HDAC8 gene. Using an HDAC8-mutated cell line from a male proband, we performed ChIP-Seq of RAD21 and SMC3-ac and noted that there is a dramatic reduction in total cohesin binding. We then demonstrated that HDAC8 is required for efficient recycling of cleaved cohesin from chromosomes in anaphase. Without HDAC8 activity, the cleaved RAD21 N-terminal fragment is not released from SMC3 and the dysfunctional complex remains associated with chromosomes. This work suggests that acetylated SMC3 strongly binds RAD21 and that disruption of its dissociation interferes with stable cohesin reloading during G1. This work demonstrates a central role of HDAC8 as a vertebrate SMC3 deacetylase, refines the mechanism of acetylation in the regulation of cohesin and furthermore, is the first to demonstrate a mechanistic role for a lysine deacetylase in a human developmental disorder.

28

Mutations in *RAD21* as a cause of a new cohesinopathy. *F.J. Kaiser[1], D. Braunholz[1], J. Wilde[2], M.C. Gil-Rodriguez[1, 3], M. Albrecht[1], D. Clark[2], A. Rampuria[2], W. Xu[1], I.D. Krantz[2, 4], G. Gillessen-Kaesbach[1], H. Xu[5], J.A. Horsfield[6], M.A. Deardorff[2,4].* 1) Institut für Humangenetik, University of Lübeck, Lübeck, Germany; 2) The Children's Hospital Division of Genetics, Philadelphia, USA; 3) Unit of Clinical Genetics and Functional Genomics, Departments of Pharmacology-Physiology, Medical School, University of Zaragoza, Spain; 4) The University of Pennsylvania School of Medicine Department of Pediatrics, Philadelphia, USA; 5) Research Division, Peter MacCallum Cancer Centre, Melbourne, Australia; 6) University of Otago Department of Pathology, Dunedin, New Zealand.

Cornelia de Lange syndrome (CdLS) is a dominantly inherited multisystem developmental disorder that includes distinctive facial features, growth retardation, a range of limb defects and variable cognitive delay. Mutations in the *NIPBL* gene have been identified in approximately 75% of patients with severe classical CdLS but have been found in a far lower percentage of patients with mild or atypical features. NIPBL, along with several other regulatory proteins, is required for the establishment of sister chromatid cohesion through the loading of Cohesin, a multimer consisting of SMC1 and SMC3 and two clasp proteins, RAD21 and STAG. Following the identification of *NIPBL*, several mutations in *SMC1A* and one mutation in *SMC3* were detected in a small percent of variant CdLS cases. These patients have a milder phenotype with no major structural anomalies, although all have significant mental retardation. Using a unique collection of about 300 mutation-negative individuals with CdLS and overlapping clinical phenotypes, we undertook a genome-wide copy number variation (CNV) analysis. We have identified patients with chromosomal variants and clinical phenotypes that overlap CdLS and implicate *RAD21* as a candidate cohesinopathy gene. We extended this data to characterize several additional patients with overlapping deletions and have identified two de novo *RAD21* mutations in patients with features that diverge from classical CdLS. We have measured the expression and activity of these missense mutations in protein interaction, DNA damage and zebrafish developmental assays. In short, one of these *RAD21* mutations appears to function in a dominant-negative manner, increases binding to STAG proteins, and results in a more severe clinical phenotype than the other loss-of-function *RAD21* mutation. These data correlate with the features observed in the patients, helping to explain why patients with deletions of *RAD21* may have subtle clinical phenotypes. We will discuss the current understanding of the mechanisms by which Cohesin mutations may alter transcriptional profiles to result in multiple variant developmental phenotypes and speculate on clinical features that appear to be unique and common between different disorders of Cohesin.

29

Pitt Hopkins syndrome: further delineation of the neurological phenotype. Description of 32 novel patients and proposition of a Diagnostic Criteria Score. *S. Whalen[1], A. Jacquette[1, 2], T. Gaillon[3], O. Moldovan[4], M. Rossi[5], F. Devillard[6], F. Giuliano[7], G. Soares[8], M. Mathieu-Dramard[9], A. Afenjar[1, 10], C. Mignot[1, 10], L. Burglen[11], L. Van Maldergem[12], S. Aftimos[13], G. Mancini[14], P. Dias[4], N. Philip[15], A. Goldenberg[16], M. Le Merrer[17], D. Josifova[18], A. Van Hagen[19], D. Lacombe[20], P. Edery[5], S. Dupuis-Girod[5], MP. Alex[5], D. Sanlaville[21], M. Goossens[3], J. Amiel[17], D. Héron[1,2], I. Giurgea[3].*
1) Département de Génétique et Cytogénétique, Groupe Hospitalier Pitié Salpêtrière, Paris, France; 2) Centre de Référence Maladies Rares Déficiences Intellectuelles de causes rares; 3) Service de Biochimie et Génétique and INSERM U955 IMRB équipe 11, Hôpital Henri Mondor, Créteil; 4) Serviço de Genética Médica, Hospital de Santa Maria, Lisboa, Portugal; 5) Service Génétique Clinique, Groupement Hospitalier Est, Hôpital Femme Mère Enfant, Bron, France; 6) Laboratoire et Consultations de Génétique Chromosomique, CHU Grenoble, Grenoble, France; 7) Service de Génétique Médicale, CHU de Nice, France; 8) Centro Genética Médica Jacinto Magalhães, INSA, IP - Porto, Portugal; 9) Service de Néonatologie, Attachée de Génétique Clinique, Centre Hospitalier d'Amiens, Saint Quentin, France; 10) Service de Neuropédiatrie, Hôpital Armand Trousseau, Paris, France; 11) Service de Génétique et Embryologie Médicales, Hôpital Armand Trousseau, Paris, France; 12) Service de Génétique Médicale, CHU de Besançon, Besançon, France; 13) Northern Regional Genetic Service, Auckland City Hospital, Auckland, New Zealand; 14) Department of Clinical Genetics, Erasmus MC, Universitair Medisch Centrum Rotterdam, Rotterdam, Pays Bas; 15) Département de Génétique Médicale, CHU de Marseille, Hôpital La Timone, Marseille; 16) Service de Génétique Clinique, CHU de Rouen, Hôpital Charles Nicolle, Rouen; 17) Département de Génétique, Hôpital Necker-Enfants Malades, Paris, France; 18) Clinical genetics, Guy's and St Thomas Hospital, London, UK; 19) Klinische Genetica, VU Medisch Centrum, Amsterdam, Pays Bas; 20) Service Génétique Clinique, CHU Pellegrin, Bordeaux, France; 21) Laboratoire de Cytogénétique Constitutionnelle, Groupement Hospitalier Est, Hôpital Femme Mère Enfant, Bron, France.

Pitt Hopkins syndrome (PHS, MIM#610954) patients present with severe intellectual disability (ID) with no speech, typical facial gestalt and hyperventilation episodes. In 2007, *TCF4* was identified as the disease causing gene with de novo heterozygous mutations or deletions. Since, 63 patients have been reported. Here, we report a novel series of 32 patients diagnosed with PHS, compared to previously reported cases, and refine the neurological and morphological phenotype. Within a two years period we identified heterozygous de novo *TCF4* gene mutations or deletions in 32 patients with severe intellectual disability and the PHS facial gestalt. All patients were clinically evaluated with special attention to neurological presentation, with films whenever possible, and available brain-MRIs were reviewed. All patients in this series had severe ID with no speech. Thirty have facial features typical of the syndrome. A majority of the patients presented stereotypic movements (arm flapping, rapid movement of fingers, hand nibbling or rubbing and head rotation) and restless movements increasing with anxiety and agitation. Hyperventilation is frequent and often triggered by anxiety or excitement. Epilepsy and microcephaly are very inconstant findings. Cerebral MRI was abnormal in 50% of patients but show minor morphological changes only. PHS is not rare among patients with severe intellectual deficiency. We report the clinical features of 32 PHS patients, with special emphasis on the neurological phenotype and key diagnostic features in the Rett-like group of patients. A Diagnostic Criteria Score will be suggested including major features (severe ID with absent langage and delayed walking, morphological features, stereotypic movements), minor features (ataxic gait, hypotonia, hyperventilation, behavioural traits, constipation, strabismus, myopia) and exclusion criteria (severe microcephaly). Genotype-phenotype correlation will also be discussed.

30

Epidemiological features of Costello Syndrome and Cardio-facio-cutaneous Syndrome: findings from the first nationwide survey. *Y. Abe[1], Y. Aoki[1], S. Kuriyama[2], H. Kawame[3], N. Okamoto[4], K. Kurosawa[5], H. Ohashi[6], S. Mizuno[7], T. Ogata[8], S. Kure[9], T. Niihori[1], Y. Matsubara[1].*
1) Dept Med Genet, Tohoku Univ Sch Med, Sendai, Japan; 2) Dept Mol Epidemiolo, Tohoku Univ Sch Med, Sendai, Japan; 3) Dept Genet Counseling, Ochanomizu Univ, Tokyo, Japan; 4) Dept Med Genet, Osaka Med Ctr & Res Inst for Maternal & Child Health, Izumi, Osaka, Japan; 5) Dev Med Genet, Kanagawa Children's Med Ctr, Yokohama, Japan; 6) Dev Med Genet, Saitama Children's Med Ctr, Saitama, Japan; 7) Dept Pediatr, Central Hosp, Aichi Human Service Ctr, Kasugai, Aichi, Japan; 8) Dept Pediatr, Hamamatsu Univ Sch Med, Hamamatsu, Shizuoka, Japan; 9) Dept Pediatr, Tohoku Univ Sch Med, Sendai, Japan.

Costello syndrome and cardio-facio-cutaneous (CFC) syndrome are a group of genetic disorders which result from dysregulation of the RAS/MAPK cascade. Germline mutations in HRAS are causative for Costello syndrome and those in KRAS, BRAF and MAP2K1/2 (MEK1/2) cause CFC syndrome. Since our discovery of HRAS mutations in Costello syndrome (2005) and KRAS/BRAF mutations in CFC syndrome (2006), approximately 200 patients of each syndrome have been reported. However, clinico-epidemiological features of these disorders remain to be elucidated. In order to assess the prevalence, natural history, prognosis and tumor incidence, we conducted a nationwide prevalence study of patients with Costello syndrome and CFC syndrome in Japan. The study consisted of two-stage questionnaire surveys, which were distributed to a total of 1127 departments, including randomly selected pediatric and genetic departments at hospitals and institutions for severely-retarded children. The first survey inquired about the number of mutation-positive patients as well as clinically suspected patients with Costello syndrome and CFC syndrome. The second asked for detailed clinicoepidemiological information of each patient reported. The response rate of the first-stage survey was 76% (856/1127). Sixty-three patients with HRAS-positive Costello syndrome and 64 patients with CFC syndrome with mutations with KRAS, BRAF or MAP2K1/2 were reported, including 9 new patients with Costello syndrome and 8 new patients with CFC syndrome analyzed in the current study. The total numbers of patients with Costello syndrome and CFC syndrome in Japan were estimated as 105 (95% confidence interval, 84 to 127) and 160 (95% confidence interval, 88 to 232), respectively. The secondary survey revealed age distribution and the frequency of heart defects, mental retardation and tumor association. Sixteen adult patients aged 18-32 years were reported. Ten of the sixteen patients can walk alone and one of them had developed recurrent bladder papillomata. The prevalence of Costello syndrome and CFC syndrome were estimated as 1 in 1,230,000 population and 1 in 790,000 population, respectively. This is the first epidemiological study of both disorders. Identification of patients older than 32 years of age and follow up of the reported patients in the current study will be important to estimate the precise prevalence and natural history of these disorders.

31

Torwards the dissection of marfanoid syndromes with intellectual disability. *L. Faivre[1], S. Lambert[1], H. Dindy[2], C. Ragon[1], M. Payet[1], C. Francannet[3], Y. Sznajer[4], A. Megarbane[5], N. Philip[6], P. Collignon[6], S. Odent[7], L. Pasquier[7], A. Toutain[8], R. Missotte[9], C. Baumann[10], MA. Delrue[11], C. Goizet[11], C. Thauvin-Robinet[1], S. Julia[12], Y. Dulac[12], G. Jondeau[13], B. Aral[1], C. Boileau[2], P. Callier[1].* 1) Dept Genetics, Hosp d' Enfants, Dijon, France; 2) Hôpital Ambroise Paré, Paris, France; 3) CHU Clermont-Ferrand, France; 4) Hôpital Universitaire, Brussels, Belgium; 5) Université Saint-Joseph, Beirut, Lebanon; 6) Marseille, France; 7) CHU Rennes, France; 8) CHU Tours, France; 9) CHU, La Martinique, France; 10) Hôpital Robert Debré, Paris, France; 11) CHU Bordeaux, France; 12) CHU Toulouse, France; 13) Hôpital Bichat, Paris, France.

The term “marfanoid phenotype” is used to describe patients with skeletal signs suggestive of Marfan syndrome but who do not meet the international criteria. The association of a marfanoid phenotype and intellectual disability has been reported in the literature, in Lujan-Fryns syndrome (LJS), in Shprintzen-Goldberg syndrome (SGS) and in rare cases of Loeys-Dietz syndrome (LDS). Mutations of the X chromosome ZDHHC9, UPF3B and MED12 genes have been reported in a very small number of familial cases with LJS and de novo mutations in the FBN1 and TGFBR2 genes have been described in rare observations of SGS. Finally, chromosomal imbalances have also been reported in sporadic observations. The implication of each gene in marfanoid syndromes with intellectual disability is unknown and genetic counselling is difficult, particularly in sporadic cases. 100 patients (74 males, 26 females) with marfanoid phenotype and intellectual disability were recruited through the framework of the reference centres for rare diseases in France. A 244K specially designed CGH array (AgilentÒ) study has been performed in all patients, as well as direct sequencing of MED12, ZDHHC9, UPF3B, FBN1, TGFBR1, and TGFBR2 genes, and X-inactivation study in females. In the population of 75 patients with LJS, a de novo chromosomal rearrangement was found in 8 patients, including one 1.4 Mb 3q29 deletion that has already been reported in association with arachnodactyly and long/thin habitus; a de novo FBN1 mutation was found in 1 patient; a skewed X inactivation was found in 40% of females. In the population of 22 patients with cardiac or ocular features of the MFS spectrum, a de novo chromosomal rearrangement was found in 4 patients, including one 10 Mb del15q21.1q21.3 including FBN1, and 4 pathogenic FBN1 mutations. Interestingly, 2 of these FBN1 mutations were coexisting with a chromosomal microrearrangement. No abnormalities were found in the 3 patients with SGS. In conclusion, we identified a cytogenetic or molecular defect in 16% of patients. We showed that MED12, ZDHHC9, UPF3B genes are not major disease genes of such phenotypes and that submicroscopic rearrangements are the most prevalent anomalies. Finally, the search for aortic complications is mandatory even if the marfanoid habitus is associated with MR since FBN1 mutations were found in a subset of patients.

32

Disruption of a long-distance regulatory region of *SOX9* in isolated 46,XX and 46,XY disorders of sex determination. *S. Benko[1,2,3], C.T. Gordon[1,2], R. Sreenivasan[4], D. Mallet[5], C. Tauvin-Robinet[6], A. Brendehaug[7], S. Thomas[1,2], O. Bruland[7], D. David[8], D. Sanlaville[9], P. Callier[10], F. Huet[11], A. Molven[12,13], A. Munnich[1,2,14], L. Faivre[6], J. Amiel[1,2,14], V. Harley[4], G. Houge[7,15], Y. Morel[5], S. Lyonnet[1,2,14].* 1) INSERM U-781, Hôpital Necker-Enfants Malades, Paris, France; 2) Université Paris Descartes, Faculté de Médecine, Paris, France; 3) Department of Structural and Chemical Biology, Mount Sinai School of Medicine, New York, NY, USA; 4) Molecular Genetics and Development, Prince Henry's Institute of Medical Research, Monash Medical Centre, Clayton, Australia; 5) Endocrinologie Moléculaire et Maladies Rares, Centre de Biologie et Pathologie Est, Hospices Civils de Lyon, Bron, France; 6) Centre de Génétique, Hôpital d'Enfants, Dijon, France; 7) Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway; 8) Endocrinologie Pédiatrique, HFME, Hospices Civils de Lyon, Bron, France; 9) Cytogénétique, Centre de Biologie et Pathologie Est, Hospices Civils de Lyon, Bron, France; 10) Cytogénétique, Plateforme de Biologie, CHU Dijon; 11) Pédiatrie, Hôpital d'Enfants, CHU Dijon; 12) The Gade Institute, University of Bergen, Bergen, Norway; 13) Department of Pathology, Haukeland University Hospital, Bergen, Norway; 14) AP-HP, Département de Génétique, Hôpital Necker-Enfants Malades, Paris, France; 15) Department of Clinical Medicine, University of Bergen, Bergen, Norway.

The early gonad is bipotential and can differentiate into either testis or ovary. In XY embryos, the SRY gene located on the Y chromosome triggers the testicular differentiation and a subsequent male development via its action on a single gene, *SOX9*. However, during development, *SOX9* is expressed in a broad variety of tissues, a diversity mostly achieved by means of a complex regulatory region extending at least 1.5 Mb upstream and downstream of its coding sequence and harbouring tissue-specific transcriptional regulatory elements. The mapping of deletions and duplications in patients with isolated disorders of sex development (DSD; one family with a 46,XY male-to-female and three families with 46,XX female-to-male sex reversals) identified a 78 kb critical sex-determining region located in a gene desert ~500 kb upstream of the SOX9 promoter. This data supports the long-distance cis-ruption model at the SOX9 locus as the genetic basis of a significant proportion of isolated 46,XX and 46,XY DSD. Moreover our data point towards the existence of the sex-determining gonad-specific SOX9 enhancer(s) among highly conserved elements comprised in the 78 kb region critical for SOX9 gonadal expression and testicular development whose activity would be super-regulated by an activating (“male”) or repressing (“female”) chromatin environment. We thus propose that cis-ruptions of long distance gonad specific *SOX9* transcriptional enhancer(s) can result in genomic imbalance sufficient to activate or inactivate SOX9 gonadal expression, switch sex determination, and result in DSD in humans.

33

Mutations in ANKRD11 cause KBG syndrome, a syndrome of intellectual disability, skeletal malformations and macrodontia. *M. Tekin[1,2], A. Sirmaci[1,2], M. Spiliopoulos[1], F. Brancati[3,4,5], E. Powell[1,2], D. Duman[6], A. Abrams[1,2], G. Bademci[1,2], E. Agolini[3], S. Guo[1,2], B. Konuk[6], A. Kavaz[6], S. Blanton[1,2], M.C. Digilio[7], B. Dallapiccola[7], J. Young[1,2], S. Zuchner[1,2].* 1) Dept Human Genetics, University of Miami, Miami, FL; 2) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 3) IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Mendel Laboratory, Rome, Italy; 4) Department of Biopathology and Diagnostic Imaging, Tor Vergata University, Rome, Italy; 5) Department of Biomedical Sciences, d'Annunzio University, Chieti, Italy; 6) Division of Pediatric Genetics, Ankara University School of Medicine, Ankara, Turkey; 7) IRCCS Bambino Gesù Children Hospital, Rome, Italy.

KBG syndrome is characterized by intellectual disability associated with macrodontia of the upper central incisors as well as distinct craniofacial findings, short stature and skeletal anomalies. While believed to be genetic in origin, the specific underlying defect is unknown. Through whole exome sequencing we identified deleterious heterozygous mutations in ANKRD11, encoding ankyrin repeat domain 11 also known as ankyrin repeat-containing cofactor 1. A splice site mutation, c.7570-1G>C, co-segregated with the disease in a family with three affected members, while in a sporadic patient a de novo truncating mutation, c.2305delT (p.Ser769GlnfsX8), was detected. Sanger sequencing revealed additional de novo truncating ANKRD11 mutations in three other sporadic cases. ANKRD11 is known to interact with nuclear receptor complexes to modify transcriptional activation. Our results demonstrate that mutations in ANKRD11 cause KBG syndrome and outline a fundamental role of ANKRD11 in craniofacial, dental, skeletal and central nervous system development and function.

34

De novo nonsense mutations of ASXL1 cause Bohring-Opitz (Oberklaid-Danks) syndrome. *A. Hoischen[1], B.W.M. van Bon[1], B. Rodríguez-Santiago[1, 2], C. Gilissen[1], L.E.L.M. Vissers[1], P. de Vries[1], I. Janssen[1], B. van Lier[1], R. Hastings[3], S.F. Smithson[3], R. Newbury-Ecob[3], S. Kjaergaard[4], J. Goodship[5], R. McGowan[6], D. Bartholdi[7], A. Rauch[7], M. Peippo[8], J.M. Cobben[9], D. Wieczorek[10], G. Gillessen-Kaesbach[11], J.A. Veltman[1], H.G. Brunner[1], B.B.B.A. de Vries[1].* 1) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 2) Unitat de Genètica, Universitat Pompeu Fabra, and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Barcelona, Spain; 3) Department of Clinical Genetics, St Michael's Hospital, Bristol, UK; 4) Deptartment of Clinical Genetics, University Hospital of Copenhagen, Rigshospitalet, Denmark; 5) Institute of Human Genetics, Newcastle University, Newcastle Upon Tyne, UK; 6) Department of Medical Genetics, Yorkhill Hospitals, Yorkhill, Glasgow, UK; 7) Institute of Medical Genetics, University of Zurich, Schwerzenbach, Switzerland; 8) Department of Medical Genetics, The Family Federation of Finland, Helsinki, Finland; 9) Department of Pediatric Genetics, Emma Kinderziekenhuis AMC,Amsterdam, The Netherlands; 10) Institut für Humangenetik, Universitätsklinikum Essen, Essen, Germany; 11) Institut für Humangenetik Lübeck, Universität zu Lübeck, Lübeck, Germany.

Bohring syndrome is characterized by severe intellectual disability, distinctive facial features and multiple congenital malformations. Whole exome sequencing and follow-up by Sanger sequencing revealed heterozygous de novo nonsense mutations in the additional sex combs like 1 gene (ASXL1) in seven patients with Bohring syndrome. ASXL1 belongs to the Polycomb group (PcG) and trithorax complexes family. This family of genes is required for maintenance of both activation and silencing of Hox genes, which determine segmental identity in the developing embryo. Somatic mutations in ASXL1 have previously been reported with leukemia and myelodysplactic syndrome. Thus, our findings suggest a new link between severe developmental disorders and cancer.

35

Rapid detection of copy number imbalances and balanced translocations in multiple myeloma by translocation array CGH. *S.A. Morton[1], R.A. Schultz[1], M.L. Slovak[1,8], L. McDaniel[2], A. Furrow[1], J-H. Han[3], J. Fink[4], U. Surti[5], N. Berry[6], K. Fagan[6], T.C. Brown[7], V. Cawich[1], C. Valentin[1], S. Minier[1], N.J. Neill[1], S. Byerly[1], T. Sahoo[1,9], L.G. Shaffer[1], B.C. Ballif[1].* 1) Signature Genomic Laboratories, PerkinElmer Inc., Spokane, WA; 2) Sacred Heart Medical Center, Spokane, WA; 3) Dong-A University College of Medicine, Busan, South Korea; 4) Hennepin County Medical Center, Minneapolis, MN; 5) University of Pittsburgh, Magee-Womens Hospital, Pittsburgh, PA; 6) Pathology North (Hunter), John Hunter Hospital, Newcastle, Australia; 7) CSI Laboratories, Alpharetta, GA; 8) Current affiliation: Quest Diagnostics Nichols Institute, Chantilly, VA; 9) Current affiliation: Quest Diagnostics Nichols Institute, San Juan Capistrano, CA.

Multiple myeloma (MM) is a plasma cell malignancy for which identification of cytogenetic abnormalities impacts treatment decisions, but conventional cytogenetic methods have technical limitations. Low plasma cell counts may mask critical aberrations during karyotyping, with low resolution even in successful studies; FISH limits the number of regions interrogated in one assay and only detects abnormalities in targeted regions. To improve detection of clinically significant MM aberrations, we designed a 135K-feature whole-genome oligonucleotide microarray targeting over 1,800 genomic regions associated with cancer. To evaluate array performance and optimize sample type, we ran array CGH (aCGH) on 54 MM samples: 10 CD138-enriched and 44 unenriched with >30% plasma cells. aCGH on enriched samples was more likely to identify all abnormalities seen by standard methods (e.g. 13q and 17p deletions and hyper/hypodiploidy) versus more limited success with unenriched samples. Variability of relative log ratio shift within abnormalities in the same samples presented clearer evidence of clonality in enriched than unenriched samples. Both sample types revealed many abnormalities not detected by conventional methods (e.g., 6q, 8p, and 14q deletions and chromosome 1 aberrations). We also developed a microarray-based balanced translocation assay (tCGH) that couples linear DNA amplification with aCGH to detect six prognostically-significant MM-associated balanced IgH translocations (*CCND1*, *CCND3*, *MAF*, *FGFR3*, *IRF4*, and *MYC*) and validated the assay's ability to detect translocations using selected enriched samples and cell lines with known translocations. Our results show that tCGH can identify translocations relevant to MM, provide more precise resolution of breakpoints (within a few hundred base pairs) and identify cryptic submicroscopic imbalances at translocation breakpoints. tCGH also has the potential to detect translocations involving more rare partners to the targeted genes (e.g., *MYC*/*IGL*), thus expanding the number of potential translocations detectable. tCGH is anticipated to improve testing for MM patients by targeting multiple translocations in a single assay versus the common tiered FISH approach. The data demonstrate that aCGH and tCGH on CD138-enriched samples provides consistent results relative to standard methods and facilitates the detection of submicroscopic, clinically significant copy changes and translocations with substantially improved resolution.

36

Development and validation of a CGH microarray for clinical diagnosis of hematological malignancies. *S.A Yatsenko[1,2], S.M. Gollin[2,3], J. Hu[1], M. Sathanoori[1], U. Surti[2], A. Rajkovic[1].* 1) OB/GYN & Reproductive Sci, Univ Pittsburgh, Magee-Womens Hospital of the UPMC; 2) Department of Pathology; 3) Department of Human Genetics, Univ Pittsburgh, Graduate School of Public Health, Pittsburgh, PA.

Genome instability is a hallmark of cancer. Accumulation of genetic alterations in cells is associated with loss of tumor suppressor genes and the improper activation of oncogenes, which trigger uncontrolled cellular proliferation and the development of malignant cells. Most mutations are large genomic alterations (deletions, amplifications, and translocations) that contribute to cancer pathogenesis, progression, and have prognostic significance. We developed a custom genome-wide microarray to assist in clinical diagnosis of hematological malignancies. This microarray contains 180,000 oligonucleotide probes to detect chromosome abnormalities, including gains and losses, with an average 150 kb genomic resolution and an enhanced 5-10 kb resolution for 900 genes involved in carcinogenesis. This custom-designed microarray has been validated for clinical use by analyzing 20 bone marrow specimens of pediatric and adult leukemias and lymphomas. All specimens had a total of 49 aberrations, both numerical (n = 35) and structural (n = 14), previously characterized by classical cytogenetic analysis and/or fluorescence *in situ* hybridization (FISH) studies. We identified 67 acquired copy number changes of diagnostic or prognostic significance, 18 changes of unclear clinical significance involving oncogenes or tumor suppressor genes, and 23 benign copy number variants. In eleven specimens (55%), microarray analysis revealed additional pathogenic aberrations that were not seen by karyotype or FISH studies. These included heterozygous and homozygous deletions in genes associated with T- or B-cell malignancies, such as *PAX5*, *CDKN2A*, *CDKN2B*, *MLL*, and *ETV6*. Examination of genomic integrity at translocation or inversion breakpoints in 14 apparently balanced structural rearrangements revealed cryptic deletions in 4/14 (28%) rearrangements including recurrent t(9;22), inv(16), and t(15;17). Microarray analysis is more sensitive than FISH analysis in diagnosis of multiple myelomas, lymphomas, and myelodysplastic syndromes. This custom genome-wide microarray has been implemented for clinical testing of hematological malignancies, and additional studies are underway. The application of array CGH for clinical diagnosis offers significant advantages in studying patient specimens beyond the resolution of classical G-banding chromosome studies and FISH analyses, which may be hampered by “normal” karyotypes, poor chromosome morphology, and/or *in vitro* growth failure.

37

Identification of a novel palindrome mediated translocation associated with the t(3;8) of hereditary renal cancer. *T. Kato[1], M.B. Sheridan[1], A.M. Hacker[1], H. Inagaki[2], T.W Glover[3], S.E Plon[4], H.A Drabkin[5], R.M Gemmill[5], H. Kurahashi[2], B.S. Emanuel[1].* 1) Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Molecular Genetics, ICMS, Fujita Health University, Toyoake, Japan; 3) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 5) Division of Hematology-Oncology, Medical University of South Carolina, Charleston, SC.

It has emerged that palindrome-mediated genomic instability generates genomic rearrangements. Previously, palindrome-mediated recurrent translocations of 22q11 including the constitutional t(11;22), t(8;22) and t(17;22) have been reported. The presence of palindromic AT-rich repeats (PATRRs) at 22q11.2, as well as within the 11q23, 17q11 and 8q24 regions, suggested a palindrome-mediated mechanism involved in the generation of recurrent constitutional 22q11.2 translocations. To date, all reported PATRR mediated translocations have involved PATRR22. However, in this study we identified a new PATRR-mediated translocation in two unrelated families which does not involve PATRR22, the t(3;8)(p14.2;q24.1). It has been associated with hereditary renal cell carcinoma. Detailed sequence analysis of the t(3;8) junction shows that the chromosome 8 breakpoint lies in the first intron of the *RNF139* gene, which is the location of the breakpoint of the t(8;22). Based on sequence analysis we found that the chromosome 3 breakpoint is encompassed by an AT-rich palindromic sequence in intron 3 of *FHIT* (PATRR3). Sequencing the region surrounding the PATRR3 in DNA from normal individuals revealed VNTRs and (AT)n repeats. The region varies in size (from 624 to 2030bp) due to variation in the number of VNTRs and length of the (AT)n repeats. We have previously demonstrated *de novo* t(11;22)s and t(8;22)s in sperm using translocation-specific PCR. Applying this method, we attempted to detect PATRR3-involved *de novo* translocations. Although the t(3;8) was never observed, the t(3;22) and t(3;11) were seen in individuals carrying a symmetrical PATRR3. Thus, palindrome mediated translocation appears to be common and not unique to chromosome 22. An analysis of the PATRR3 in non-human primates revealed that anthropoid PATRR3 sequences are short, similar to humans. Old world monkeys do not have any AT-rich sequences and inverted repeat sequences were not found in primates. Thus, the PATRR3 region appears to have elongated with inverted duplication during primate evolution. This suggests that in the future, new PATRRs might arise and existing PATRRs may be altered as genomic evolution continues.

38

The ETS gene family members: relative positioning in normal human epithelial prostate cells and induction of nuclear reorganization. *I. Tereshchenko[1], A. Vazquez[2], N. Kane-Goldsmith[1], D. Dvorzhinski[2], S. Huhn[1], S. Serrano[1], R. DiPaola[2], J. Tischfield[1].* 1) Department of Genetics, Rutgers University, Piscataway, NJ; 2) Cancer Institute of New Jersey, New Brunswick, NJ.

Spatial organization of the interphase nucleus has an integral role in genome function. The relative proximity of specific loci in interphase nuclei has been established as a key factor in oncogenic rearrangements which lead to hematologic and mesenchymal malignancies. Recurrent chromosomal translocations involving the AR-regulated (androgen receptor-regulated) TMPRSS2 gene and members of the ETS transcription factor family are a distinguishing feature of prostate cancer. However, little is known about the mechanism of these genetic rearrangements in intact prostate cells. Using multicolor 3D fluorescence in situ hybridization (3D FISH), confocal microscopy and computational three-dimensional reconstruction, we determined the relative position of four members of the ETS gene family (ERG, ETV1, ETV4 and ETV5) and the TMPRSS2 gene in interphase nuclei of normal human primary prostate epithelial cells (PrECs, Lonza). Each gene was represented by two appropriate BAC clones, providing approximately 400 kb of genome coverage. According to our q-PCR analysis, a molecular profile of PrECs exhibits a combination of basal epithelial markers and low levels of AR. The three-dimensional positions of the ETS genes were non-random in interphase nuclei of the PrECs. At least one ETV4 locus was positioned in proximity to the ETV5 and ERG genes in 8% of nuclei ($p<0.00001$). The ERG and TMPRSS2 genes, which are located on NOR-bearing chromosome 21 and separated by ~3 Mb, were detected in close proximity (≤0.6 μm) in 27% of PrECs. Moreover, in 10% of the cells, two homologous TMPRSS2 loci were non-randomly positioned in close proximity to each other. Androgen stimulation (via application of charcoal-stripped serum containing media for 48 hr followed by treatment with 10^{-7}M dihydrotestosterone for 1hr) induced nuclear reorganization in PrECs that included the repositioning of the ETS genes toward the nuclear periphery and separation of the homologous chromosome 21 pair. In addition, the treatment was probably related with the reduction of TMPRSS2 and ERG loci pairing. Nuclear reorganization in basal PrECs followed by the treatment may help to explain ineffective attempts to induce specific chromosomal translocation in the cells. However, the spatial nuclear organization of epithelial prostate cells may be a principal contributor to recurrent gene fusion involving TMRSS2 and members of the ETS gene family.

39

Trisomy 21: a disease of chromatin architecture? *A. Letourneau[1], S.B. Montgomery[1], D. Gonzalez[2], D. Robyr[1], C. Borel[1], E. Migliavacca[1,3], Y. Hibaoui[4], L. Farinelli[5], M. Gagnebin[1], E. Falconnet[1], S. Deutsch[1], S. Dahoun-Hadorn[6], JL. Blouin[6], A. Feki[4], R. Guigo[2], E.T. Dermitzakis[1], S.E. Antonarakis[1].* 1) Genetic Medicine & Development, University of Geneva Medical School, Geneva, Switzerland; 2) Center for Genomic Regulation, University Pompeu Fabra, Barcelona, Spain; 3) Swiss Institute of Bioinfomatics, Switzerland; 4) Stem Cell Research Laboratory, University Hospitals of Geneva, Geneva, Switzerland; 5) FASTERIS SA, Plan-les-Ouates, Switzerland; 6) University Hospitals of Geneva, Geneva, Switzerland.

Down syndrome (DS) results from trisomy of human chromosome 21 (T21). It is the most frequent live-born aneuploidy and a leading cause of mental retardation. It is likely that the majority of the phenotypes are related to alterations of gene expression. In order to assess the perturbations of gene expression in T21 tissues, we studied the transcriptome of fibroblasts derived from a pair of monozygotic twins discordant for T21. The use of these samples eliminates the bias of genome variability and most of the differences observed will likely be related to the supernumerary chromosome 21. The transcriptome was studied by mRNA-Seq; more than 200 million 75bp paired-end reads were generated and mapped using GEM (Ribeca et al, unpublished). The expression level of all protein-coding genes was compared between the discordant twins. Surprisingly, for most of the chromosomes, we identified large chromosomal regions that are either up- or downregulated in the trisomic twin. This observation suggests that differential expression is organized in domains along the chromosomes. The same domains of gene expression dysregulation have been found when we compared the transcriptomes of induced pluripotent stem (iPS) cells derived from the twins fibroblasts. This indicates that iPS cells retain a specific memory of their origin. Control comparisons of normal monozygotic twins as well as 20 DS and normal unrelated individuals were also performed. Interestingly, the position of those chromosomal domains significantly correlates with the lamina-associated domains (LADs) described by Guelen et al (2008). Those genomic regions associated with the nuclear lamina are known to contain mostly repressed genes. Our data suggest that the genes located within those LADs may be de-repressed in trisomic cells. Mapping of the LADs in the twins' fibroblasts will help to further explore this hypothesis. In addition, we identified genes that differentially utilize alternative spliced isoforms between the trisomic and normal samples and genes that show differential allele specific expression. Our study reveals a unique finding in genome regulation of T21 nuclei and may be important for the molecular pathogenesis of whole and partial chromosome aneuploidies.

40

A model for the formation of 11q segmental amplifications or hsr based on structure defined by aCGH in leukemia. *K. Reddy.* Kaiser Permanente Southern California, CA.

Segmental amplification of chromosome 11q is observed in AML and is considered to have a poor prognosis. Two intra chromosomal 11q amplifications were examined using cytogenetics, FISH with MLL probe and 120K oligoarray CGH. The first case had 4 similar G-band segments with MLL signals on a dicentric chromosome, dic(11;13). By aCGH the segments spanned 11q21q25 (95612033-134432465 bp) of ~ 27.9 Mb, with a 0.8 Mb loss proximal to MLL. Inter segment composition was dissimilar, all 4 segments involved the MLL gene region and 2 segments also involved regions proximal and distal to MLL gene. The second case had apparent recurring G-band pattern. By FISH, a paracentric inversion or insertion of 11q23 MLL to proximal 11q11.2 band, an in situ deletion of 3'MLL signal, followed by additional 3 equally spaced MLL signals and two terminal inversion duplication or triplication of MLL segment, was observed. In total, the amplified 11q included ~8-10 complete and 1 partial MLL signal. In aCGH, the segments spanned 11q14.1q25 (78109652-134432465 bp), involving 5 fragments ranging from 1.3 to 12.4 Mb totaling 29.7 Mb deleted, intervening fragments totaling to 28.9 Mb were amplified and fused in various combinations. Only some fragments were in all segments and others were in some segments. Overall, the putative cancer gene pool involved in the two cases of 11q segmental amplifications is considerably different. Episome model for hsr formation, a chromosome segment is deleted, circularized and externally amplified. We have shown in both our cases one or more fragments within a segment were deleted probably prior to amplification. Amplification by rolling circle replication reeling out multiple segments, with the fragments excluded in some segments looped out from the replicating circle is invoked for its efficiency. The amplified segment was integrated into chromosomes. Post integration change in fragments is suggested in aCGH. Understanding the cause and effect for differences in the number of amplicons may have clinical implications. The occasional metaphases with additional markers and a chromosome 11 with shorter segmental amplifications were probably the breakage fusion bridge products of mitosis.

41

Mechanism of recurrent translocation t(11;22) initiated by cruciform conformation of palindromic sequences. *H. Inagaki[1], T. Ohye[1], H. Kogo[1], M. Tsutsumi[1], T. Kato[1,2], M. Tong[1], B.S. Emanuel[2], H. Kurahashi[1].* 1) Div Molecular Genetics, ICMS, Fujita Health Univ, Toyoake, Aichi, Japan; 2) Div Human Genetics, The Children's Hospital of Philadelphia, and Dept Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA.

The t(11;22)(q23;q11) is one of the most frequent recurrent constitutional translocations in humans. The breakpoints have been located within characteristic sequences, palindromic AT-rich repeats (PATRRs) on both chromosomes. Previously, we established a plasmid-based model system for the PATRR-mediated translocation using a human cell line as a host. Using this system, we demonstrated that cruciform formation of the PATRR is essential for induction of double-strand-breaks (DSBs) that lead to the translocation-like rearrangement. Further, we found key nucleases, GEN1 and Artemis, participate in this reaction: Holliday junction (HJ) resolvase GEN1 resolves the cruciform structure into closed hairpin DNA ends, and Artemis opens the hairpins into DSB ends, which are joined via illegitimate nonhomologous end joining. Recently, a novel HJ resolvase SLX4-SLX1 complex was identified in addition to the MUS81-EME1 complex and GEN1, and this has raised a possibility for redundancy of the resolution pathway. In this study, we examined the levels of the first cleavage of the PATRR cruciform by double knockdown of the three HJ resolvases to determine whether these resolvases work coordinately or independently. The result of the knockdown of the SLX1 and MUS81 combination did not show any effect. The combination of a GEN1 and SLX1, or GEN1 and MUS81 knockdown, on the other hand, showed decreased cleavage equal to that of a single knockdown of GEN1. The results indicate that only GEN1 acts independently in the first cleavage of the PATRR cruciform. Our data suggest that, although GEN1 fulfills overlapping and backup roles with other resolvases to maintain genomic integrity through the homologous recombination pathway, GEN1 acts as a double-edged sword and also participates in the production of recurrent chromosomal translocations in humans.

42

Ectopic synapsis as a mediator of ectopic crossing over: inferences from genomic disorders. *P. Liu[1], M. Lacaria[1], F. Zhang[1,4], M. Withers[1], P.J. Hastings[1], J.R. Lupski[1,2,3].* 1) Molec & Human Gen, Baylor College Med, Houston, TX, USA; 2) Dept of Pediatrics, Baylor College of Med, Houston, TX, USA; 3) Texas Children's Hospital, Houston, TX, USA; 4) MOE Key Laboratory of Contemporary Anthropology, School of Life Science, Fudan Univ, Shanghai, China.

Genomic disorders constitute a class of diseases that are associated with DNA rearrangements resulting from region-specific genome instability; i.e. genome architecture incites genome instability. Non-allelic homologous recombination (NAHR) or ectopic recombination in human meiosis can result in recurrent deletions or duplications causing genomic disorders. Previous studies of NAHR have focused on description of the phenomenon, but it remains unclear how NAHR occurs during meiosis and what factors determine its frequency. Here we assembled two patient cohorts with reciprocal genomic disorders; deletion associated Smith-Magenis syndrome (SMS [MIM 182290]) and duplication associated Potocki-Lupski syndrome (PTLS [MIM 610883]). By assessing the full spectrum of rearrangement types from the two cohorts, we find that complex rearrangements (those with more than one breakpoint) are more prevalent in copy number gains (17.7%) than in copy number losses (2.3%); an observation that supports a role for replicative mechanisms in complex rearrangement formation. Interestingly, for NAHR mediated recurrent rearrangements, we show that crossover frequency is positively associated with the flanking low-copy repeat (LCR) length and inversely influenced by the inter-LCR distance. To explain this, we propose that the probability of ectopic chromosome synapsis increases with increased LCR length, and that this is a necessary precursor to ectopic crossing-over.

43

Intragenic copy number changes detected by an exon-targeted array in patients with neurodevelopmental disorders. *S.W. Cheung[1, 2], C. Shaw[1, 2], S-H.L. Kang[1, 2], J. Pham[2], A. Ester[2], P. Luke[4], P.M. Hixson[2], A.N. Pursley[1, 2], P.M. Boone[1], C.A. Bacino[1, 2, 3, 4], S. Lalani[1, 2, 3, 4], F. Probst[1, 3, 4], W. Bi[1, 2], A.L. Beaudet[1, 2, 3, 4], J.R. Lupski[1, 2, 3, 4], A. Patel[1, 2], P. Stankiewicz[1, 2, 5].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Medical Genetics Laboratories, Baylor College of Medicine, Houston, TX; 3) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 4) Texas Children's Hospital, Houston, TX; 5) Department of Medical Genetics, Institute of Mother and Child, Warsaw, Poland.

It has been suggested that rare copy number variations (CNVs) containing genes involved in neurotransmission or synapse formation and maintenance are associated with autism spectrum disorders (ASD) and intellectual disability (ID) and that disruption of these shared biological pathways can result in neurodevelopmental disorders (PMID 20421931; 19736351). Recently, we demonstrated the utility of a custom-designed exon-targeted oligo array to detect intragenic copy-number changes in patients with various clinical phenotypes (PMID 20848651). 10,362 patients have been evaluated by this array with indications provided for 8,898 cases. 5,138 patients studied have a clinical indication of cognitive impairment (primarily developmental delay, mental retardation and autism). From a list of 255 disease or candidate genes thought to play a role in ASD and ID (PMID 20531469), our exon-targeted clinical array has coverage for 187 (73%) of these genes with an average of 81 oligonucleotide probes per gene. We identified 243 intragenic CNVs encompassing or disrupting 60 of these genes. Examples of exonic CNVs occurring within a gene known to be causative for the observed clinical phenotype include: *MECP2* and Rett syndrome; *CREBBP* and *EP300* and Rubinstein-Taybi syndrome; *NF1* and neurofibromatosis type 1; and *ZDHHC9* and X-linked mental retardation. The more commonly observed CNVs detected in this cohort include those affecting known and candidate genes for ASD such as *NRXN1* (38 cases), *CNTNAP2* (5), *NLGN4X* (4), *A2BP1* (22), *CNTN4* (11), *CDH18* (13), and *TMEM195* (8); and genes associated with cognitive impairment such as *TSPAN7* (13), *IL1RAPL1* (5), and *PAK3* (3). Detection of recurring intragenic CNVs in this patient population will support reclassifying candidate genes implicated in neurodevelopmental disorders as disease-causing genes. For example, of the six cases with exonic CNVs in the candidate gene, *AUTS2*, indications were provided for five cases, all of which include developmental delay. Additionally, parental studies performed to date have revealed no evidence of the CNV in a parent (3 cases are de novo, 1 case is not present in the mother). In summary, our approach allows enhanced detection of known genetic conditions and has the potential to elucidate new disease genes. We demonstrate the utility of a custom-designed exon-targeted DNA microarray to detect small CNVs thus improving the diagnostic capability in patients with ASD and ID.

44

Genomewide testing the most frequent genetic diseases optimizes abnormal gene identity and test reliability. *R.V. Lebo[1], W.W. Grody[2].* 1) Dept Pathology, Akron Children's Hosp, Northeast Ohio College of Medicine; 2) Dept Pathology, University of California, Los Angeles.

Targeted genomewide testing for the most frequent genetic disease targets can provide medically important, unambiguous results that define the etiology of a substantial proportion of disease phenotypes. Ranking over 500 human genetic disease frequencies listed in GeneTests found 146 genetic diseases with a frequency exceeding 1 in 100,000 in the general population that together affect 2% (1 in 49) of individuals worldwide. Testing 40 autosomal recessive disease genes in the U.S. reproductive population would detect 1 disease mutation in 53% of tested U.S. patients, compared to 1 mutation in 3% of Caucasian cystic fibrosis patients as in current practice. When the identified carrier's partner is tested for the same mutant gene, 2% (1 in 51) couples will both carry the same mutated gene and can be alerted to their 1-in-4 risk of an affected fetus. By similar calculations, 1 in 203 reproductive patients are carriers for 1 of 24 X-linked genetic diseases,1 in 84 patients are affected with 1 of 72 autosomal dominant diseases, 1 in 444 patients are affected with 1 of 9 listed common gene region deletions, and 1 in 8500 individuals are affected with syndromic mitochondrial mutations at 1 of 12 mtDNA sites. These high-yield disease target lists can be adapted to test any of the following patient categories: pregnant or prepregnant couples, products of conception, prenatal samples, newborns, affected patients of any age, and asymptomatic, presymptomatic or symptomatic patients seeking medical advice. Appropriate targets include de novo genetic diseases in newborns, late-onset genetic diseases in adults, and testing pregnant couples for newborn diseases requiring immediate treatment. Impressive advances in testing platforms have enabled single nucleotide mutation analysis by rapid sequencers and by microarrays. Excluding the 20% of balanced fetal chromosomal abnormalities detected by karyotyping, nearly all of the remaining 80% of abnormalities that are unbalanced could be detected by testing 37 polymorphic sites molecularly. Developing platforms for each category can continue to optimize future patient care by incorporating additional recurrent disease gene mutations from single nucleotide changes to multigene deletions and duplications. Developed disease gene panels would be used to test patients with nonspecific genetic disease symptoms, while those with recognized symptomatology can be tested immediately for the specific relevant gene(s).

45

Bringing current research technology to the clinic: The Manton Center for Orphan Disease Research Gene Discovery Core. *M.C. Connolly, J. Picker, I.A. Holm, A.H. Beggs, P.B. Agrawal.* Division of Genetics and Program in Genomics, The Manton Center for Orphan Disease Research, Children's Hospital Boston, Harvard Medical School, Boston, MA, United States.

The Manton Center for Orphan Disease Research is an interdisciplinary program established to provide the infrastructure needed to promote gene discovery in patients with orphan diseases, thereby leading to new diagnostic and therapeutic approaches against rare diseases. The Center's state-of-the-art “Gene Discovery Core” (GDC) provides resources to investigators, including an IRB-approved protocol for patient/family enrollment, genetic studies, expertise, advice and logistical support and access to the latest research technologies, including Next Generation Sequencing (NGS). The GDC links phenotypic data (i.e., medical and family history) to genomic data and biological samples (blood, tissues, etc.). Investigators can access samples and data for further analysis, and The Manton Center provides funding to support genetic studies on a competitive basis through internal grants. To date, investigators from 10 Divisions/Departments within Children's Hospital Boston have enrolled patients in the GDC, including the Divisions of Genetics, Gastroenterology and Newborn Medicine, and the Departments of Neurology, Urology and Cardiology. In total, 173 individuals, including those with rare or unknown disorders and their family members, have enrolled. Several individuals with strong evidence for a genetic disease, but without an identified genetic cause despite extensive clinical testing, have undergone NGS to identify the causative genetic alterations. In one notable example, a patient with ROHHAD syndrome, a newly described and rare disorder characterized by Rapid onset of Obesity, Hypothalamic dysfunction, Hypoventilation and Autonomic Dysregulation, was identified in clinic and referred to the study. Working with the family, the GDC staff identified 11 additional families with ROHHAD via online patient networks. Six of these families are currently enrolled and several are now undergoing NGS to determine the genetic basis for this condition. Importantly for the participants, the GDC has a mechanism to report research results back to the participants through The Center's genetic counselor and a health care provider of the participant's choosing. The Manton Center GDC is a novel model that brings state-of-the-art research technologies to patients with rare diseases. The GDC reduces obstacles between the bedside and bench thereby helping to identify new disease conditions, etiologies and potential therapeutic targets.

46

Genetic diagnostics using next generation sequencing: The CEO Genome Project. *C.T. Caskey, M.L. Gonzalez-Garay.* University of Texas, Medical School, Institute of Molecular Medicine, Houston, TX.

Replacing traditional methods of genetic testing for heritable disorders with next generation sequencing (NGS) will reduce the cost of analysis and increase the knowledge available to patients. NGS will become an invaluable resource for patients and physicians especially if the patient sequences and analyses are stored properly, enabling future re-analysis and analysis comparisons as bioinformatics tools and annotations improve. On the other hand, high-throughput sequencing is continually improving, but it is currently full of false positive and negative calls and requires a complex infrastructure and specialized personnel to properly analyze the sequences. To develop firsthand experience with NGS as a diagnostic tool, last October 2010, we organized a local event directed to educate influential members of our community about Genetics and NGS. During the process we recruited 80 non-related individuals from the Young Presidents' Organization (YPO) Houston Chapter and 10 non-related volunteers to participate in our project under an approved IRB. We sequenced all of the volunteers using whole exome sequencing (WES). In addition, the 10 non-YPO volunteers were also sequenced using an orthogonal technology (Complete Genomics Inc.) to validate variants. After bioinformatics analysis we found that each volunteer had over 10,000 non-synonymous coding variants (NSCV) on their exome, with an average of 271 private NSCV. To detect disease causing mutations we annotated the volunteers' NSCV using the human gene mutation database (HGMD) and the GET-Evidence database from the Personal Genome Project. We found a large number of inconsistencies between databases, so we took a very conservative approach and selected only those considered disease causing by HGMD with a low frequency and with high sequencing coverage and mapping values. Under our stringent conditions we identified a dozen variants per volunteer of mostly carrier status but in some cases we identified variants for serious medical conditions.

47

An internet-based approach to enhance genetic data discovery in ALS. *C. Brownstein, T. Vaughan, P. Wicks.* Research & Development, PatientsLikeMe, Cambridge, MA.

Purpose: Patients with familial ALS (FALS) have a strong need to understand their illness and the impact of genetics on their outcomes. To meet this need, a novel disease-focused social network (PatientsLikeMe.com) was used to investigate the variability in clinical phenotype and disease progression due to different ALS-causing mutations, with the aim of sharing this information with patients. Methods: Genetic data-capture capabilities were added to PatientsLikeMe.com, and all active members with ALS were asked to submit their genetic mutation, if known. A search engine was launched on the site so patients could identify others with the same mutation. Clinical features of the mutation-carrying cases were summarized and compared with the published literature. Results: Fifty-two percent of Familial ALS (FALS, N=164) patients responded to a request to tell PatientsLikeMe any genetic information they had on their disease, including “no mutations found”. Twenty-three community members with 14 different mutations submitted their genetic mutations and self-report clinical outcome scores. The most frequently reported mutations were SOD1 A4V (N=9), SOD1 D90A (N=5), and the VAPB mutation P56S (N=5). Preliminary analysis indicate that D90A patients lost 0.24 points per month on the self-report Amyotrophic Lateral Sclerosis Functional Rating Scale, Revised (ALSFRS-R)(SE 0.08, CI95%:-0.45, -0.03), SOD1 A4V patients lost an average of 1.7 points per month (SE 0.4, CI95%: -2.6, -0.84), and VAPB P56S mutation patients had an average loss of 0.31 points per month (SE 0.12, CI95%: -0.66, 0.03). By comparison, Sporadic ALS patients with no genetic information (N=1342) had a mean loss of 1.04 points per month (SE 0.08, CI95%: -1.20, -0.87). Conclusions: Even with a small sample size, trends rapidly become visible in ALS progression rates due to different mutations. Our hope is that sharing this information to patients who want it will help them plan for and anticipate future disease progression. Using the internet to aggregate groups of patients by specific mutation criteria allows for the international sharing of predictive outcomes based on historical data. This presents the possibility of real time distribution of new information to improve clinical understanding and communication. Making patients the driver of this communication medium means that they become educated and engaged, and can more effectively contribute to research and their own clinical management.

48

National Thalassaemia Registry - An Effective Strategy for Thalassaemia Control. *I. Ng[1,2], G.P. Tan[2], Y.M. Tan[2], H.Y. Law[1,2].* 1) Genetics Service, Dept of Paed, KK Women's & Children's Hospital, Singapore; 2) National Thalassaemia Registry, Singapore.

Introduction: Thalassaemia is the commonest genetic disease in Singapore. The incidence is about 4.45% ((-thalassaemia 3.0%,) -thalassaemia 0.90%, HbE 0.55%). Population screening for thalassaemia has been advocated since 1988. The National Thalassaemia Registry (NTR) was established in 1992 by Ministry of Health to register all individuals with thalassaemia. NTR receives notifications of thalassaemia from different sources such as hospitals, polyclinics, private practitioners, Singapore Armed Forces and self-referred cases. Free thalassaemia counseling and subsidized screening are offered to registered families. Methods: A standard notification form is used by doctors to refer index cases upon obtaining consent. Patients and families are invited to attend counseling sessions and offered thalassaemia screening. Couples identified prospectively to be at- risk for a child with thalassaemia major are then offered the option of prenatal diagnosis. Results: Data was collected from 43471 registrants from Year 1992-2010. This is made up of 72.6% Chinese, 17.6% Malay, 4.8% Indian and 5.1% of other ethnicity, similar to the ethnic composition of Singapore population. NTR received 18482 notifications and screened 24989 individuals. A total of 30394 individuals with thalassaemia were identified. These were (-thalassaemia minor (46.48%),) -thalassaemia minor (38.0%), HbE trait (12.02%), HbH disease (1.71%), concurrent (and) -thalassaemia carriers (1.16%), HbE with) -thalassaemia (0.33%) and) -thalassaemia major (0.29%). With this proactive approach, the number of new patients with thalassaemia major has decreased from about 10-15/year to 0-1/year since 1997. Conclusion: The NTR has been effective in significantly decreasing the incidence of thalassaemia major in Singapore, with decrease in the healthcare burden. Resources can then be used to optimize the care of existing patients with thalassaemia major. Recent challenges include the increasing diversity of the population, evolving public perspectives of genetic diseases, bio-ethical considerations and advances in the management of thalassaemia major patients. Thalassaemia control in the regional countries can leverage on the experience of NTR.

49

Understanding the psychosocial impact of Klinefelter Syndrome. *A.S. Herlihy[1,2,3,4], R.I. McLachlan[2,3,4], L. Gillam[5,6], J.L. Halliday[1,5].* 1) Public Health Genetics, Murdoch Childrens Research Institute, Parkville, Victoria , Australia; 2) Andrology Australia, Clayton, Victoria, Australia; 3) Monash University, Clayton, Victoria, Australia; 4) Clinical Andrology, Prince Henry's Institute of Medical Research, Clayton, Victoria, Australia; 5) The University of Melbourne, Parkville, Victoria, Australia; 6) Children's Bioethics Centre, Royal Childrens Hospital, Flemington Road, Parkville, Victoria, Australia.

Klinefelter Syndrome (KS) is a genetic condition affecting males, resulting in a spectrum of clinical features ranging from infertility, androgen deficiency and gynaecomastia to social, behavioral and learning difficulties. The prevalence is estimated at 1:650, yet up to 70% remain undiagnosed. This is be partly due to the variable phenotype, which may be influenced by genetic factors. Whilst medical aspects of KS have been well-researched, very little exists regarding the personal impact. This study aimed to determine the psychosocial impact of KS, the influence of genetic factors on this, and the attitudes of men with KS towards optimal diagnosis age and acceptability of screening. A population-based sample was recruited from a broad range of sources. 87 participants completed a written questionnaire, 79 provided a saliva sample for genetic analysis and 77 agreed to an interview, of which 37 took place. Questionnaire outcomes were analysed using multivariate analysis. DNA samples were examined for androgen receptor (AR) CAG repeat length and X-inactivation. Interview sessions were audio-taped, transcribed and analysed for content, interactions and themes. The mean age of participants was 43 years (range 19 to 76). Compared to population normative data, there was strong evidence ($p<0.001$) for the KS cohort having poorer outcomes for subjective well-being, body image, self-esteem, and mental health. The strongest predictors of this decreased psychosocial health were phenotypic severity, unemployment and low social support. Although skewed X-inactivation was observed, no association was found between AR CAG repeat length and the psychosocial or biomedical outcomes measured. Seventy-two percent of participants were diagnosed as adults, of which 66% wish they had been diagnosed earlier. Two thirds of all participants supported population screening for KS. Interview findings revealed a wide range of reactions to the diagnosis of KS and a variable personal impact of living with the condition. For those diagnosed as adults, the diagnosis of infertility was devastating, but the additional diagnosis of KS provoked much reflection: “How much is me and how much is the condition?” Often this occurred in the absence of access to resources or support. For men with KS, there is a measurably negative psychosocial impact of having the condition. The potential benefits of early diagnosis through screening should be considered, further support resources developed.

50

The influence of genetic risk information on parental role identity in adolescent girls and young women from families with fragile X syndrome. *A. McConkie-Rosell, E.M. Heise, G.A. Spiridigliozzi.* Pediatrics, Duke University Medical Center, Durham , NC.

Using a multi-group cross-sectional design, we explored self-concept related to parental role salience and enactment in 53 young women (14 to 24 yrs) with knowledge they were either carriers, non-carriers, or could be a carrier of fragile X syndrome (FXS). Parental role salience included the participants' desire “to be a mother” and the importance they placed on this role. Enactment focused on the participants' views regarding ways to become a mother (reproductive options), parenting a child affected by FXS, and the development of partner relationships (marriage). Participants completed the FXS Adolescent Interview and the FX-Visual Analog Scale. Participants' knowledge of their genetic risk status appears to have influenced both salience and enactment of the parental role. The effect varied based on carrier status. While the majority reported wanting a biological child, those who knew they were at-risk had a reduced desire to be a mother ($p = 0.037$) and had less exploration of options related to enactment than those who were carriers or non-carriers. There was general support for adoption and prenatal diagnosis; however, these options were not often viewed as personal choices. Prenatal testing was endorsed if it could prevent a pregnancy with an affected child or would be used to better prepare to parent an affected child. Only 3 participants, no carriers, would consider pregnancy termination of an affected child. Participants in all groups expressed some conflict between terminating a pregnancy because of FXS and positive feelings for their affected relative(s). For many, their personal definition of “mother” includes parenting a child with special needs. The majority acknowledged this possibility, expressed acceptance, were confident they would be successful in parenting a child with FXS, and identified the importance of finding a partner who had positive characteristics to help them meet future challenges. For many knowledge of genetic risk appears to have led to reappraisal, redefinition, and re-engagement with the goal of becoming a parent. This process was prominent in those who are carriers and less so in those who were at-risk and did not typically occur in those who were non-carriers. Findings offer a valuable insight into the impact of genetic risk information on developing perceptions of the parental role and offer new directions for genetic counseling in adolescents and young women with a family history of FXS.

51

Multivariate linkage analysis points to a shared genetic component between mycobacteria-triggered TNF production and innate resistance to tuberculosis infection. *A. Cobat[1], E. Hoal[2], E. Schurr[1], A. Alcais[3]*. 1) McGill Centre for the Study of Host Resistance & Departments of Human Genetics and Medicine, McGill University, Quebec, Canada; 2) Molecular Biology and Human Genetics, MRC Centre for Molecular and Cellular Biology, DST/NRF Centre of Excellence for Biomedical TB Research, Faculty of Health Sciences, Stellenbosch University, Tygerberg, South Africa; 3) Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Santé et de la Recherche Médicale, U.550, Paris 75015, France, EU and Rockefeller Branch, The Rockefeller University, New York, 10065 NY, USA.

Tuberculosis (TB) is a chronic infectious disease caused by Mycobacterium tuberculosis (Mtb) that remains a major public health challenge worldwide, with 9 million new cases and 2 million deaths each year (1). Tumour necrosis factor (TNF) is a key immune regulator of TB resistance as is shown by the highly increased risk of TB disease among individuals receiving TNF-blocker therapy (2). Heritability of TNF production triggered by mycobacterial antigens has been estimated at 30% to 60% (3, 4). Moreover, major gene control of Mtb-triggered TNF production has been detected by complex segregation analysis of tuberculosis cases and their household contacts in Uganda (5). Taken together, these data suggested mycobacteria-triggered TNF production as an attractive phenotype for detailed molecular genetics follow-up studies. We determined the extent of TNF production in 134 nuclear families from an area hyperendemic for TB in South Africa. In an in-vitro assay we stimulated whole blood obtained from children with Mycobacterium bovis -Bacille Calmette Guerin (BCG) or with BCG plus interferon gamma (IFN-.). We performed a bivariate genomewide linkage scan of TNF production after stimulation by BCG alone and BCG plus IFN-. . Linkage analysis was conducted using the new MLB-QTL method (6) extended to multivariate analysis, following an approach proposed by Mangin et al (7). We detected a major locus (p < 10-5) on chromosomal region 11p15 that controls TNF production after stimulation by both BCG alone and BCG plus IFN-. . Interestingly, this major locus overlaps the TST1 locus, previously identified in the same family sample, that controls T-cell-independent resistance to Mtb infection (8). Our results suggest a previously unknown link of genetically controlled TNF production and resistance to infection with Mtb. 1. WHO report 2009. 2. Gardam et al. 2003. Lancet Infect Dis 3:148-155. 3. Cobat et al. 2010. J Infect Dis 201:15-19. 4. Stein et al. 2003. J Infect Dis 187:1679-1685. 5. Stein et al. 2005. Hum Hered 60:109-118. 6. Cobat et al. 2011. Genet Epidemiol 35:46-56. 7. Mangin et al. 1998. Biometrics 54:88-99. 8. Cobat et al. 2009. J. Exp. Med. 206:2583-2591.

52

Major histocompatibility complex association to rheumatoid arthritis is explained by polymorphic amino acids in the binding grooves of HLA-DR, HLA-B, and HLA-DP. *P.I.W. de Bakker[1,2,3], E.A. Stahl[1,2], X. Jia[1,2], L. Alfredsson[4], L. Padyukov[4], K.A. Siminovitch[5], L. Klareskog[4], J. Worthington[6], R.M. Plenge[1,2], P.K. Gregersen[7], S. Raychaudhuri[1,2]*. 1) Brigham & Women's Hosp, Boston, MA; 2) Broad Institute of Harvard and MIT, Cambridge, MA; 3) University Medical Center Utrecht, The Netherlands; 4) Karolinska Institutet, Stockholm, Sweden; 5) University of Toronto, Mount Sinai Hospital and University Health Network, Toronto, Ontario, Canada; 6) University of Manchester, UK; 7) Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, New York, USA.

Background: Rheumatoid arthritis is a disabling auto-immune disease that affects 1% of the adult population worldwide. While the RA association at the MHC has been noted for 20 years, there has been controversy about associations outside *DRB1* and the causal alleles within *DRB1* itself.

Methods: In order to fine-map the MHC association, we used a previously described method (Pereyra *et al*., Science, 2010) to impute classical types and amino acid polymorphisms for the classical *HLA-A*, *-B*, *-C*, *DPA1*, *DPB1*, *DQA1*, *DQB1*, and *DRB1* loci along with 2,537 SNPs across the MHC in 5,018 anti-CCP+ cases and 14,974 controls from 6 GWAS data sets (Stahl *et al*., Nat. Genet., 2010). Using logistic regression across all 6 data sets, we performed conditional haplotype analyses to identify independent variants, correcting for 5 principal components in each GWAS.

Results: We compared imputed classical alleles to *DRB1* 4-digit genotypes for 1150 overlapping individuals from two cohorts; the imputation was 96.5% and 85.4% accurate for 2-digit and 4-digit types, respectively. We also observed that classical *DRB1* associations were identical to previously reported effect sizes. The strongest signal mapped to the presence of a valine or leucine in position 11 of DRB1 (OR=3.7, $p<10^{-500}$), accounting for 9% of the phenotypic variance. With conditional haplotype analysis we found that most of the *DRB1* signal can be explained by 3 amino acid positions (11, 71, 74), extending the "shared epitope" hypothesis to position 11, located at the floor of the binding groove. We also confirmed an independent effect due to a single amino acid polymorphism in HLA-B (OR=2.1; $p=2x10^{-38}$, Asp at position 9), corresponding to *HLA-B*08*, and identified a novel independent effect at HLA-DPB1 (OR=1.3; $p=5x10^{-22}$, Phe at position 9). Controlling for all of these variants leaves no additional significant associations within the MHC ($p>10^{-6}$).

Conclusions: In each case, the amino acids at these positions point into the binding groove of the HLA molecule, highlighting the importance of these residues in presenting antigens to the immune system. In aggregate, these independent variants account for almost all of the known MHC signal in RA, and explain 15% of the phenotypic variance, a significant increase relative to the most recent genome-wide meta-analysis in RA.

53

Mulitple signals of association at known loci explain additional phenotypic variation and reveal complex patterns of association. *T.M. Frayling[1], A.R. Wood[1], D.G. Hernandez[2,3], M.A. Nalls[2], H. Yaghootkar[1], J.R. Gibbs[2], L.W. Harries[4], S. Chong[2], M. Moore[2], M.N. Weedon[1], J.M. Guralnik[5], S. Bandinelli[6], A. Murray[1], L. Ferrucci[7], A.B. Singleton[2], D. Melzer[4].* 1) Peninsula Medical Sch, Exeter Univ, Exeter, United Kingdom; 2) Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; 3) Department of Molecular Neuroscience and Reta Lila Laboratories, Institute of Neurology, UCL, London, UK; 4) Institute of Biomedical and Clinical Sciences, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter, UK; 5) Department of Epidemiology and Public Health, University of Maryland School of Medicine, MD, USA; 6) Geriatric Unit, Azienda Sanitaria di Firenze, Florence, Italy; 7) Clinical Research Branch, National Institute on Aging NIA-ASTRA Unit, Harbor Hospital, MD, USA.

Allelic heterogeneity could explain additional phenotypic variance ("missing heritability") of common traits, and help identify causal genes. Despite this, few studies have systematically tested for multiple independent signals at individual loci. We examined gene expression levels as a potential model trait because of the large number of strong genetic effects acting in cis. Using blood based expression profiles from 613 individuals, we performed genome-wide SNP analyses to identify cis-eQTLs, and conditional analysis to identify second signals. We examined patterns of association when accounting for multiple SNPs at a locus and when including additional SNPs from the 1000 Genomes Project. We identified 1298 cis eQTLs at FDR~0.01, of which 118 (9%) showed evidence of a second independent signal. Accounting for two signals resulted in an average 31% increase in phenotypic variance explained ($p<0.0001$) at these 118 loci. At some loci, apparent second signals could be attributed to LD as weak as $r2 = 0.03$ with a primary signal. At other loci, evidence for a second independent signal increased despite LD as high as $r2 = 0.26$. The association of SNPs with cis gene expression could increase, stay similar, or decrease in significance when accounting for linkage disequilibrium with second signals at the same locus. Pairs of SNPs increasing in significance tended to have gene expression increasing alleles on opposite haplotypes, whereas pairs of SNPs decreasing in significance tended to have gene expression increasing alleles on the same haplotypes. Adding in data from the 1000 Genomes Project showed that several apparently independent signals could be potentially explained by a single association signal. Our results show that accounting for multiple variants at a locus will increase the variance explained in a substantial fraction of loci, but that allelic heterogeneity will be difficult to define without resequencing loci and functional work.

54

Expression quantitative trait locus analysis in an experimental cohort identifies multiple loci strongly associated with the human response to influenza vaccination. *L. Franco[1,4], K.L. Bucasas[2], J.M. Quarles[6], J.M. Wells[3], N. Arden[6], D. Niño[3], R.B. Couch[3,4], C.A. Shaw[1], J.W. Belmont[1,5].* 1) Molec & Human Gen, Baylor College Med, Houston, TX; 2) Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX; 3) Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX; 4) Department of Medicine, Baylor College of Medicine, Houston, TX; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 6) Department of Microbial and Molecular Pathogenesis, Texas A&M University System Health Sciences Center, College Station, TX.

We report the results of the first human expression quantitative trait study employing an in vivo experimental stimulus. The goal of the study was to identify regions of human DNA that are associated with the transcriptional response to influenza vaccination. We immunized an ethnically homogeneous cohort of 119 healthy male adults with a trivalent influenza vaccine. DNA samples were collected at enrollment and blood RNA samples were obtained before and on days 1, 3, and 14 after vaccination. Genome-wide genotyping was performed on Illumina HumanOmniExpress microarrays. Global gene expression was assayed on Illumina Human-HT12v3 microarrays. Mixed-effects analysis of variance was employed to account for the effects of genotype, day, person, and day-genotype interactions. The latter are strong candidates as genetic regulators of vaccine-induced responsiveness. Local eQTL analysis was performed, using intervals of 1 megabase up- and downstream of each reporter sequence. We identified 28 loci with evidence of a day effect (variation in expression following the vaccine stimulus) and a highly significant genotype effect (p-values for association < 1 x 10e-15). Content analysis revealed that a majority of these are involved in peptide binding, antigen processing, and antigen presentation. These results provide a striking view of the regions of human DNA that are associated with individual differences in the magnitude of transcriptional responses to influenza vaccines. More generally, they provide a compelling example for human in vivo studies using global gene expression following an experimental stimulus as a powerful approach to complex trait gene mapping.

55

Meta-analysis on eQTL mapping identify common and tissue specific eQTLs in LCL, PBMC and skin tissues. *H. Chen[1], J. Esparza[2], J. Ding[3,4], G. Abecasis[3], Y. Lee[2], M.F. Moffatt[5], W.OC Cookson[5], L. Liang[1].* 1) Dept Epidemiology, Havard School of Public Health, Boston, MA; 2) Max-Delbrück-Center for Molecular Medicine, Berlin, German; 3) Center for Statistical Genetics, Department of Biostatistics, University of Michigan, MI, USA; 4) Laboratory of Genetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; 5) National Heart and Lung Institute, Imperial College London, London, UK.

Large scale genome-wide association studies (GWAS) for gene expression quantitative trait loci (eQTL) mapping have been primarily focused on Epstein-Barr virus transformed lymphoblastoid cell lines (LCL) and have helped interpret findings from GWAS for a variety of complex diseases and traits. However, tissue specific gene expression regulations are also important to understand the genetics of gene expression and provide direct insight into the function of GWAS hits. In this study, we systematically investigate eQTLs from three different tissue types, including LCL, peripheral blood mononuclear cell (PBMC) and skin by a formal statistical approach to test and estimate the common and tissue specific effects for each gene and for each SNP. The 405 LCL samples are from siblings of British descent ascertained through an asthma proband and genotyped using the Illumina 300K panel, the 240 PBMC samples are from 47 German families were genotyped using Affymetrix 6.0 and 500K panels, the 60 skin samples are from a recently published psoriasis GWAS through dbGaP with 400K SNPs genotyped by Perlegen Sciences. All samples were imputed using the latest 1000 Genomes SNP panels (phase1 release). With a total sample size of >800, the power of discovery for eQTL shared across tissues can be substantially increased while tissue specific eQTLs would be more distinctly stand out. For example, for genes in the asthma associated chromosome17 , we find that expression levels for 320 (FDR<5%) genes are associated with at least one SNP within 1Mb of the gene and 130 genes appear to have tissue specific eQTL (FDR<5%). Among these, a proportion were missed by previous eQTL mapping based on LCL alone and appear to betissue specific (FDR<5%). We will illustrate how this newly developed database for common and tissue specific eQTLs could help explain the association to available disease phenotypes, including asthma, atopic dermatitis and psoriasis.

56

Simultaneous Clustering and Significance Testing of Transcription Factor Binding Predictions From Gene Expression Data. *K.S. Kompass[1], D.C. Beebe[2], J.S. Witte[1].* 1) Epidemiology & Biostatistics, UCSF, San Francisco, CA; 2) Ophthalmology and Visual Sciences, Washington University School of Medicine, St. Louis, Missouri.

Computational prediction of cis-regulatory elements and the transcription factors (TFs) that bind them is currently of great interest. The need for computational methods stems partly from the fact that experimental screening and investigation of these mechanisms is costly and labor intensive, while public genome sequence and gene expression databases can be readily used to make computational predictions. Many consensus DNA binding sites for transcription factors have been identified and modeled by position weight matrices (PWMs), which describe the probability of a given base pair at each position of the consensus sequence. A recent method, the `Promoter Analysis Pipeline,' produces ranked scores of the evidence for TF binding based on the frequency of consensus PWM sequences within conserved promoter regions across clusters of co-regulated genes (Chang et al., 2006). The significance of these predictions for a particular gene set is determined by the probability of observing a score more extreme than random, replaced draws of same-sized sets of genes. Because the database of both unique and overlapping PWMs will continue to grow, here we consider an approach that uses all available information from all PWM scores. We applied this method to a gene expression dataset of normal and neoplastic prostate tissue (Wang et al., 2010; Jia et al., 2011). Our results show that individual PWM score significance testing does not hold up well to multiple testing adjustment, and, subsequently, only a small number of PWM scores remained significant after correction when testing them individually. Decomposing the PWM scores with the singular value decomposition produced meta-scores that contained a large percentage of the variation in PWM scores across the gene clusters. This gave a more complete picture of predicted TF binding. The benefits of this approach - a reduction of the multiple testing burden and prediction of jointly acting transcription factors - suggest that it is a useful technique for the interpretation of TF binding predictions from gene expression data.

57

Quantifying significance of phenotype-genotype relationships when various sources of high throughput data on the same individuals are integrated. *H.K. Im[1], E.R. Gamazon[2], M.E. Dolan[2], R.S. Huang[2], N.J. Cox[2,3].* 1) Dept. Health Studies, University of Chicago, Chicago, IL; 2) Dept. Human Genetics, University of Chicago, Chicago, IL; 3) Dept. Medicine, University of Chicago, Chicago, IL.

Given the recent advances in the generation of high-throughput data such as whole genome genetic variation and transcriptome expression, it is critical to come up with novel methods to integrate these data, and more importantly to assess the significance of identified phenotype-genotype relationships. Recent studies show that genome-wide association findings are likely to fall in loci with gene regulatory effects such as expression quantitative trait loci (eQTLs) demonstrating the utility of such integrative approaches. When genotype and gene expression data are available on the same individuals we developed methods where top phenotype-associated genetic variants are prioritized if they are associated, as eQTLs, with gene expression traits that are themselves associated with the phenotype. There has been no method to determine an overall p-value for the findings that arise specifically from the integrative nature of the approach. We propose a computationally feasible permutation method that accounts for the assimilative nature of the method and the correlation structure among gene expression traits and among genotypes. We apply the method to data from a study of cellular sensitivity to etoposide, one of the most widely used chemotherapeutic drugs. Integrative analyses of SNPs, cellular sensitivity to etoposide and transcription level expression were performed in the International HapMap CEU samples. At an overall false discovery rate (FDR) less than 0.05, we identify 13 significant SNPs that associate with cellular sensitivity to etoposide via gene expression. To our knowledge, this study is the first statistically sound quantification of the significance of the genotype-phenotype relationships resulting from applying an integrative approach. This method can be easily extended to cases where gene expression data are replaced by other intermediate traits of interest, e.g., microRNA or proteomic data. This study has important implications for studies seeking to expand on genetic association studies by the use of “omics” data.

58

GPU accelerated genotype imputation for low-coverage high-throughput whole-genome sequencing data. *K. Wang[1,2], G.K. Chen[2].* 1) Zilkha Neurogenetic Institute, Department of Psychiatry, University of Southern California, Los Angeles, CA; 2) Department of Preventive Medicine, University of Southern California, Los Angeles, CA.

High-throughput whole-genome DNA sequencing has been increasingly used in genetic association studies. It is generally believed that a more efficient strategy is to sequence more individuals at lower coverage (<5X) than less individuals at higher coverage, under fixed sequencing costs. When many individuals are sequenced, it is possible to combine sequence data across individuals to impute genotype calls from inferred haplotype patterns using Hidden Markov Model (HMM) based approaches. For example, when *m* individuals with *2m* haplotypes are sequenced, one may define a HMM with (2*m*)^2 states representing all possible pairs. In standard implementations of HMM training with *n* variants, the runtime complexity scales quadratically in the number of states: O(*nm*^2). Thus for the goal of imputation, we are computationally bound by sample size rather than the number of variants. Given the heavy computational burden, various heuristics are often used to maintain tractability (e.g., significantly reducing the number of states). Here we propose a parallel implementation that carries out parameter estimation on Graphical Processing Unit (GPU) devices, where in each step of the algorithm, run time complexity is reduced from quadratic to log2, and/or linear to constant time. Our benchmarks guarantee over 150 fold speedups relative to a serial implementation. We propose a solution to genotype imputation as follows: 1) Investigators first specify a minimal minor allele frequency (MAF) cutoff *a* that they are interested in (for example, 0.5%), which dictates the number of hidden states as 1/*a*^2. 2) Using only genotype calls with posterior probability higher than a set cutoff alpha, we then perform simple whole-genome haplotype phasing into 1/*a*^2 representative haplotypes, that guarantees the inclusion of all variants with MAF higher than cutoff values. Using our parallel algorithm, we re-calculate posterior probability of genotype calls for all potential variants, then re-iterate until convergence. Our algorithm allows one to include a large number of external samples in the imputation procedure (e.g., all subjects from 1000 Genomes Project) in order to increase imputation accuracy and enhance inference in admixed populations. Our GPU accelerated imputation software may also be useful for large-scale sequencing-based GWAS and association studies of targeted regions with complex haplotype patterns.

59

Rare *de novo* and common variations implicate *CARD14* as psoriasis susceptibility locus 2 (PSORS2). *C.T. Jordan[1], E.D.O. Roberson[1], L. Cao[1], K.C. Pierson[2], C.F. Yang[3], C. Ryan[4], R.P. Nair[5], Y. Liu[6], K.C. Duffin[7], Y. Chen[6], B. Feng[7], W.L. Hwu[8], P.E. Stuart[5], G. Hayashi[9], J.Y. Wu[3], C.R. Pullinger[10], J.P. Kane[11], C. Wise[12], L. Peddle[13], V. Chandran[14], P. Rahman[13], D. Gladman[14], G.G. Krueger[7], J.T. Elder[5], W. Liao[9], Y.T. Chen[3], R. Goldbach-Mansky[6], M.A. Lowes[2], A. Menter[4], A.M. Bowcock[1].* 1) Division of Human Genetics, Washington University, Saint Louis, MO, United States of America; 2) Laboratory for Investigative Dermatology, The Rockefeller University, New York, NY, United States of America; 3) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; 4) Psoriasis Research Institute, Baylor University Medical Center, Dallas, TX, United States of America; 5) Department of Dermatology, University of Michigan, Ann Arbor, MI, United States of America; 6) National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH, Bethesda, MD, United States of America; 7) Department of Dermatology, University of Utah, Salt Lake City, UT, United States of America; 8) Department of Pediatrics and Medical Genetics, National Taiwan University Hospital and National Taiwan University College of Medicine, Taipei, Taiwan; 9) Department of Dermatology, University of California, San Francisco, CA, United States of America; 10) Cardiovascular Research Institute and Department of Physiological Nursing, University of California, San Francisco, CA, United States of America; 11) Cardiovascular Research Institute, Department of Medicine, and Department of Biochemistry and Biophysics, University of California, San Francisco, CA, United States of America; 12) Sarah M. and Charles E. Seay Center for Musculoskeletal Research, Texas Scottish Rite Hospital for Children, Dallas, TX, United States of America; 13) Department of Medicine, Division of Rheumatology, Memorial University of Newfoundland, St. John's, Newfoundland, Canada; 14) University of Toronto and Toronto Western Hospital, Toronto, Ontario, Canada.

Psoriasis is a common inflammatory disorder of the skin and other organs including joints (psoriatic arthritis). Genome-wide association studies have identified over 20 risk loci accounting for only a fraction of disease heritability. Linkage analysis localized psoriasis susceptibility locus 2 (PSORS2) to chromosome 17q25.3-qter in one large family of Northern European origin with multiple cases of psoriasis and psoriatic arthritis. Genomic capture and NextGen DNA sequencing of this region in family members revealed a mutation within caspase recruitment domain protein 14 (*CARD14*) mapping to chromosome 17q25 that segregated with disease. A second large 17q25-linked psoriasis family from Taiwan harbored a different mutation in *CARD14*. Both mutations affected the same splice donor site and lead to insertion of 22 amino acids between the critical CARD and coiled-coil domains. A *de-novo* germline *CARD14* mutation (p.Glu138Ala), located in the coiled-coil domain, was identified in a sporadic case of early-onset generalized pustular psoriasis. CARD14 regulates apoptosis and NF-kB signaling. With NF-kB luciferase reporter assays, we demonstrated that psoriasis-associated mutations increased NF-kB activation relative to wildtype CARD14 *in vitro*. Re-sequencing of all coding exons of *CARD14* in 192 cases and 96 controls revealed 7 additional novel missense mutations, but further analysis of 4,000-6,000 cases and controls revealed that very few were associated with disease. Those that were associated also altered NF-kB activation in-vitro. Meta-analysis of six independent case/control cohorts revealed significant association with rs11652075 (p.Arg820Trp). This was also associated with psoriasis in Asians. There was also evidence for genetic interaction of rs11652075 with the MHC (PSORS1). Immunostaining of skin confirmed the presence of CARD14 in epidermal keratinocytes and revealed an atypical distribution in psoriatic skin. We conclude that, following an inflammatory signal, altered regulation of the anti-apoptotic and inflammatory pathways mediated by CARD14 in skin keratinocytes and immune cells lead to psoriasis. These findings enhance our understanding of the pathogenesis of the disease and highlight the value of studying families with Mendelian inheritance of common diseases with complex genetic inheritance.

60

Clinical transcriptome sequencing and analysis of a patient with spiny follicular hyperkeratoses of unknown etiology. *K.V Fuentes Fajardo[1], C.E. Mason[3, 4], M. Huizing[2], M. Nehrebecky[1], M. Turner[5], P. Zumbo[3,4], F. Gill[1], C.F. Boerkoel[1], A.R. Cullinane[2], W.A. Gahl[1,2].* 1) Undiagnosed Diseases Program, OCD, NHGRI/NIH, Bethesda, MD; 2) Medical Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 3) Department of Physiology and Biophysics, Weill Medical College, Cornell University, New York, NY 10065, USA; 4) HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Medical College, Cornell University, New York, NY 10065, USA; 5) Dermatology Branch, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, MD.

Advances in Next Generation Sequencing technology (NGS) have made transcriptome sequencing a practical reality for clinical research applications. Exome sequencing as a tool for finding genetic causes of diseases has become prominent in the literature, and is chosen as the major NGS technology when following up cases in the NIH Undiagnosed Diseases Program (UDP) Protocol. To date, no diagnosis has been arrived at through leads from transcriptome research. The UDP's first attempt at transcriptome sequencing in a clinical setting involved a 50 year-old Caucasian woman with a unique phenotype that included widespread spiny follicular hyperkeratosis resulting in scarring alopecia, follicular plugging and skin abscesses. Her initial milder lesions were exacerbated by a combination of UV-A light treatment and oral retinoids. Thorough dermatologic and medical evaluations to rule out infectious, endocrinologic and paraneoplastic etiologies were unhelpful in delineating an etiology and in clearing the skin problem. Histology and candidate gene sequencing were not diagnostic. Here we present the analysis of transcriptome sequence variations found in RNA from keratinocytes and fibroblasts from involved skin contrasting and comparing them to those found in RNA from fibroblasts and keratinocytes from unaffected skin, blood, and the reference human genome sequence. We identified 18 mutations unique to the keratinocytes and fibroblasts from affected areas, including two genes that indicate a “second-hit,” causing additional damage to a gene already harboring a missense mutation. We also observed that deleterious germline mutations were enriched for genes in cell adhesion, cell transport and immune response (p=0.00002, Bonferroni). These data demonstrate the ability to use RNA-Seq to find genes that carry deleterious mutations and demonstrate dysregulation of expression. These findings can then guide follow-up targets for diagnostic and treatment decisions.

61

Exome sequencing identifies rare recessive IL1F5 variants in Generalised Pustular Psoriasis. *A. Onoufriadis[1], M.A. Simpson[1], A.E. Pink[1], C.H. Smith[1], J. Knight[1], S.L. Spain[1], A.D. Burden[2], F. Capon[1], R.C. Trembath[1], J.N. Barker[1].* 1) Division of Genetics and Molecular Medicine, King's College London, London, United Kingdom; 2) Glasgow Western Infirmary, Glasgow, United Kingdom.

Generalised pustular psoriasis (GPP) is a rare and severe subtype of psoriasis, characterised by the formation of sterile pustules on the skin. It can either occur as a single entity or overlap with other forms of psoriasis. Although infection, pregnancy and drugs are well recognized disease triggers, the genetic architecture of GPP remains poorly understood. To explore the hypothesis that rare penetrant variants contribute to GPP, we sequenced the exomes of five unrelated patients who had no history of psoriasis vulgaris, the form of the disease which affects the majority of patients. Since one affected case was the offspring of consanguineous parents, we compared the exome variant profiles under a rare autosomal recessive model. Only protein altering variants with an estimated population frequency of <0.01 were considered as candidate pathogenic mutations. Under this model, a homozygous S113L substitution of the IL1F5 gene was identified in two individuals, with a third subject being a compound heterozygote for S113L and R48W. Homology searches demonstrated that both mutations affect evolutionary conserved residues that are located in close proximity to critical binding residues in loop 7 and 8 in three dimensional space. All three individuals with IL1F5 mutations shared a highly consistent phenotype, having severe, intermittent, generalised disease, frequently associated with malaise, fever and a marked neutrophilia. None of them had chronic plaque psoriasis or acral involvement. Sequencing of the four coding exons and associated splice sites of IL1F5 in 30 patients with heterogeneous pustular forms of psoriasis (including palmar-plantar pustulosis) did not reveal any coding variants. Having established that IL1F5 mutations contribute to this rare form of psoriasis, we explored the potential role of common IL1F5 variation in psoriasis vulgaris. Using data we had generated in a recent genome-wide association study, we carried out association tests under additive and recessive models, on all genotyped and imputed SNPs lying within 100 kb of IL1F5 (n=690). None gave significant p-values. Taken together, these findings identify mutation of IL1F5 to underlie GPP and implicate defects of the innate immunity pathway in this disorder with a genetic basis that appears distinct from common forms of psoriasis. The IL1F5 protein participates in IL-1 signalling and as such highlights this pathway as a potential target for therapeutic intervention in GPP.

62

Genomic Deletions in Phospholipase C$_2$ Define a New Syndrome of Cold Urticaria, Antibody Deficiency and Susceptibility to Both Autoimmunity and Infection. *M.J. Ombrello[1], E.F. Remmers[1], G. Sun[2], A. Freeman[3], H. Komarow[2], I. Aksentijevich[1], S. Datta[2], P. Torabi-Parizi[4], N. Subramanian[4], N. Romberg[5], T.D. Bunney[6], R.W. Baxendale[6], H.S. Kim[7,8], J. Ho[9], D.C. Douek[10], C. Gandhi[11], A.A. Wanderer[12], H. Lee[13], S. Nelson[13], K.V. Shianna[14], E.T. Cirulli[14], D.B. Goldstein[14], E. Long[7], S. Moir[9], E. Meffre[5], S. Holland[3], M. Katan[6], H. Hoffman[11,15,16], J.D. Milner[2], D.L. Kastner[1].* 1) Inflammatory Disease Section, National Human Genome Research Institute, NIH, Bethesda, MD; 2) Laboratory of Allergic Diseases, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD; 3) Laboratory of Clinical Infectious Diseases, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD; 4) Laboratory of Systems Biology, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD; 5) Department of Immunobiology, Yale University School of Medicine, New Haven, CT; 6) Chester Beatty Laboratories, The Institute of Cancer Research, London; 7) Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD; 8) Department of Medicine, Graduate School, University of Ulsan, Korea; 9) Laboratory of Immunoregulation, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD; 10) Vaccine Research Center, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD; 11) Division of Allergy, Immunology, and Rheumatology, Department of Pediatrics and Medicine, University of California, San Diego, CA; 12) University of Colorado Health Sciences Center, Aurora, CO; 13) Department of Human Genetics, University of California Los Angeles, CA; 14) Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC; 15) Rady Children's Hospital of San Diego, San Diego, CA; 16) Ludwig Institute of Cancer Research, San Diego, San Diego, CA.

Immune dysregulation occurs when the balance of finely-tuned immune regulatory networks is altered, which may lead both to autoimmunity and immune deficiency. Genetic analyses of Mendelian disorders manifesting these seemingly antithetical features may provide new insights into the molecular mechanisms that govern immune responses. We have identified three families with a novel, dominantly inherited complex of cold-induced urticaria, antibody deficiency, and susceptibility to autoimmunity and infection through whom we sought to better understand the interface between autoimmunity and immunodeficiency. Cold urticaria was present in all affected members of each family (n=27), and other immunologic abnormalities were found in 26 of 27 patients. These included antibody deficiency (72%), recurrent infection (59%), atopy (52%), and autoimmunity (48%). Affected subjects had depressed serum immunoglobulins M and A, decreased numbers of circulating NK and class-switched memory B cells, and impairment of B cell central tolerance. Patient B cells also had impaired ligand-mediated activation, which was restored at subphysiologic temperatures. Using SNP genotyping and linkage analysis, we identified a single 7 Mb candidate interval on chromosome 16q (LOD=4.2) in one family, which overlapped by 3.5 Mb a disease-associated haplotype identified in another family. Given its importance in B, NK, and mast cells, *PLCG2* was selected from our candidate interval for mutational screening. Sanger sequencing of *PLCG2* cDNA from purified B cells revealed heterozygous deletions of exon 19 in two families, and exons 20-22 in the third. Long-range PCR and genomic sequencing identified three family-specific deletions in *PLCG2* (4.8-8.2 kb) that co-segregated perfectly with cold urticaria, but were not detected in over 400 healthy control chromosomes. Of note, five of the six deletional breakpoints were within *Alu* or LINE repetitive elements, suggesting a role for repetitive element-mediated recombination in their genesis. The deletions, which are within an autoinhibitory domain of the PLC$_2$ protein, caused constitutive phospholipase activity, but paradoxically resulted in diminished activation of downstream signaling pathways. We describe a novel immunodysregulatory syndrome in which deletions in *PLCG2* cause signaling abnormalities in multiple leukocyte subsets and a pleiotropic phenotype encompassing both excessive and impaired immune function.

63

IL-36 receptor antagonist deficiency causes auto-inflammation and generalized pustular psoriasis. *A. Smahi[1], P. Guigue[1], S. Marrakchi[2], BR. Renshaw[3], A. Puel[4], X.Y. Pei[5], S. Fraitag[11], J. Zribi[2], E. Bal[1], C. Cluzeau[1], M. Chrabieh[4], J.E. Towne[3], J.P. Douangpanya[3], C. Pons[6], S. Mansour[1], V. Serre[1], H. Makni[2], N. Mahfoudh[2], F. Fakhfaf[2], C. Bodemer[1], J. Feingold[1], S. Hadj-Rabia[1], M. Favre[6], E. Genin[8], M. Sahbatou[7], H. Turki[2], J.L. Casanova[4,9], J.E. Sims JE[3], H. Bachelez[10], A. Munnich[1].* 1) Hospital Necker-Enfants Malade, U781 INSERM and University Paris-Descartes, PARIS, France; 2) Department of Dermatology and laboratory of Immunology, Hedi Chaker Hospital, Sfax, Tunisia; 3) Inflammation Research, Amgen, Seattle, WA 98119; 4) Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM U980; 5) Department of Biochemistry, University of Cambridge, Cambridge, UK; 6) Institut Pasteur, Unité de Génétique, Papillomavirus et Cancer Humain, F-75015 Paris; 7) Fondation Jean Dausset, Centre d'Etude du Polymorphisme Humaine (CEPH), 27 rue Juliette Dodu, 75010 Paris, France; 8) Inserm UMRS 946, Institut Universitaire d'Hématologie, Université Paris-Diderot, Paris, France; 9) St. Giles Laboratory of Human Genetics of Infectious Diseases, the Rockefeller University, New York, USA; 10) INSERM U976 and Department of Dermatology. University Paris Diderot. Hôpital Saint-Louis, Paris, France; 11) Department of Pathology, Hôpital Necker-Enfants Malades, Paris, France.

Generalized pustular psoriasis (GPP) is a life-threatening disease of unknown etiology, characterized by the repetition of sudden, triggered episodes of high fever, generalized skin rash and disseminated pustules, with hyperleucocytosis and raised C reactive protein serum level, which may be associated with plaque-type psoriasis. We performed homozygosity mapping and direct sequencing on 9 Tunisian multiplex families with autosomal recessive GPP. Impacts of mutation on protein expression and conformation, stability and function were assessed. We identified a highly significant linkage to a 1,2Mb interval on 2q13-q14.1 and a homozygous missense mutation in the IL-36 Receptor Antagonist (IL-36Ra, previously IL-1F5) anti-inflammatory cytokine gene affecting an evolutionarily conserved residue (L27P). Homology-based structural modelling of human IL-36Ra predicted that the proline at position 27 affects both the stability of IL-36Ra and its interaction with its receptor IL-1RL2 (IL-1Rrp2). Over-expression of IL-36Ra in HEK293T cells and HaCat keratinocytes showed that the L27P mutation greatly reduces the expression of IL-36Ra protein, when compared with either L27R or L27A. Moreover, the L27P mutant was much less potent than wild-type IL-36Ra in inhibiting the IL-36. -induced response in an interleukin-8 (IL-8) reporter assay, leading to enhanced production of inflammatory cytokines by the patients' keratinocytes, IL-8 in particular. These results reveal the key role of IL-36Ra-regulated auto-inflammation in the pathogenesis of GPP, via unregulated IL-1 family inflammatory cytokine secretion in the skin. The identification of GPP as a novel auto-inflammatory syndrome opens new therapeutic perspectives of targeted immunomodulation in both GPP and more common psoriasis forms.

64

Mutations underlying exfoliative ichthyosis reveal a key role for the protease inhibitor cystatin A in keratinocyte adhesion. *H.C. Hennies[2], D.C. Blaydon[1], D. Nitoiu[1], K.M. Eckl[2], R. Cabral[1], P. Bland[1], A. Zvulunov[3], D.P. Kelsell[1].* 1) Centre for Cutaneous Research, The Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom; 2) Division of Dermatogenetics, Cologne Center for Genomics, University of Cologne, Cologne, Germany; 3) Pediatric Dermatology Unit, Schneider Children's Medical Center, Petah-Tiqva, Israel.

Autosomal recessive exfoliative ichthyosis is characterized by fine peeling of the skin, predominantly on the palms and soles, without visible erythema and believed to be induced by mechanical stress. Using whole-genome homozygosity mapping, candidate gene analysis and deep sequencing, we have identified loss-of-function mutations in the gene for protease inhibitor cystatin A (CSTA) as the underlying genetic cause of exfoliative ichthyosis. We found two homozygous mutations, a splice-site and a nonsense mutation, in two consanguineous families of Bedouin and Turkish origin. Electron microscopy of the patients displayed widened intercellular spaces, thickening of keratin filaments, and desmosomal irregularities in the basal layers of the epidermis. By utilising an in-vitro cell stretch assay combined with robust siRNA knockdown, we demonstrate that loss of CSTA leads to destabilised intercellular connections suggesting a possible role in maintaining cell-cell adhesion via desmosomes in the basal keratinocytes. As CSTA is known to inhibit proteases including cathepsins B, H and L, we have investigated the expression of these proteases in the epidermis and cultured keratinocytes. IHC of the three cathepsins in normal skin shows that cathepsin L is expressed throughout all layers of the epidermis whilst cathepsin B and H are more strongly expressed in the suprabasal layers with cathepsin B expressed at lower levels in the basal layer. We have generated 3D models of the disease by knockdown of CSTA in normal primary keratinocytes and organotypic tissue culture. These models demonstrated regular epidermal differentiation, slight hyperkeratosis and a mild disturbance of basal epidermal architecture. The epidermal barrier function was intact, in agreement with a cell adhesion defect observed in basal layers but in contrast to the findings in other disorders characterized by generalized peeling such as peeling skin disease caused by mutations in corneodesmosin. Ongoing work is further looking at how the activity of the target protease(s) is affected by the presence or absence of CSTA in keratinocytes with or without mechanical stress. In summary, our data describes the first report of CSTA with an inherited skin disease and reveals a previously unknown key role for CSTA in basal/suprabasal keratinocyte adhesion.

65

Ablation of TNF or TNFR1 rescues the skin phenotypes of heterozygous female mutant mice. *Y. Gu, D.L Nelson.* Dept Molec & Human Gen, Baylor Col Med, Houston, TX.

Nuclear factor (NF)-2B transcription factor is important for regulating cell survival and anti-apoptosis. Disruption of NF-2B pathway leads to embryonic lethality in NF-2B subunits and its regulator IKK complex knockout mouse models. It has been demonstrated that tumor necrosis factor and its receptor (TNF-TNFR1) are responsible for lethality of these mouse models and ablation of TNF or TNFR1 can rescue of lethality of NF-2B subunit knockouts as well as IKK) knockouts. NF-2B essential modulator (NEMO) or IKK. is essential for activating NF-2B and human females with mutations of this Xq28-linked gene exhibit variable symptoms of Incontinentia Pigmenti (IP). Knockout of *NEMO* in the mouse results in lethality during early embryonic stages in hemizygous males and skin phenotypes in heterozygous females. We have created two different NEMO mutant mouse models by introduction of point mutations into exon 10 of the *Nemo* gene. These model mutations in human *NEMO* that have been identified in patients with atypical forms of IP. Each model showed early embryonic lethality as found in IKK) and NF-2B subunit knockout models. In addition, the stop codon mutation model showed two stages of skin phenotypes in females, similar to the complete Nemo knockout. The 1161C insertion model shows only very mild dermatitis in the older adult females. Both models also displayed partial lethality in heterozygous females. To investigate the mechanism of pathology in the *Nemo* mutations and to determine whether TNF-TNFR1 pathway is responsible for all phenotypes, we placed both models on TNF or TFNR1 knockout backgrounds. Skin symptoms at the early stage (postnatal day 7-10) disappeared and symptoms at the adult later stage were reduced dramatically or occurred later in both TNF and TNFR1 knockout/heterozygous backgrounds. The lethality in the heterozygous *Nemo* females of both models was also rescued by TNF or TNFR1 knockout. However, ablation of TNF or TNFR1 was unable to rescue the lethality observed in hemizygous mutant males at the embryonic stages. These findings demonstrate that the mechanism of skin abnormality of heterozygous females is likely to be distinct from that of lethality of hemizygous mutant males. To further understand the mechanism of two Nemo mutant proteins' effects on NF-2B activity *in vivo*, we are currently making use of NF-2B gal reporter mice from Dr Philip Barker to assess the NF-2B activity i*n vivo* in our two *Nemo* mutant mice in different backgrounds.

66

A mutation in *KEAP1* causes familial multinodular goiter. *R. Teshiba[1,2], T. Tajiri[1], K. Sumitomo[3], K. Masumoto[4], T. Taguchi[1], K. Yamamoto[2].* 1) Department of Pediatric Surgery, Reproductive and Developmental Medicine, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan; 2) Department of Molecular Genetics, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan; 3) Department of Pediatric Surgery, Shimonoseki City Central Hospital, Yamaguchi, Japan; 4) Division of Thoracic, Endocrine and Pediatric Surgery, Fukuoka University Faculty of Medicine, Fukuoka, Japan.

Purpose: Nontoxic multinodular goiter (MNG) is characterized by nodular enlargement of the thyroid gland without thyroid dysfunction and inflammation. Although several studies have mapped linkage regions for familial MNG, the causal mutations have remained unclear. The aim of this study was to identify a responsible gene for familial MNG in a Japanese family. Materials and Methods: A five-generation Japanese MNG family was identified. We performed a genome-wide linkage study by genotyping 13 family members (8 affected and 5 unaffected) using a high density single nucleotide polymorphism (SNP) array (Illumina 370K). After sample and SNP genotype quality checks, multipoint parametric linkage analysis was performed using the GeneHunter program with different marker intervals at 1.0 and 0.2 cM. Subsequently, next-generation sequencing was employed for a proband to identify the causative mutation in the linkage region. The sequences of all exons and intron-exon boundaries of 662 genes within this candidate region were compared with reference genome sequence (hg19). The detected candidate variants were subsequently examined by direct sequencing for all members of the family. Results: A single peak of linkage signal was obtained by multipoint parametric analysis (LOD score, 2.41) on approximate 25 Mb region of chromosome 19. Next-generation sequencing analysis followed by direct sequencing identified c.879C>A, del880G, S293S, fs23X mutation in exon 3 of kelch-like ECH-associated protein 1 gene (*KEAP1*) in a proband. Direct sequencing for family members confirmed that the mutation co-segregated with all affected individuals. Conclusions: We identified a germline mutation in *KEAP1* responsible for a Japanese familial MNG. *KEAP1*, together with nuclear factor-erythroid 2-related factor 2 (*NRF2*), has been shown to play a central role in the protection of cells against oxidative stress. Interestingly, somatic *KEAP1* mutations, which are all different from the mutation described here, have been reported in lung and gall bladder cancers. This is the first report of a germline mutation of *KEAP1* in human, and our result could provide an insight about the etiology in MNG and about the role of KEAP1 pathway in tumorigenesis.

67

***NBEAL2* is mutated in Gray Platelet Syndrome and is required for biogenesis of platelet alpha-granules.** *T. Vilboux[1], T.C. Falik-Zaccai[2,3], Y. Zivony-Elboum[2], F. Gumruk[4], M. Cetin[4], M. Khayat[2], C.F. Boerkoel[1], N. Kfir[2], Y. Huang[1], D. Maynard[1], H. Dorward[1], K. Berger[1], R. Kleta[1], M. Arat[5], A.S. Freiberg[6], B.E. Kehrel[7], K. Jurk[7], P. Cruz[8], J.C. Mullikin[8], J.G. White[9], M. Huizing[1], W.A. Gahl[1,10], M. Gunay- Aygun[1,10].* 1) MGB, NHGRI, NIH, Bethesda, MD; 2) Institute of Human Genetics, Western Galilee Hospital, Naharia, Israel; 3) Ruth and Bruce Rappaport Faculty of Medicine, Technion, Israel Institute of Technology, Haifa, Israel; 4) Pediatric Hematology Unit, Hacettepe University Children's Hospital, Ankara, Turkey; 5) Department of Hematology, Ankara University Faculty of Medicine, Ankara, Turkey; 6) Division of Pediatric Hematology/Oncology, PennState Hershey Children's Hospital, Hershey, PA; 7) Department of Anaesthesiology and Intensive Care, Experimental and Clinical Haemostasis, University Hospital Münster, Münster, Germany; 8) NIH Intramural Sequencing Center, NIH, Bethesda, MD; 9) Department of Laboratory Medicine, University of Minnesota, Minneapolis, MN; 10) Office of Rare Diseases Research, Office of the Director, National Institutes of Health, Bethesda, MD.

Gray Platelet Syndrome (GPS) is a rare autosomal recessive disorder characterized by bleeding tendency, myelofibrosis, thrombocytopenia, and large platelets that lack (-granules. The causative gene has been sought for decades. We mapped the locus for GPS to a 9.4Mb interval on 3p21.1-22.1 that included 197 protein-coding genes. We sequenced these genes using a combination of next generation and Sanger sequencing in 15 independent GPS families. We identified 15 different mutations in *NBEAL2* (neurobeachin-like 2); 5 missense, 4 frameshift, 3 nonsense and 3 splice site mutations. The protein encoded by *NBEAL2* has no known function, yet; however, it belongs to a family of proteins that contains 3 domains (BEACH, ARM and WD40) that are crucial for protein-protein interactions, membrane dynamics and vesicle trafficking. Another protein from this family, LYST, a lysosomal trafficking regulator protein, is defective in Chédiak-Higashi syndrome, a disorder associated with platelet dense granule abnormalities in addition to giant secretory granules in leukocytes and other cell types. RNA analysis showed that at least 7 *NBEAL2* mRNA transcripts are expressed in hematopoietic cells, including megakaryocytes and platelets. Mass spectrometry of discontinuous sucrose gradient platelet fractions localized NBEAL2 protein to the platelet dense tubular system (endoplasmic reticulum). Microarray data in GPS fibroblasts showed overexpression of fibronectin, essential for proplatelet formation in cultured megakaryocytes and critical for megakaryocyte-matrix interactions. This could explain the myelofibrosis seen in GPS patients and therefore can be explored as a therapeutic target. Understanding NBEAL2 function will likely lead to the discovery of novel pathways of organelle formation and maturation.

68

A novel human peroxisome biogenesis disorder affecting peroxisome division. *H.R. Waterham[1], M.S. Ebberink[1], J. Koster[1], I. Stolte-Dijkstra[2], F.J. van Spronsen[3], G.P.A. Smit[3], G. Visser[4], R.J.A. Wanders[1].* 1) Lab Gen Metabolic Dis (FO-222), Univ Amsterdam/Acad Med Ctr, Amsterdam, Netherlands; 2) Dept of Clinical Genetics, Univ Groningen/ Univ Med Ctr, Groningen, Netherlands; 3) Dept of Metabolic Dis, Univ Groningen/ Univ Med Ctr, Groningen, Netherlands; 4) Dept Metabolic Diseases, Wilhelmina Children's Hospital, Utrecht, Netherlands.

Objective Human peroxisomes play an important role in various essential metabolic pathways, among which the biosynthesis of ether phospholipids and the (- and) -oxidation of fatty acids. Consequently, defects in genes encoding peroxisomal proteins can lead to a variety of different peroxisomal disorders, including the peroxisome biogenesis disorders (PBDs). Peroxisomes are dynamic organelles that divide continuously and adjust their protein content in response to the metabolic need. In humans, currently 16 different peroxins (PEX), encoded by PEX genes, have been implicated in various stages of peroxisome biogenesis, protein import, and division. The division of peroxisomes involves elongation of existing peroxisomes followed by membrane constriction and finally fission of peroxisomal tubules. PEX11 has been implicated in peroxisome elongation. In mammals, three PEX11 isoforms(PEX11(, PEX11) and PEX11.) have been identified. Here, we report the identification of the first patient with a defect in PEX11) . **Results** The male patient, who is currently 22 years old, has a clinical picture of congenital cataract, mild mental retardation, hearing loss, sensory nerve involvement and encephalopathic attacks. Based on this clinical presentation he was suspected to suffer from a peroxisomal or mitochondrial disorder, but biochemical peroxisomal as well as mitochondrial parameters were all normal, except for occasional slightly elevated plasma lactate levels. However, immunofluorescence microscopical analysis of patient cells revealed peroxisomes that were often enlarged or elongated and frequently arranged in rows, suggesting a fission defect. After excluding mutations in all 12 PEX genes currently implicated in PBDs, we identified a homozygous nonsense mutation in the PEX11) gene. When the patient cells were cultured at 40°C, the majority of cells lost matrix protein-containing peroxisomes. Overexpression of wild-type PEX11) in such cells reverted the mutant phenotype to normal, while overexpression of PEX11. in such cells resulted in partial reversion to the peroxisomal phenotype observed at 37°C. **Conclusion** Our identification of a novel peroxisomal defect in peroxisome division, which cannot be diagnosed by standard laboratory analysis expands the clinical and genetic spectrum of peroxisomal disorders and indicates that additional peroxisomal defects exist that can only be detected by specific diagnostic workups including microscopical analysis.

69

Mutations in *ABCD4* cause a new inborn error of vitamin B_{12} metabolism. *J.C. Kim[1], D. Coelho[2,3], I.R. Miousse[1], S. Fung[1], M. du Moulin[4], I. Buers[4], T. Suormala[2,3], M. Stucki[2], P. Nürnberg[5], H. Thiele[5], N. Longo[6], M. Pasquali[6], E. Mengel[7], D. Watkins[1], E.A. Shoubridge[1], F. Rutsch[4], J. Majewski[1,8], M. Baumgartner[2], B. Fowler[2,3], D.S. Rosenblatt[1].* 1) Department of Human Genetics, McGill University, Montreal, Canada; 2) Division of Metabolism, University Children's Hospital, Zurich, Switzerland; 3) Metabolic Unit, University Children's Hospital, Basel, Switzerland; 4) University Children's Hospital, Münster, Germany; 5) Cologne Center for Genomics, Cologne, Germany; 6) University of Utah and ARUP Laboratories, Salt Lake City, United States; 7) University Children's Hospital, Mainz, Germany; 8) McGill University and Genome Québec Innovation Centre, Montreal, Canada.

Two patients presented with methylmalonic aciduria and hyperhomocysteinemia. Patient 1 presented at birth following an abnormal newborn screen with hypotonia, lethargy, poor feeding and bone marrow suppression. Patient 2 presented in the newborn period with poor feeding, macrocytic anemia and heart defects. Studies of cultured fibroblasts from both patients showed decreased function of the cobalamin (vitamin B_{12}) dependent enzymes methionine synthase and methylmalonyl-CoA mutase. There was increased uptake of labeled cyanocobalamin (CNCbl) but decreased synthesis of the cobalamin coenzymes methylcobalamin (MeCbl) and adenosylcobalamin (AdoCbl), with accumulation of "free" (i.e. non-protein bound) CNCbl in the cells. The cellular phenotype mimicked that of the *cblF* disorder caused by mutations in the *LMBRD1* gene encoding the lysosomal membrane protein LMBD1 that is thought to play a role in transfer of cobalamin across the lysosomal membrane into the cytoplasm. However, cells from both patients complemented those from all known complementation classes, including *cblF*, and no mutations in *LMBRD1* were found. Whole exome capture (patient 1) and microcell-mediated chromosome transfer and exome capture of chromosome 14 (patient 2), led to the identification of two mutations in the *ABCD4* gene in each patient: c.955A>G (p.Tyr319Cys) and c.1747_1748insCT (p.Glu583LeufsX9) in patient 1 and c.542+1G>T and c.1456G>T (p.Gly486Cys) in patient 2. All mutations were predicted to be deleterious. Transfection of patient fibroblasts with wild type *ABCD4* led to rescue of the abnormal cellular phenotype. Transfection with c.1456G>T did not rescue function confirming the functional significance of this mutation. We conclude that this novel disorder, tentatively named "*cblJ*", is an autosomal recessive disorder caused by mutations in *ABCD4*. We suggest that ABCD4, a presumed ABC transporter, is another essential component of intracellular cobalamin metabolism and might interact with LMBD1 to allow transport of vitamin B_{12} out of the lysosome.

70

A homozygous mutation in the ganglioside biosynthetic enzyme, ST3GAL5, results in a severe autosomal recessive neurocutaneous condition and altered glycosphingolipids and O-linked glycan expression. *C. Schwartz[1, 2], L. Boccuto[1], Q. Zhang[3], F. Bartel[1], K. Aoki[4], X. Fan[4], R. Saul[1], A. Chaubey[1], H. Wang[3], R. Steet[4], M. Tiemeyer[4], X. Yong[3].* 1) Greenwood Genetic Center, Greenwood, SC, USA; 2) Department of Genetics and Biochemistry, Clemson University, Clemson, SC, USA; 3) Beijing Genomics Institute, Shenzhen,China; 4) Complex Carbohydrate Center, University of Georgia, Athens, GA, USA.

In 1983 we identified a family consisting of 3 siblings with abnormal pigmentation and severe intellectual disability which we proposed represented a novel neurocutaneous syndrome we called "Salt and Pepper syndrome" (Saul et al. Proc. Greenwood Genetic Center, 2:6-9, 1983). Clinical features included hyper and hypopigmentation, midface hypoplasia, prognathism, choreoathetosis, severe ID, abnormal EKG and scoliosis. Based on the presentation of the pedigree, we assumed autosomal recessive inheritance and conducted LOH (loss of heterozygosity) studies on two siblings. We identified 4 regions of LOH overlap on chromosomes 2, 6, 8 and 16. Simultaneously, whole exome sequencing of one individual was undertaken after which, in combination with the LOH data, we narrowed our candidate gene search to chromosome 2. We confirmed, by Sanger sequencing, two homozygous mutations: c.994G>A (p.E332K) in the ST3GAL5 gene and c.1150-4C>G in the TCF7L1 gene. The latter appears in the SNP database (rs56311472) as a polymorphism.

The p.E332K mutation in ST3GAL5 was not observed in >1600 chromosomes and bioinformatic analyses indicated it was likely damaging. ST3GAL5 is a sialytransferase which catalyses the initial step in the biosynthesis of complex-type gangliosides. Mass spectrometry analysis of glycosphingolipids (GSLs) extracted from primary fibroblasts cultured from one individual revealed the absence of GM3 ganglioside, the product of ST3GAL5, as well as its biosynthetic derivatives with a concomitant increase in isoglobo-series GSLs. The p.E332K mutation also impacted protein O-glycosylation, giving rise to the production of highly sialylated O-linked glycans. Creation of a zebrafish morpholino knockdown model of ST3GAL5 exhibited similar effects on ganglioside biosynthesis, as well as increased neuronal cell death. Further analysis of ST3GALV-deficient zebrafish and human fibroblast cultures may help identify developmental processes sensitive to altered glycan expression. Our data clearly indicates the neurocutaneous Salt and Pepper syndrome results from a homozygous loss of function missense mutation in ST3GAL5 and is allelic to the Old Amish infantile epilepsy syndrome reported by Simpson et al. (Nat Genet 36:1225-1229, 2004). To date, these conditions represent the only human diseases resulting from a disruption of ganglioside biosynthesis.

71

Mutations within a component of the oligosaccharyltransferase (OST) complex identified by next generation sequencing are implicated in congenital disorders of glycosylation (CDG). *M.A. Jones[1], B.G. Ng[2], P. He[2], M-E. Losfeld[2], S. Bhide[1], D. Rhodenizer[1], E.L.H. Chin[1], M. He[1], H.H. Freeze[2], M.R. Hegde[1].* 1) Emory University School of Medicine, Atlanta, GA; 2) Sanford-Burnham Medical Research Institute, La Jolla, CA.

Congenital disorders of glycosylation (CDG) are comprised of over 30 autosomal recessive disorders caused by deficient N-glycosylation. Multiple organs are affected in these disorders and symptoms can include ataxia, severe developmental delay, seizures, coagulopathies, liver fibrosis and retinopathy. Significant morbidity and mortality is associated with these disorders with 20% of children not surviving beyond 5 years of age. Those that do survive continue to have significant medical problems throughout the rest of their lives. CDG patients are classified as having either Type I or Type II CDG depending on whether the defect resides in the synthesis or processing (Type I) or modification (Type II) of glycans. A diagnosis of CDG is based on biochemical testing but this method cannot identify the specific gene defect. Over 40% of patients lack a confirmatory molecular diagnosis due to limited molecular diagnostic testing available for these disorders and patients having mutations in genes not yet known to cause CDG. We are using combined biochemical and whole exome sequencing (WES) approaches to identify new genes associated with CDG. With the use of next generation sequencing (NGS) technology, we have identified two heterozygous changes in a gene that is a component of the oligosaccharyltransferase (OST) complex. We identified a missense change which is predicted to change the amino acid glycine to aspartic acid at codon 650 (p.G217D) and a 22 bp deletion (c.1265_1286del22) resulting in a stop codon 7 amino acids downstream of the deletion. Sanger sequencing confirmed the NGS results and parental testing confirmed these changes were inherited. The OST complex is part of the N-glycosylation biosynthesis pathway and is responsible for transferring the dolichol-linked precursor glycan onto nascent polypeptides in the ER lumen. Immunoblots using OST48 antibody showed a 50% reduction in protein expression and three protein markers were hypoglycosylated in the patient's cells. FACS, Western blot and immunostaining showed that the plasma membrane marker, ICAM-1, is greatly reduced in patient fibroblasts, and FACS analysis showed similar ICAM-1 reduction in the patient's PBMCs. Our results highlight the power of combining whole exome sequencing with biochemical insights into glycosylation to identify the causative gene in CDG-Ix patients. This should improve molecular diagnosis for this group of disorders.

72

Whole exome sequencing identifies a novel mitochondrial enzyme as the gene responsible for Combined Malonic and Methylmalonic Aciduria. *J.L. Sloan[1], J.J. Johnston[2], I. Manoli[1], R.J. Chandler[1,3], C. Krause[2], N. Carrillo-Carrasco[1], S.D. Chandrasekaran[1], J.R. Sysol[1], K. O'Brien[4], N.S. Hauser[1], J.C. Sapp[2], H.M. Dorward[4], M. Huizing[4], N.I.H. Intramural Sequencing Center[2], B.A. Barshop[5], S. Berry[6], P.M. James[7], N.L. Champaigne[8], P. de Lonlay[9], V. Valayannopoulos[9], M.D. Geschwind[10], D.K. Gavrilov[11], W.L. Nyhan[5], L.G. Biesecker[2], C.P. Venditti[1].* 1) Genetics and Molecular Biology Branch, NHGRI, Bethesda, MD; 2) Genetic Disease Research Branch, NHGRI, Bethesda, MD; 3) Department of Pharmacology, IBS, The Georgetown University, Washington, DC; 4) Medical Genetics Branch, NHGRI, Bethesda, MD; 5) Department of Pediatrics, Biochemical Genetics Laboratory, University of California San Diego, La Jolla, CA; 6) Department of Pediatrics, University of Minnesota, Minneapolis, MN; 7) Department of Pediatrics, Harvard Medical School, Children's Hospital Boston, Boston, MA; 8) Greenwood Genetics Center, Greenwood, SC; 9) Reference Center for Inherited Metabolic Disorders and Necker-Enfants Malades Hospital, Université Paris Descartes, Paris, France; 10) Department of Neurology, University of California San Francisco, San Francisco, CA; 11) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN.

Combined methylmalonic and malonic aciduria (CMAMMA) is characterized by increased urinary methylmalonic acid (MMA) and malonic acid (MA), with MMA>MA, and normal malonyl-CoA decarboxylase activity. CMAMMA was first reported in a child with failure to thrive, seizures and immunodeficiency and a dog with neurodegeneration but the molecular etiology was unknown. We performed exome sequencing of a patient with CMAMMA and her parents and filtered variants to identify 12 candidate genes. One of these genes had a putative mitochondrial leader. We sequenced this gene in 9 patients and found two mutations in this gene in 8 of 9 patients. DNA from the dog also had a homozygous mutation in this gene. Mutations included 9 missense, 1 in-frame deletion and 1 nonsense mutation. Eight missense mutations and the in-frame deletion were in functional motifs conserved among members of this enzyme family. One patient with 2 mutations was identified in an exome cohort of subjects not ascertained for metabolic disease and had the distinct biochemical features of CMAMMA. The age of diagnosis and symptoms in the nine subjects with CMAMMA were highly variable. MMA and MA were elevated in plasma and urine using a new GC-MS assay developed to measure MA. Fibroblasts from 4 subjects had a cellular metabolic defect; increased production of MMA (2.4- to 6-fold) compared to controls, after loading with 5 mM propionate. Viral overexpression of the candidate gene, but not GFP, corrected the metabolic defect. Immunostaining of fibroblasts overexpressing a C-terminal GFP fusion protein or the native enzyme showed a mitochondrial distribution and co-localized with a mitochondrial antibody. These data establish the causative gene for CMAMMA and describe the first disease association with a member of this enzyme family. Mutant alleles occur with a minor allele frequency (MAF) of 0.0058 in ~1,000 control individuals predicting a CMAMMA population incidence of ~1:30,000. This predicts that CMAMMA is one of the most common forms of MMAemia, and perhaps, one of the more common inborn errors of metabolism. The spectrum of symptoms and natural history of this disorder are highly variable and require further delineation. The identification of an affected using exome sequencing highlights an interesting and alternative diagnostic approach because CMAMMA is not identified through routine newborn screening by elevated propionylcarnitine.

73

Identification of mutations causing mitochondrial DNA depletion and translation defects by Next Generation Sequencing. *S.C. Lim[1,2], S.E. Calvo[3,4,5], A.G. Compton[1], S.G. Hershman[3,4,5], T. Yamazaki[1,6], J. Sceneay[1], C. Sugiana[1], A. Laskowski[1], V.K. Mootha[3,4,5], D.R. Thorburn[1,2].* 1) Murdoch Childrens Research Institute, Royal Children's Hospital, Flemington Road, Parkville, Victoria 3052, Australia; 2) Department of Pediatrics, University of Melbourne, Victoria 3010, Australia; 3) Broad Institute of Harvard and MIT, 7 Cambridge Center, Cambridge, MA 02141, USA; 4) Center for Human Genetic Research, Massachusetts General Hospital, 185 Cambridge Street CPZN 5-806, Boston, MA 02114, USA; 5) Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA; 6) Department of Pediatrics, Saitama Medical University, Saitama, Japan.

Mitochondrial oxidative phosphorylation (OXPHOS) disorders affect 1:5000 live births and can be caused by mutations in either the mitochondrial DNA (mtDNA) or nuclear DNA (nDNA). Many patients with OXPHOS disorders have an isolated enzyme defect affecting just one of the five OXPHOS complexes. However, some of the most common diagnoses are `combined enzyme' defects, affecting multiple complexes, such as combined complex I, III and IV deficiency. Identifying the genetic basis of combined enzyme defects is particularly challenging since they can potentially be caused by mtDNA deletions, mutations in mtDNA tRNA genes and in several hundred nDNA genes encoding proteins required for mitochondrial biogenesis and mtDNA maintenance and expression. In addition to this genetic complexity, there is also a wide spectrum of clinical presentations and tissue specificity within these disorders. A surprisingly high proportion of genetic diagnoses can be obtained in some subgroups. For example, we identified autosomal recessive mutations in 43 of 50 patients with combined enzyme defects and hepatic mtDNA depletion by candidate gene analysis of 4 genes (34 *POLG*, 4 *DGUOK*, 3 *MPV17* and 2 *C10orf2* patients). However, in most patients with combined enzyme defects the plethora of candidate genes makes sequential analysis of candidate genes by Sanger sequencing inefficient. We therefore performed a targeted next-generation sequencing analysis of the complete mtDNA and 1034 nDNA genes encoding the known mitochondrial proteome in 16 children with a combined OXPHOS enzyme defect. Sequence variants were filtered on the basis of rarity in healthy individuals, prediction to modify protein function and being consistent with a recessive inheritance model , as previously described (Calvo et al, 2010, Nat. Genet. 42:851-8). We identified clearly pathogenic mutations in 6 patients with combined enzyme defects in 4 known disease genes (*POLG*, *TSFM*, *GFM1* and the recently described *AARS2*) encoding proteins required for mtDNA replication or translation. Pathogenicity of mutations was proven by analyses of mRNA, protein and segregation in the family. An additional 6 patients with combined enzyme defects had recessive-type mutations in 5 novel genes representing new candidates for mitochondrial disease. Proving pathogenecity in those variants remains a substantial challenge in patients lacking a relevant family history, but will become easier as catalogs of benign and pathogenic variants improve.

74

Achieving a novel dynamic in mitochondrial diseases: identification of mitochondrial microRNAs. *A. Henrion Caude[1], S. Bandiera[1], S. Ruberg[2], M. Girard[1], S. Hanein[1], A. Munnich[1], S. Lyonnet[1].* 1) Genetics, Inserm U781 - Necker Hospital, PARIS, France; 2) Miltenyi Biotec GmbH, Bergisch Gladbach, Germany.

MicroRNAs (miRNAs) are endogenous small non-coding RNAs that associate with Argonaute proteins and regulate gene expression by interfering with translation or stability of target transcripts. To date, genetic variations in miRNAs and/or their mRNA targets that have been shown to disrupt their interaction and thereby contribute to Mendelian disorders concern in particular hearing loss and neurodegenerative diseases. Intriguingly, mitochondrial defects are common to most of those diseases. We thus hypothesized that miRNAs may directly target mitochondria to provide a sensitive and rapid mechanism by which to adjust the mitochondrial energy production. This implies a localization of miRNAs and the related machinery to mitochondria to achieve a miRNA mediated post-transcriptional regulation of mitochondrial genome. To enable identification of mitochondrial-enriched miRNAs, we profiled the mitochondrial and cytosolic RNA fractions from the same cells by miRNA microarray analysis. We found 57 miRNAs differentially expressed in mitochondria and cytosol. Of these 57, a signature of 13 nuclear-encoded miRNAs was reproducibly enriched in mitochondrial RNA and validated by RT-PCR for hsa-miR-494, hsa-miR-1275 and hsa-miR-1974. The significance of their mitochondrial localization was investigated by characterizing their genomic context, cross-species conservation and instrinsic features such as their size and thermodynamic parameters. Interestingly, the specificities of mitochondrial miRNAs versus cytosolic miRNAs were underlined by significantly different structural and thermodynamic parameters. The functional relevance of miRNAs at mitochondria was further supported by the finding of Argonaute 2 localization to mitochondria revealed by immunoblotting, confocal and electronic microscopy, and also validated by the co-immunoprecipitation of the mitochondrial transcript COX3. Targeting analysis of most mitochondrial miRNAs revealed not only nuclear but also mitochondrial-encoded targets. The novel localization of RNA interference components in human mitochondria unravels the molecular bases for a novel layer of crosstalk between nucleus and mitochondria. Those mitochondrial miRNAs that we termed `mitomirs' provide a new and dynamic mean to revise the pathogenic mutations of the mitochondrial genome.

75

Multiple novel loci conferring risk to coronary artery disease identified in a study of 63,253 cases and 126,820 controls. *S. Kanoni[1], C. Willenborg[2], J. Thompson[3], K. Stirrups[1], C. Nelson[3], J. Danesh[4], J. Erdmann[2], A. Hamsten[5], S. Kathiresan[6], J.S. Kooner[7], R. Roberts[8], U. Thorsteinsdottir[9], H. Watkins[10], H. Schunkert[2], N. Samani[3], P. Deloukas[1] for the CARDIoGRAMplusC4D consortium.* 1) Wellcome Trust Sanger Institute, Hinxton, UK; 2) Universität zu Lübeck, Medizinische Klinik II, Lübeck, Germany; 3) Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester, UK; 4) Department of Public Health and Primary Care, University of Cambridge, UK; 5) Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden; 6) Cardiovascular Research Center and Cardiology Division, Massachusetts General Hospital, Boston and Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 7) Hammersmith Hospital, National Heart and Lung Institute, Imperial College London, London, UK; 8) The John and Jennifer Ruddy Canadian Cardiovascular Genetics Centre, University of Ottawa Heart Institute, Ottawa, Ontario, Canada; 9) deCODE Genetics, 101 Reykjavik, Iceland; 10) Cardiovascular Medicine and Wellcome Trust Centre for Human Genetic, Oxford, UK.

Genome wide association studies (GWAS) have identified 30 loci influencing susceptibility to coronary artery disease (CAD) and its main complication myocardial infarction (MI). Recent large scale GWAS meta-analyses have also yielded a large number of additional associations that haven't been evaluated systematically. The Metabochip array is a custom iSELECT chip containing 196,725 SNPs, designed for two purposes (i) follow-up of putative associations in several cardiometabolic traits and (ii) fine mapping (FM) of confirmed loci in these traits. CARDIoGRAM has provided the SNP content for CAD/MI: 5658 SNPS for replication and 21440 for FM (22 loci). In the CARDIoGRAMplusC4D consortium we undertook a two stage meta-analysis to test for disease risk 5658 SNPS associated with CAD at P<0.01 in CARDIoGRAM (22,233 cases and 64,762 controls). Stage 2 involved 34 studies totaling 41,020 cases and 62,058 controls of European or South Asian descent ~60% of which were typed with the Metabochip array and the remaining had GWA data imputed using HAPMAP. We applied study-wise standard QC and corrected for population stratification using 4310 SNPs (associated with long QT but at least 5Mb away from established CAD loci) if λ_{GC} 1.05. We then combined p-values (2-sided) for all 5658 SNPs from Stage 1 (CARDIoGRAM discovery phase) with their respective Stage 2 p-values (1-sided) using Fisher's method. In the All cases vs controls model (adjusted for sex and age) 50 SNPs representing 18 novel loci had a combined p value < 10^{-6}. Of those, 12 loci reached genome wide significance (WGS; p<5 x 10^{-8}): IL6R, APOB, VAMP8-GGCX, SLC22A4, GUCY1A3, KCNK5, PLG, TRIB1, ABCG5, FURIN, and FLT1. Four of the novel WGS CAD loci APOB, TRIB1, ABCG5, and FURIN as well as EDNRA (p=1.08 x 10^{-7}) are known lipid loci. Interestingly, IL6R is a key molecule in the inflammatory cascade and together with the recent report of impairment of IFN.-response signaling via the 9p21 locus there is growing evidence linking inflammation to CAD risk - the most associated IL6R variant was intronic (p=3.52 x 10^{-8}). An additional new CAD locus at 2p24.1 (gene desert near FLJ40608) reached genome wide significance in the male only analysis (adjusted for age). The 13 new loci reported here increase our insight in to the pathogenesis of CAD.

76

Twenty-nine Common Variants Associated with Blood Pressure and Cardiovascular Disease Risk: The International Consortium for Blood Pressure Genome-Wide Association Studies (ICBP-GWAS). *G.B. Ehret[1], P.B. Munroe[2], K.M. Rice[3], M. Bochud[4], A.D. Johnson[5], D.I. Chasman[6], A.V. Smith[7], M.D. Tobin[8], G.C. Verwoert[9], V. Pihur[10], N.R.G. Shrine[8], L.V. Wain[8], M. Boehnke[11], M.G. Larson[5], M.R. Järvelin[12], B.M. Psaty[13], G.R. Abecasis[14], A. Chakravarti[10], P. Elliott[12], C.M. van Duijn[9], C. Newton-Cheh[15], D. Levy[5], M.J. Caulfield[2], T. Johnson[2] on behalf of the ICBP-GWAS consortium.* 1) Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA and Cardiology, Geneva University Hospital, 1211 Geneva, Switzerland; 2) Clinical Pharmacology and The Genome Centre, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK; 3) Department of Biostatistics, University of Washington, Seattle, WA, USA; 4) Institute of Social and Preventive Medicine (IUMSP), Centre Hospitalier Universitaire Vaudois and University of Lausanne, Bugnon 17, 1005 Lausanne, Switzerland; 5) Framingham Heart Study, Framingham, MA, USA; 6) Division of Preventive Medicine, Brigham and Women's Hospital, 900 Commonwealth Avenue East, Boston MA 02215, USA; 7) Icelandic Heart Association, Kopavogur, Iceland; 8) Department of Health Sciences, University of Leicester, University Rd, Leicester LE1 7RH, UK; 9) Department of Epidemiology, Erasmus Medical Center, PO Box 2040, 3000 CA, Rotterdam, The Netherlands; 10) Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 11) Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, 48109, USA; 12) Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK; 13) Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle, WA, USA; 14) Center for Statistical Genetics, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI 48103, USA; 15) Center for Human Genetic Research, Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, 02114, USA.

Blood pressure (BP) is a classic heritable trait under the control of multiple biological pathways and environmental stimuli. Even small increments in BP are associated with increased risk of cardiovascular events and the risk is much higher in the presence of hypertension (BP ⁄ 140 mm Hg systolic [SBP] or ⁄ 90 mm Hg diastolic BP [DBP]), a trait that affects roughly one third of the general population. We performed a genome-wide association study of SBP and DBP in a multi-stage design using 200,000 individuals of European descent and identified 16 novel BP-associated loci. Our experiment confirmed all but one of 13 associations previously reported by our groups. A total of 29 independent associations (P < 5x10-8) were at 28 loci (ADM, ATP2B1, BAT2-BAT5, C10orf107, CACNB2, CYP17A1-NT5C2, CYP1A1-ULK3, EBF1, FES, FGF5, FLJ32810-TMEM133, GNAS-EDN3, GOSR2, GUCY1A3-GUCY1B3, HFE, JAG1, MECOM, MOV10, MTHFR-NPPB, NPR3-C5orf23, PLCE1, PLEKHA7, SH2B3, SLC39A8, SLC4A7, TBX5-TBX3, ULK4, ZNF652). Twenty-two loci did not contain genes that were a priori strong biological candidates for BP phenotypes. Thirteen of 29 SNPs were associated with altered gene expression levels. For several of these SNPs, the SNP (or proxy) most strongly associated with BP at a locus was the same as the SNP most strongly associated with transcript levels. Eight of the 29 SNPs were in high LD with a non-synonymous coding SNP. Each of the 29 SNPs has an effect size of 0.3-1.1mmHg per risk allele in the GWAS and the 29 SNPs explain 0.9% of the phenotypic variance of SBP and DBP. We genotyped individuals of East Asian (N=29,719), South Asian (N=23,977), and African (N=19,775) ancestries and found significant associations at specific loci in Asians. Genetic risk scores incorporating all 29 SNPs were strongly associated with SBP and DBP in each non European ancestry group. In an independent sample of 23,294 women, an increase of 1 SD in the genetic risk score was associated with a 21% increase in the odds of hypertension. The genetic risk score was significantly associated with left ventricular wall thickness, occurrence of stroke, and CAD. The same genetic risk score was not, however, associated with chronic kidney disease or measures of kidney function, despite having a similar sample size to other outcomes tested. Our findings provide new insights into the genetics and biology of BP, and suggest potential novel therapeutic pathways for cardiovascular disease prevention.

77

A large-scale genome-wide association study of PR interval identifies 41 loci associated with atrial and atrioventricular conduction. *A. Isaacs[1,2], J. van Setten[3], D.E. Arking[4], E.J. Rossin[5], D.S. Evans[6], V. Gudnason[7], C. Hayward[8], A.A. Hicks[9], Y. Jamshidi[10], K. Kerr[11], B. Krijthe[12], T. Lehtimäki[13], M. Müller[14], S. Padmanabhan[15], A. Parsa[16], O. Polasek[17], R.B. Schnabel[18], K. Stefansson[19,20], T. Tanaka[21], A. Teumer[22], K. Tarasov[23], S. Trompet[24], S. Ulivi[25], Vand. Genomic Electronic Records[26], P. van der Harst[27], X. Yin[28], P.I.W. de Bakker[29], N. Sotoodehnia[30] on behalf of the PRIMA Consortium.* 1) Genetic Epidemiology Unit, Dept. of Epidemiology, Erasmus MC Rotterdam, Rotterdam, Netherlands; 2) Centre for Medical Systems Biology, Leiden, the Netherlands; 3) Dept. of Medical Genetics, University Medical Center Utrecht, Utrecht, the Netherlands; 4) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 5) Harvard Medical School, Boston, MA, USA; 6) Dept. of Medicine, University of California, San Francisco, San Francisco, CA, USA; 7) Icelandic Heart Association, Kopavogur, Iceland; 8) Medical Research Council (MRC) Human Genetics Unit, Institute of Genetics and Molecular Medicine, Edinburgh, UK; 9) Institute of Genetic Medicine, European Academy Bozen-Bolzano (EURAC), Bolzano, Italy, affiliated institute of the University of Lübeck, Germany; 10) Dept. of Twin Research and Genetic Epidemiology Unit, St. Thomas' Campus, King's College London, St. Thomas' Hospital, London, UK; 11) Department of Biostatistics, University of Washington, Seattle, WA, USA; 12) Dept. of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands; 13) Laboratory of Atherosclerosis Genetics, Dept. of Clinical Chemistry, University of Tampere, Tampere, Finland; 14) Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology and Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany; 15) Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, University Place, Glasgow, UK; 16) Dept. of Medicine, University of Maryland, Baltimore, MD, USA; 17) Andrija Stampar School of Public Health, Medical School, University of Zagreb, Zagreb, Croatia; 18) Dept. of General and Interventional Cardiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; 19) deCODE Genetics, Reykjavik, Iceland; 20) Faculty of Medicine, University of Iceland, Reykjavik, Iceland; 21) National Institute on Aging, NIH, DHHS, Baltimore, MD, USA; 22) Interfaculty Institute for Genetics and Functional Genomics, Ernst Moritz Arndt University Greifswald, Greifswald, Germany; 23) Laboratory of Cardiovascular Science, National Institute on Aging, NIH, DHHS, Baltimore, Maryland, USA; 24) Dept. of Cardiology, Leiden University Medical Center, Leiden, the Netherlands; 25) Medical Genetics, IRCCS-Burlo Garofolo/University of Trieste, Trieste, Italy; 26) Vanderbilt University Medical Center, Nashville, TN, USA; 27) Dept. of Cardiology, University Medical Center Groningen, University of Groningen, the Netherlands; 28) Boston University and National Heart, Lung, and Blood Institute's (NHLBI) Framingham Heart Study, Framingham, Massachusetts, USA; 29) Division of Genetics, Dept. of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; 30) Division of Cardiology, Dept. of Medicine, University of Washington, Seattle, Washington, USA.

PR interval duration (PR), determined electrocardiographically, is a measure that reflects intra-atrial and atrioventricular conduction. Its prolongation is a demonstrated risk factor for atrial fibrillation and is associated with the development of conduction abnormalities such as heart block. PR is heritable; population-based studies have estimated narrow-sense heritabilities of ~30%. In the current study, a genome-wide screen for common variants associated with PR among individuals of European descent was conducted. Twenty-nine population-based cohorts participated, leading to an aggregate maximum effective sample size of 87,543 after exclusions (for atrial fibrillation or use of Type I or III anti-arrhythmics). Each study analyzed the relationship between PR and ~2.5 million genotyped and HapMap imputed SNPs using linear regression models (linear mixed models for family studies) under the assumption of an additive genetic effect. These results were synthesized with an inverse variance weighted fixed effects meta-analysis. More than 30 novel loci were identified at genome-wide significance levels ($P < 5x10^{-8}$). Additionally, all previous PR GWAS loci (*SCN5A/SCN10A*, *CAV1/CAV2*, *NKX2-5*, *SOX5*, *WNT11*, *MEIS1*, *TBX5/TBX3*, and *ARHGAP24*) were successfully replicated. A number of loci, including *SCN5A/SCN10A*, *ARHGAP24*, *CAV1/CAV2*, *TBX5/TBX3*, *c12orf67*, *ID2* and *TMEM182*, appear to harbor multiple independent associations ($r^2 < 0.05$). Novel loci included multiple genes and pathways previously implicated in cardiac (patho)physiology including: *TTN*, implicated in cardiomyopathies, and *OBSCN*, which interacts with *TTN*, suggesting the importance of that pathway in PR determination; a myosin binding protein (*MYBPHL*); *FRMD4B*, associated with heart failure; *MYOCD*, expressed in aorta and involved in regulating the expression of cardiac specific genes; *CAMK2D*, a calcium dependent protein kinase; *EOMES*, a t-box containing transcription factor, similar to *TBX5/TBX3*; and a number of genes involved in signal transduction. The products of these genes are connected via a protein-protein network ($P < 0.015$). Together, these loci (considering only the most significant SNP per locus) account for approximately 5% of the variation in PR (~16% of the heritability). These analyses provide insights into the biology of atrial and atrioventricular conduction and may eventually help in the prediction, prevention and treatment of PR-associated disease such as heart block or atrial fibrillation.

78

Association meta-analysis of correlated EKG traits detect novel loci that influence electrically active left ventricular cardiac mass. *J. van Setten[1], A. Isaacs[2], D.E. Arking[3], J.C. Chambers[4], M. Eijgelsheim[5], C. Hayward[6], S. Hwang[7], Y. Jamshidi[8], K. Kerr[9], T. Lehtimäki[10], I. Mateo Leach[11], M. Mueller[12], O. Meirelles[13], A. Pfeufer[14], O. Polasek[15], N.J. Samani[16], A.V. Smith[17], R. Sorice[18], N. Sotoodehnia[19], T. Tanaka[20], A. Teumer[21], S. Trompet[22], S. Ulivi[23], P.I.W. de Bakker[1], P. van der Harst[11] on behalf of the QRS Voltage Consortium.* 1) Dept. of Medical Genetics, University Medical Center Utrecht, Utrecht, Netherlands; 2) Genetic Epidemiology Unit, Dept. of Epidemiology, Erasmus University Medical Center, Rotterdam, Netherlands; 3) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; 4) Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK; 5) Dept. of Epidemiology, Erasmus University Medical Center, Rotterdam, Netherlands; 6) Medical Research Council (MRC) Human Genetics Unit, Institute of Genetics and Molecular Medicine, Edinburgh, UK; 7) National Heart, Lung, and Blood Institute's (NHLBI) Framingham Heart Study, Framingham, MA, USA; 8) Dept. of Twin Research and Genetic Epidemiology Unit, St. Thomas' Campus, King's College London, St. Thomas' Hospital, London, UK; 9) Department of Biostatistics, University of Washington, Seattle, WA, USA; 10) Dept. of Clinical Chemistry, University of Tampere, Tampere, Finland; 11) Dept. of Cardiology, University Medical Center Groningen, Groningen, Netherlands; 12) Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology and Chair of Epidemiology, Ludwig-Maximilians University, Munich, Germany; 13) Laboratory of Genetics, National Institute on Aging, NIH, DHHS, Baltimore, MD, USA; 14) Institute of Genetic Medicine, European Academy Bozen-Bolzano, Bolzano, Italy; 15) Andrija Stampar School of Public Health, Medical School, University of Zagreb Zagreb, Croatia; 16) Dept. of Cardiovascular Sciences, University of Leicester, Leicester, UK; 17) Icelandic Heart Association, Kopavogur, Iceland; 18) IGB-ABT, National Research Council (CNR), Rome, Italy; 19) Division of Cardiology, Dept. of Medicine, University of Washington Seattle, WA, USA; 20) National Institute on Aging, NIH, DHHS, Baltimore, MD, USA; 21) Interfaculty Institute for Genetics and Functional Genomics, Ernst Moritz Arndt University Greifswald, Greifswald, Germany; 22) Dept. of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, Netherlands; 23) Medical Genetics, IRCCS-Burlo Garofolo/University of Trieste, Trieste, Italy.

Background: Increased left ventricular mass (LVM) is a common manifestation of preclinical cardiovascular disease associated with increased morbidity and mortality. Although previous studies found a number of loci strongly associated with LVM, they explain only a small proportion of the trait variance in the population. In order to detect novel LVM loci, we performed a large GWAS meta-analysis of three correlated EKG traits based on QRS duration*voltage products as surrogates for LVM. Methods: Twenty-four population-based cohorts participated with >55,000 samples of European descent were available after exclusions for atrial fibrillation and other conditions, QRS duration >120 msec, and Type I and III anti-arrhythmics drug use. Each study performed a GWAS for Sokolow-Lyon product (SL), Cornell voltage product (CV), and 12-lead sum product (12LS) with ~2.8 million SNPs using linear regression models adjusted for age, gender, height, and BMI. Cohort-level results were meta-analyzed using a fixed-effects model. Genome-wide significance was adjusted for genomic inflation and for the number of traits studied. Results: In total, 1,261 SNPs, spanning 29 genomic regions, were significantly associated with QRS duration*voltage products (13 loci with SL, 13 with CV, and 18 with 12LS). Eleven loci were associated with more than one phenotype. Six loci included well-studied genes previously implicated in hypertrophic cardiomyopathies, including *TTN, MYBPC3, PLN*, and *TNNT2*. Eight were known QRS duration loci, including *HAND1, TBX3, SCN5A* and *CDKN1A*. Novel loci included the Z-line protein *MYOZ1, DPYSL5, CDH13*, and *CTNNA3*. The latter two are cadherin and cadherin-associated proteins, both highly expressed in the heart. *MYOZ1* is involved in myofibrillogenesis, and mutations in a closely related gene, *MYOZ2*, were previously associated with myopathies. Cross-trait analyses highlighted another 12 loci that will require replication. Conclusion: This study, coupled with on-going functional experiments, provides novel insights into the genetic basis of electrically active myocardium and may eventually aid efforts to predict, prevent and treat LVM associated diseases such as hypertrophic cardiomyopathy.

79

Metabonomic and genetic fine mapping reveal wide range of associations across metabolites in known lipid loci. *T. Tukiainen[1,2,3,4], J. Kettunen[1,5], A.J. Kangas[2], L.-P. Lyytikäinen[6], P. Soininen[2,7], A.-P. Sarin[1,5], E. Tikkanen[1,5], P.F. O'Reilly[4], M.J. Savolainen[2,8], K. Kaski[3], A. Jula[5], T. Lehtimäki[6], M. Kähönen[9], J. Viikari[10], M. Jauhiainen[5], J.G. Eriksson[5,11,12,13,14], O. Raitakari[15,16], V. Salomaa[5], M.-R. Järvelin[4,8,17,18,19], M. Perola[5], A. Palotie[1,20,21,22], M. Ala-Korpela[2,7,8], S. Ripatti[1,5].* 1) Institute for Molecular Medicine Finland FIMM, FIN-00014 University of Helsinki, Helsinki, Finland; 2) Computational Medicine Research Group, Institute of Clinical Medicine, University of Oulu and Biocenter Oulu, FIN-90014 University of Oulu, Oulu, Finland; 3) Department of Biomedical Engineering and Computational Science, Aalto University School of Science, FIN-00076 Aalto, Espoo, Finland; 4) Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom; 5) Department of Chronic Disease Prevention, National Institute for Health and Welfare, FIN-00271 Helsinki, Finland; 6) Department of Clinical Chemistry, University of Tampere and Tampere University Hospital, FIN-33521 Tampere, Finland; 7) NMR Metabonomics Laboratory, Department of Biosciences, University of Eastern Finland, FIN-70211 Kuopio, Finland; 8) Department of Internal Medicine and Biocenter Oulu, Clinical Research Center, FIN-90014 University of Oulu, Oulu, Finland; 9) Department of Clinical Physiology, University of Tampere and Tampere University Hospital, FIN-33521 Tampere, Finland; 10) Department of Medicine, University of Turku and Turku University Hospital, FIN-20521 Turku, Finland; 11) Department of General Practice and Primary Health Care, FIN-00014 University of Helsinki, Finland; 12) Unit of General Practice, Helsinki University Central Hospital, Helsinki, Finland; 13) Folkhälsan Research Centre, Helsinki, Finland; 14) Vaasa Central Hospital, Vaasa, Finland; 15) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, FIN-20521 Turku, Finland; 16) Department of Clinical Physiology, Turku University Hospital, FIN-20521 Turku, Finland; 17) Institute of Health Sciences, University of Oulu, Oulu, Finland; 18) Biocenter Oulu, University of Oulu, Oulu, Finland; 19) Department of Lifecourse and Services, National Institute of Health and Welfare, Oulu, Finland; 20) Department of Human Genetics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK; 21) Department of Medical Genetics, University of Helsinki and the Helsinki University Hospital, Helsinki, Finland; 22) The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

A recent genome-wide meta-analysis of over 100,000 individuals identified 95 genetic loci to associate with plasma levels of cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol (HDL-C) or triglycerides. For many of these loci the biological function and causal variants remain unknown. We performed a genetic and metabonomic fine mapping of these regions in 8330 individuals from five population-based Finnish cohorts utilizing the latest NMR-based serum metabonomics platform with 216 metabolites and derived variables, including a broad panel of lipoprotein subclasses, and a dense set of genotyped and imputed genetic markers. Twenty-five of the 95 loci were associated (p-value < $5*10^{-8}$) with one or more metabolite measures, the associations on the lipoprotein subclasses highlighting specific parts of the apolipoprotein B metabolic pathway or HDL metabolism. The greatest heterogeneity underlay the loci previously associated with HDL-C; for example, there were several loci where the same allele lowered the levels of the smallest HDL particles but elevated the larger HDL particles. The subclass profiling enabled a detailed characterization of the gene effects on the lipoprotein cascade and revealed novel associations for several genes. For example, *PPP1R3B*, a gene previously identified to primarily modulate HDL-C levels, displayed the strongest associations to intermediate-density lipoprotein lipids. A few loci, including *GCKR* and *APOA1*, had associations beyond the lipoprotein measures: the associations to alanine, branched-chain amino acids and lactate point to the wider involvement of these genes in energy metabolism. Twelve of the lipid loci harbored multiple, statistically independent variants. As an example, the independent variants in *APOB* associated with two different metabolite profiles that can be related to the different functional domains of the gene. Our data show that there is considerable diversity in association patterns between loci originally identified through associations with total lipid measures and reveal metabolic profiles reaching far beyond routine clinical lipid measures. Further understanding of these processes may open up new possibilities to understand mechanisms involved in metabolic diseases. The results also advocate the use of cost-effective metabonomics-based biomarker profiles in genetics and epidemiological studies to improve understanding of underlying disease mechanisms.

80

Meta-analysis of Metabochip SNPs identifies novel waist-hip ratio loci involved in adipogenesis and chronic inflammation disorders. *D.C. Croteau-Chonka[1], T. Ferreira[2], D. Shungin[3,4,5], T.W. Winkler[6], R. Magi[2], R.J. Strawbridge[7], A.E. Locke[8], K. Fischer[9], T. Workalemahu[10], P.J. Griffin[11], C.C. White[11], A.U. Jackson[8], F. Day[12], M.C. Zillikens[13], I. Barroso[14,15], C.S. Fox[16], E. Ingelsson[17], M.I. McCarthy[18], K.E. North[19], E.K. Speliotes[20,21], P.W. Franks[3,5,10], L.A. Cupples[11], L. Qi[10], I.M. Heid[6], K.L. Mohlke[1], C.M. Lindgren[2], R.J.F. Loos[12], The Genetic Investigation of Anthropometric Traits (GIANT) Consortium.* 1) Department of Genetics, University of North Carolina, Chapel Hill, NC; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 3) Genetic and Molecular Epidemiology Group, Department of Public Health and Clinical Medicine, Section for Medicine, Umeå University, Umeå, Sweden; 4) Department of Odontology, Umeå University, Umeå, Sweden; 5) Department of Clinical Sciences, Skåne University Hospital, Lund University, Malmö, Sweden; 6) Regensburg University Medical Center, Department of Epidemiology and Preventive Medicine, Regensburg, Germany; 7) Cardiovascular Genetics and Genomics Group, Karolinska Institute, Stockholm, Sweden; 8) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 9) Estonian Genome Center, University of Tartu, Estonia; 10) Department of Nutrition, Harvard School of Public Health, Boston, MA; 11) Department of Biostatistics, School of Public Health, Boston University, Boston, MA; 12) MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK; 13) Department of Internal Medicine, Erasmus MC Rotterdam, the Netherlands; 14) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK; 15) University of Cambridge Metabolic Research Labs, Institute of Metabolic Science Addenbrooke's Hospital, Cambridge, UK; 16) National Heart, Lung, and Blood Institute, Framingham, MA; 17) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 18) University of Oxford, Oxford, UK; 19) Department of Epidemiology and Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, NC; 20) Department of Internal Medicine, Division of Gastroenterology, and Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 21) Broad Institute, Cambridge, MA.

Waist-hip ratio (WHR), a measure of fat distribution, is associated with metabolic and cardiovascular (CV) diseases. To build on the GIANT consortium's previous meta-analysis of WHR genome-wide association (GWA) studies, we genotyped 59,904 additional individuals from 24 studies using the Metabochip custom array, which includes SNPs selected for metabolic or CV-related traits. A total of 99,024 SNPs were tested for additive association with WHR, adjusting for body mass index, a measure of overall adiposity. We then performed fixed effects meta-analyses of these SNPs combining with 72,919 individuals from the discovery GWA set (Heid *et al., Nat Gen*, 2010). All 14 previously described WHR loci replicated with consistent direction of effect (13 at $P < 5.0 \times 10^{-8}$ and one at $P = 3.0 \times 10^{-5}$). Overall, more SNPs selected to follow up suggestive WHR signals from the GWA scan showed directional consistency with the discovery GWA stage than expected by chance (2,948 of 4,887 compared to 2,444, $P_{binomial} = 1.85 \times 10^{-47}$).

We identified seven novel loci associated with inter-individual variation in WHR ($P < 5.0 \times 10^{-8}$). The strongest associations included SNPs near *PEMT* ($P = 3.5 \times 10^{-11}$), *CEBPA* ($P = 3.4 \times 10^{-10}$), *BTNL2* ($P = 1.8 \times 10^{-9}$), and *ANKRD55* ($P = 3.6 \times 10^{-9}$). The *CEBPA* signal is located near two subunits of the transcription factor C/EBP, which is essential for white adipose tissue differentiation. C/EBP also binds the promoters of adiponectin and leptin, suggesting potentially broad effects on many downstream metabolic pathways. Different variants near the *PEMT*, *ANKRD55*, and *BTNL2* genes have been associated previously with chronic inflammatory disorders, including coronary artery disease (CAD), rheumatoid arthritis, and ulcerative colitis, respectively. Among these, only the *PEMT* index SNP for CAD is present on the Metabochip; while the CAD index SNP is in low linkage disequilibrium ($r^2 = 0.33$) with the WHR index SNP, its risk allele is still nominally associated with increased WHR ($P = 1.8 \times 10^{-5}$). These results highlight the potential role of inflammatory genes in fat distribution.

Increased sample size at loci targeted by the Metabochip further elucidates the genetic underpinnings of fat distribution and provides exciting new clues regarding underlying biological mechanisms.

81

A Human Hormone Sensitive Lipase (HSL) Knockout: Window into mechanisms of dyslipidemia and diabetes. *J. Albert[1], R. Horenstein[1], L. Yerges-Armstrong[1], U. Sreenivasan[1], S. Snitker[1], J. O'Connell[1], J. McLenithan[1], A. Shuldiner[1, 2], C. Sztalryd[1, 2, 3], C. Damcott[1].* 1) Med, Univ Maryland, Balt, Baltimore, MD; 2) The Geriatric Research, Education and Clinical Center, Baltimore Maryland 21201; 3) Veterans Affairs Health Care Center, University of Maryland, Baltimore, Maryland 21201.

Impaired lipolysis is associated with both dyslipidemia and glucose intolerance/type 2 diabetes (T2D), two major risk factors for cardiovascular disease. To uncover the genetic contributions to dysfunctional lipolysis in humans, we sequenced genes in the lipolysis pathway in 24 Amish subjects with extreme serum triglyceride (TG) levels. We identified a novel 19 base pair deletion that disrupts the coding region of the *LIPE* gene, which encodes hormone sensitive lipase (HSL), the rate limiting enzyme of diacyglyceride (DAG) breakdown, the second step in the lipolysis pathway. Genotyping (n = 1893 subjects) revealed that the deletion (allele frequency = 0.02 in the Amish) was significantly associated with higher fasting insulin levels (p = 3.1X10-5), increased ectopic fat accumulation in the liver (p = 6.5X10-6), insulin resistance (p = 2.9x10-5), increased fasting TGs (p = 3.3 x 10-3), and lower HDL-cholesterol (p = 1.8 x 10-5). Western blots of white adipose tissue (WAT) from deletion homozygotes (n = 2) showed absence of HSL protein, and thin layer chromatography lipid profile analysis showed accumulation of intracellular DAG, effectively identifying the first human HSL knock-out (KO). Histologically, lack of HSL in WAT showed marked heterogeneity in fat cell size and a ~50% overall reduction in size (p = 0.026). In isolated HSL KO adipocytes, lipolysis expressed per number of cells was suppressed both in basal (PIA 20μM) and in stimulated (isoproterenol 1μM) conditions (p = 0.030 and 0.036, respectively). Most importantly, percent of insulin suppression of lipolysis at maximum dose of insulin was decrease in HSL KO adipocytes (p = 9.8X 10-4). In conclusion, HSL deficiency in humans leads to impaired lipolysis as well as WAT and systemic insulin resistance. This suggests that HSL is a key player in regulating glucose and lipid metabolism in humans.

82

The 9p21.3 coronary artery disease risk genotype is associated with elevated serum levels of interferon alpha 21. *N.A.M. Almontashiri[1,2], M. Fan[1,2], H.H. Chen[3], A.C.T. Teng[1,2], B.L.M. Cheng[2], R. McPherson[2], R. Roberts[2], A.F.R. Stewart[1,2].* 1) Biochem, Microbiol & Immunol, University of Ottawa, Ottawa, Canada; 2) Ruddy Canadian Cardiovascular Genetics Centre, University of Ottawa Heart Institute, Ottawa, Canada; 3) Ottawa Hospital Research Institute, Ottawa, Canada.

Background - The 9p21.3 locus confers risk for coronary artery disease (CAD) by an unknown mechanism. Several long-range enhancers were recently identified at CDKN2B, CDKN2A, MTAP and between the interferon =-1 (IFNW1) and interferon (-21 (IFNA21) genes, whose chromatin structure is influenced by single nucleotide polymorphisms within the 9p21.3 risk locus as far as 1 million base pairs away. Here, we tested whether protein expression from the genes in the vicinity of these long-range enhancers is correlated with the 9p21.3 risk genotype. Methods and Results - Using immunoblot analysis of protein extracts from primary cultures of human aortic smooth muscle cells (n=6) and peripheral blood lymphocytes from patients with CAD (n=18) genotyped for the 9p21.3 risk locus using SNP microarrays, we found that CDKN2A protein expression in both cell types, and CDKN2B protein expression in aortic smooth muscle cells was negatively associated with 9p21.3 risk genotype. No association between MTAP expression and 9p21.3 genotype was seen in either cell type. However, a strong positive correlation ($p<10^{-4}$) between IFNA21 levels and 9p21.3 risk genotype was seen in both cell types. Importantly, ELISA of serum from 242 genotyped CAD patients revealed highly elevated levels of IFNA21 in individuals homozygous for the 9p21.3 risk allele ($p=1.38x10^{-50}$). Conclusion - Elevated serum IFNA21 levels are a biomarker for the 9p21.3 CAD risk locus. Together with reduced CDKN2A protein expression, elevated serum IFNA21 likely accounts for the risk of CAD associated with the 9p21.3 locus.

83

A novel and common X-linked inborn error of carnitine biosynthesis associated with autism. *P.B.S. Celestino-Soper[1], S. Violante[2,3], E. Crawford[4], J. Ge[1], A.L. Hall[1], K. Lee[1], C. Lo[1], R. Luo[5], T. Moss[1], K.N. Mohan[1], R. Person[1], B. Sadikovic[1], C.A. Shaw[1], S.J. Sanders[6], M.W. State[6], D. Geschwind[5], J. Sutcliffe[4], R.J.A. Wanders[2], S.M. Leal[1], E. Cook[7], R. Goin-Kochel[1], F.M. Vaz[2], A.L. Beaudet[1,8].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 2) Laboratory Genetic Metabolic Diseases, Departments of Clinical Chemistry and Pediatrics, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands; 3) Metabolism and Genetics Group, Research Institute for Medicines and Pharmaceutical Sciences, iMed.UL, Faculdade de Farmácia, Universidade de Lisboa, Av. Prof. Gama Pinto, 1649-003 Lisboa, Portugal; 4) Department of Molecular Physiology & Biophysics, Center for Molecular Neuroscience, Vanderbilt University, Nashville, TN 37232-8548, USA; 5) Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA; 6) Program on Neurogenetics, Child Study Center and Departments of Psychiatry and Genetics, Yale University School of Medicine, New Haven, CT 06520, USA; 7) Institute for Juvenile Research, Department of Psychiatry, University of Illinois at Chicago, Chicago, IL 60608 USA; 8) Texas Children's Hospital, Houston, TX 77030, USA.

Objectives: Our goal was to discover novel genes mutated in autism. **Methods:** We used array CGH with genome-wide, exon by exon coverage to screen 100 simplex families from the Simons Simplex Collection (SSC). **Results:** One male in the first 100 trios and then numerous other males with autism were found to have inherited deletions of exon 2 of the *TMLHE* gene on the X-chromosome leading to the delineation of a novel inborn error of metabolism. Deletions of exon 2 are null mutations and were recurrent with slightly different boundaries in each family. *TMLHE* encodes the first enzyme in the pathway for biosynthesis of carnitine, and deficiency is characterized by increased trimethyllysine (TML) and decreased hydroxy-TML and . -butyrobetaine in plasma, urine, and brain. TMLHE deficiency is at least 20 times more common than PKU in males (1 in ~400), and deletion of exon 2 was 3.1 times more common in probands of male-male affected sib pairs with autism compared to control males (P = 0.026); no significant association of deletion of exon 2 with simplex autism cases was detectable. TMLHE deficiency is present in 0.5-1% of autism, and the penetrance of 20-40% in U.S. males may be dependent on carnitine intake during infancy. **Conclusions:** TMLHE deficiency is extremely common in “normal” males, although subtle phenotypes may occur in these individuals. Because *TMLHE* is highly expressed in Purkinje cells and hippocampal neurons, and carnitine is essential for mitochondrial function at the synapse, we propose a neuronal carnitine deficiency hypothesis as a risk factor for autism. The fraction of non-dysmorphic autism (NDA) associated with neuronal carnitine deficiency could be limited to TMLHE deficiency or could be much greater if nutritional carnitine deficiency was a common associated risk factor for NDA. If males are somehow more susceptible than females to impairment through neuronal carnitine deficiency, this could explain the up to 8:1 male:female ratio in NDA and/or the putative increasing in the frequency of autism and be a common risk factor for autism. Autism associated with TMLHE deficiency or with nutritional carnitine deficiency may be preventable or partially reversible using carnitine, acetylcarnitine, or butyrobetaine dietary supplementation. This work was supported by a grant from the Simons Foundation Autism Research Initiative (SFARI).

84

Next-generation sequencing in 248 families with X-linked intellectual disability. *V.M. Kalscheuer[1], H. Hu[1], S.A. Haas[2], J. Chelly[3,4], H. Van Esch[5], M. Raynaud[6,7,8,9], A. de Brouwer[10,11], T. Zemojtel[2], G. Froyen[12,13], S.G.M. Frints[14,15], F. Laumonnier[7,8,9], M.I. Love[2], N. Lebrun[3,4], M. Field[16], E. Haan[17,18], M. Corbett[17], G. Turner[16], M. Shaw[17], G. Gillessen-Kaesbach[19], U. Müller[20], A. Latos-Bieleńska[21], T. Kleefstra[10], K. Wrogemann[1,22], R. Ullmann[1], T. Jentsch[23], J. Gecz[17,18], A. Tzschach[1], H. van Bokhoven[10,11], W. Chen[1,23], H.H. Ropers[1], et.al.* 1) Human Molecular Genetics, Max Planck Institute for Molecular Genetics, Berlin, Germany; 2) Department Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin, Germany; 3) University Paris Descartes, Paris, France; 4) Institut Cochin, INSERM Unité 1016, CNRS UMR 8104, Paris, France; 5) Center for Human Genetics, University Hospitals Leuven, B-3000 Leuven, Belgium; 6) Inserm U930 “Imaging and Brain”, Tours, France; 7) University François-Rabelais, Tours, France; 8) CNRS ERL3106, Tours France; 9) Centre Hospitalier Régional Universitaire, Service de Génétique, Tours, France; 10) Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 11) Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 12) Human Genome Laboratory, Department for Molecular and Developmental Genetics, VIB, Leuven, Belgium; 13) Human Genome Laboratory, Department of Human Genetics, K.U.Leuven, Leuven, Belgium; 14) Department of Clinical Genetics, Maastricht University Medical Center, azM, Maastricht, The Netherlands; 15) School for Oncology and Developmental Biology, GROW, Maastricht University, Maastricht, The Netherlands; 16) The GOLD Service, Hunter Genetics, Waratah, New South Wales, Australia; 17) SA Pathology, Women's and Children's Hospital, Adelaide, South Australia, Australia; 18) School of Paediatrics and Reproductive Health, University of Adelaide, Adelaide, South Australia, Australia; 19) Institut für Humangenetik Lübeck, Universitätsklinikum Schleswig-Holstein, Lübeck, Germany; 20) Institut für Humangenetik, Justus-Liebig-Universität Giessen, Giessen, Germany; 21) Poznan University of Medical Sciences, Poznan, Poland; 22) Department of Biochemistry & Medical Genetics, University of Manitoba, Winnipeg, MB Canada; 23) Max-Delbrueck-Centrum fuer Molekulare Medizin, Berlin, Germany.

Genetic heterogeneity is an unappreciated factor underlying the complexity of many common disorders and intellectual disability in particular. The X-chromosome linked intellectual disability (XLID) accounts for 10% of all forms of intellectual disability and is one of the most heterogeneous traits known. In excess of 95 XLID genes have been identified, yet mutations in these genes account for only about half of the cases. To advance our understanding of the molecular causes of XLID, we have employed hybrid capture and Massively Parallel Sequencing to screen X-chromosomal exons in probands from 248 unrelated XLID families that had been recruited by the European MRX Consortium and associated groups. Out of these, 90 families were resolved; 29 families had mutations in 14 novel XLID genes. While two of the novel XLID genes encode synaptic proteins, the others play a role in diverse, fundamental cellular processes, including transcription, translation and protein degradation. Recently, various ID genes have also been implicated in autism, schizophrenia, epilepsy and related disorders, and independent evidence indicates that single gene defects play a hitherto underestimated role in these diseases, too. This study, apart from shedding new light into the pathogenesis of intellectual disability and broadening the basis for its molecular diagnosis, serves as a model for the elucidation of many other complex disorders. Supported by BMBF (MRNET), Max-Planck-Society, EU (GENCODYS), NOW, Dutch Brain Foundation, WCH Foundation and Australian NHMRC.

85

Exome sequencing implicates *de novo* mutations in the actin genes *ACTB* and *ACTG1* in Baraitser-Winter syndrome. *J.B. Riviere[1], B.W.M. van Bon[2], A. Hoischen[2], B.J. O'Roak[3], S.S. Kholmanskikh[4], A. Verloes[5], D. Pilz[6], V.M. Siu[7], M. Rossi[8], O.A. Abdul-Rahman[9], J.F. Atkin[10], M.J.M Nowaczyk[11], G.M.S. Mancini[12], M.E. Ross[4], J. Shendure[3], J.A. Veltman[2], H.G. Brunner[2], W.B. Dobyns[1,13].* 1) Center for Integrative Brain Research, Seattle Children's Hospital, Seattle, WA; 2) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 3) Department of Genome Sciences, University of Washington, Seattle, WA; 4) Department of Neurology and Neuroscience, Weill Medical College of Cornell University, New York, NY; 5) Département de Génétique, Hôpital Robert DEBRE, Paris, France; 6) Institute of Medical Genetics, University Hospital of Wales, Cardiff, UK; 7) Department of Pediatrics, University of Western Ontario, London, ON, Canada; 8) Department of Pediatrics, Federico II University, Naples, Italy; 9) Division of Genetics, University of Mississippi Medical Center, Jackson, MS; 10) Department of Pediatrics, College of Medicine, Ohio State University, Columbus, OH; 11) Department of Pathology and Molecular Medicine, McMaster University, Hamilton, ON, Canada; 12) Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands; 13) Department of Pediatrics, University of Washington, Seattle, WA.

Congenital brain malformations are individually rare diseases characterized by extreme phenotypic and genetic heterogeneity. Most overlap with developmental brain disorders having normal brain structure such as intellectual disability and autism. Similarly to those disorders, the large majority of congenital brain malformations occur sporadically and have severely reduced reproductive fitness, thus pointing towards a causative role for *de novo* mutations. Our research is aimed at uncovering the genetic basis of congenital brain malformations using massively parallel sequencing. To overcome the great genetic heterogeneity of these disorders, we carefully selected rare and severe brain malformation syndromes with multiple-matching features. In the present study, we focused on Baraitser-Winter syndrome (BWS), a rare disorder of unknown etiology characterized by intellectual disability, neuronal migration disorder, ocular colobomas, specific facial dysmorphism and variable hearing loss. In the absence of familial recurrence, we hypothesized that the genetic basis of BWS was likely to result from *de novo* protein-altering mutations. We therefore performed whole-exome sequencing in three proband-parent trios; filtered for potential causal variants based on frequency, function, and absence in parents; and identified *de novo* missense changes in the actin genes *ACTB* and *ACTG1* in one and two probands, respectively. Sequencing of coding exons of these two genes in nine additional affected individuals revealed four and five *de novo* missense mutations in *ACTB* and *ACTG1*, respectively, thus indicating that mutations in *ACTB* or *ACTG1* are the cause of BWS in all probands tested. Several of the mutations were recurrent *de novo* mutations, suggesting a gain-of-function or a dominant negative mechanism. Actin participates in many key cellular processes and is extremely conserved throughout evolution. Among the six actin genes, only *ACTB* and *ACTG1* are ubiquitously expressed; the remaining four genes are expressed primarily in muscle. The protein products of *ACTB* and *ACTG1* are nearly identical, differing by only four amino acids. Our results show that even in a scenario of limited genetic heterogeneity, whole-exome sequencing of a small number of trios combined with validation of candidate genes in additional patients is a powerful approach to identify mutations underlying sporadic congenital syndromes. These data also emphasize the crucial role of actin in human development.

86

Massive parallel sequencing of all human protein-coding genes identifies *PACS1* as a new gene for a new intellectual disability syndrome. *J.H.M. Schuurs-Hoeijmakers[1], L.E.L.M. Vissers[1], M. Holvoet[2], M.E.M. Swinkels[5], C. Gilissen[1], M.A. Willemsen[3,4], P. de Vries[1], J.A. Veltman[1], B.B.A. de Vries[1], H. van Bokhoven[1,3], A.P.M. de Brouwer[1,3], K. Devriendt[2], H.G. Brunner[1].* 1) Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences and Institute for Genetic and Metabolic Disorders, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands; 2) Center for Human Genetics, Clinical Genetics, K.U.Leuven, Leuven, Belgium; 3) Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 4) Department of Pediatric Neurology and department of Neurology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 5) Department of Medical Genetics, University Medical Center Utrecht, The Netherlands.

Intellectual disability occurs in 2-3% of the population and although its origin is often genetic, only a minority of the genes involved have been identified. Recent studies have indicated that *de novo* mutations in genes with neuronal functions are important causes of intellectual disability. We studied two patients with moderate-severe intellectual disability and a remarkably similar facial appearance. Because of the striking resemblance of their facial features we suspected an identical genetic defect. However, consultation with international syndromologists failed to provide a clinical diagnosis. To our surprise, we found an identical non-synonymous c.607C>T *de novo* mutation in the *PACS1* gene in both individuals by exome sequencing of these patients and their parents. *PACS1* has not previously been associated with developmental delay and encodes an intracellular sorting protein that mediates its effects through the binding of acidic clusters on cargo protein. In the nervous system, PACS-1 regulates subcellular protein trafficking in neurons. Our finding of an identical *de novo* nucleotide change in two boys selected only on the basis of facial appearance clearly demonstrates the power of whole exome sequencing as a diagnostic tool in the evaluation of patients with suspected genetic conditions of unknown origin, such as cognitive disorders. It also illustrates that diagnostic interpretation of these exome data is greatly facilitated by detailed clinical phenotyping and independent replication of these rare genetic observations.

87

Autism- and intellectual disability-associated variations in the transcription factor ZBTB20 differentially alter dendritic and synaptic structure. *A.K. Srivastava[1,2], Y. Luo[1,7], K.A. Jones[3], L. Dukes-Rimsky[1], S.M. Sowell[1], D.P. Srivastava[3], S. Ladd[1], B.R. DuPont[1,2], J.S. Collins[1,2], C.M. Wilson[4], C. Skinner[1], F. Gurrieri[5], R.E. Stevenson[1,2], E. Boyd[4], P. Penzes[3,6].* 1) J.C. Self Research Institute of Human Genetics, Greenwood Genetic Center, Greenwood, SC; 2) Department of Genetics and Biochemistry, Clemson University, Clemson, SC; 3) Department of Physiology, Northwestern University Feinberg School of Medicine, Chicago, IL; 4) Fullerton Genetics Center, Asheville, NC; 5) Institute of Medical Genetics, Catholic University, Rome, Italy; 6) Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL; 7) Present address: Department of Pediatrics, Emory University School of Medicine Atlanta, GA.

Autism spectrum disorder (ASD) and intellectual disability (ID) are the two most frequently reported, often co-morbid, neurodevelopmental disorders. Both disorders have a strong genetic component. Dendritic arborization and dendritic spine morphology are key determinants of neuronal connectivity and function, and are disrupted in many neurodevelopmental disorders in which cognitive function is affected. Here we have found a genetic association of the broad complex/tramtrack/bric-a-brac (BTB)-zinc finger protein ZBTB20 with ASD and ID. We initially found a disruption of the ZBTB20 gene at 3q13.2 by a chromosomal translocation t(3;12) in a patient with mild autism and Rett syndrome-like features. Subsequently, we identified three rare nonsynonymous ZBTB20 alterations in unrelated patients, two of which appeared unique to ASD and one was unique to ID. We examined the effects of the ZBTB20 alterations on dendritic architecture of cultured rat cortical pyramidal neurons. The two mutations that associated with ASD mapped to the N-terminal and central regions of the ZBTB20 protein and altered dendritic spine morphology. In contrast, the ID-linked mutation mapped to the C-terminus and affected dendritic arborization and dendritic length, but had no effect on dendritic spine morphology. Furthermore, truncation of either the extreme N- or C-terminus mimicked the corresponding point mutation-specific alterations in spine or dendritic morphology. These data suggest a potential role for ZBTB20 in ASD and ID and provide a molecular and cellular basis for phenotypic differences resulting from changes within the same gene. These findings provide further insight into molecular mechanisms that regulate dendritic and synaptic structure whose genetic disruption may contribute to specific neurodevelopmental disorders.

88

Mutation in ARID1B is a frequent cause of intellectual disability. *A. Reis[1], J. Hoyer[1], A.B. Ekici[1], S. Endele[1], B. Popp[1], C. Zweier[1], A. Wiesener[1], E. Wohlleber[2], A. Dufke[3], C. Petsch[1], I. Göhring[1], A.M. Zink[2], G. Rappold[4], E. Schröck[5], D. Wieczorek[6], O. Riess[3], H. Engels[2], A. Rauch[1,7].* 1) Inst. of Human Genetics, Friedrich-Alexander Univ, Erlangen, Germany; 2) Inst. of Human Genetics, Rheinische Friedrich-Wilhelms-Univ., Bonn, Germany; 3) Inst. of Human Genetics, Univ. of Tuebingen, Tuebingen, Germany; 4) Dept. of Human Molecular Genetics, Univ. of Heidelberg, Heidelberg, Germany; 5) Inst. für Klinische Genetik, Technische Univ. Dresden, Dresden, Germany; 6) Inst. für Humangenetik, Universitätsklinikum Essen, Essen, Germany; 7) Inst. of Medical Genetics, Univ. of Zurich, Schwerzenbach-Zurich, Switzerland.

Interstitial deletions of chr. 6q25 are rare findings associated with intellectual disability (ID), hypotonia and facial dysmorphism. CNV screening with molecular karyotyping in 1,986 patients with ID mostly of German ancestry identified a 2.5Mb *de novo* deletion in chr. 6q25 encompassing only 5 genes in an infant with severe ID, absent speech, facial dysmorphism and muscular hypotonia. We hypothesized that haploinsufficiency for one of these genes may impair normal development and screened all five positional candidates for point mutations in a subset of 121 patients with moderate to severe ID without detectable microdeletion or other known genetic cause. Following the identification of one patient with mutation in *ARID1B*, we expanded mutation screening to a total of 905 individuals with ID and found overall 7 *de novo* sequence alterations: One 11bp deletion, one 2 bp deletion, one splicing mutation, one frameshifting duplication of exons 5 and 6 and 3 nonsense mutations, all resulting in a premature translational termination. We thus identified *de novo* sequence alterations in *ARID1B* in 0.8 % of patients with unexplained ID. 7 out of 8 patients presented with moderate to severe ID, severe speech impairment especially regarding expressive speech and muscular hypotonia. The majority had short stature, minor anomalies of teeth, hands and feet, and minor facial dysmorphism such as thin upper lip, bulbous nasal tip, and low set or posteriorly rotated ears. This phenotypic spectrum resembles that reported for 6q25 deletions with the only exception of hearing loss suggesting that *ARID1B* haploinsufficiency is the major contributor to the phenotypic spectrum of this microdeletion syndrome. Our observation of several *de novo* truncating mutations and the recent report of a disruption of *ARID1B* in a translocation patient with ID, support the hypothesis of haploinsufficiency as the underlying mechanism. *ARID1B* is highly expressed in brain and encodes AT rich interactive domain 1B, a subunit of a mammalian SWI/SNF complex important for different neuronal developmental programs that was recently shown to act as an E3 ubiquitin ligase targeting histone H2B at lysine 120 for monoubiquitination *in vitro*. We conclude that *ARID1B* mutations are a relatively frequent cause of moderate to severe ID and our findings support the growing evidence for a role of altered chromatin remodeling in ID. This study is part of the German Mental Retardation Network (www.german-mrnet.de).

89

A novel dynamic mutation in AFF3 associated with developmental delay. *F. Kooy[1], S. Metsu[1], L. Rooms[1], J. Gecz[2], D.R. FitzPatrick[3].* 1) Dept Medical Genetics, Univ Antwerp, Antwerp, Belgium; 2) Women's and Children's Hospital, Neurogenetics, Genetics and Molecular Pathology, Adelaide, Australia; 3) Medical Genetics Section, MRC Human Genetics Unit, Edinburg, United Kingdom.

Many types of dynamic mutations are cytogenetically visible as fragile sites, a mechanism first described in the fragile X syndrome. Although it is known that additional fragile sites of the rare, folate sensitive type occur more frequently in cohorts of patients with cognitive disorders than in control populations, few autosomal fragile sites have been characterized at the molecular level. We studied three families with FRA2A expression. The probands were characterized by slow early motor and language development and learning disability. As the molecular basis of this fragile site we identified an elongated polymorphic CGG repeat in intron 2 of AFF3. Expansion of the AFF3 associated repeat was shown in the probands of the 3 families and in five family members using Southern blot analysis. AFF3 is a paralog of FMR2, causative of FRAXE non-syndromic mental retardation. Further members of this family of nuclear transcription factors include AF4 and AF5q31. All are involved in regulation of gene expression, cell expansion, embryonic development and fulfill a role in the central nervous system. Dynamic expansion of the intronic repeat associated with FRA2A resulted in hypermethylation of the AFF3 promoter as determined by bisulfite sequencing. By cSNP-analysis we confirmed that methylation of the promoter region silences transcription of the AFF3 gene in the patients. Most parents of the patients were not methylated and expressed both copies of the AFF3 gene. In conclusion, we describe for the first time that an intronic CGG repeat is able to repress expression of the associated gene. Furthermore, we collected compelling evidence that transcriptional silencing of the AFF3 gene could play a role in the cognitive impairment in the patients.

90

Using Next-Generation Sequencing to Investigate Rare Unidentified Congenital Disorders. *A.C. Need[1], K. Schoch[2], Y. Hitomi[1], V. Shashi[2], D.B. Goldstein[1].* 1) Center for Human Genome Variation, Duke University School of Medicine, Durham, North Carolina 27708; 2) Division of Medical Genetics, Department of Pediatrics, Duke University Medical Center, Durham, NC 27712, USA.

Approximately 1-3% of children manifest with developmental delay and/or intellectual disabilities, and at least half of these are thought to have an underlying genetic etiology. Identifying the genetic cause is important for prognosis, potential treatment options and family planning. To this end, once non-genetic causes have been ruled out, karyotypes and chromosomal microarrays are used to look for structural changes, and targeted genetic testing and metabolic tests are performed to investigate any known conditions with similar presentation. After this, approximately 50% remain undiagnosed. There is intense interest in the community now in evaluating how to use next-generation sequencing approaches to identify mutations responsible for such conditions Here we have selected 12 families with unknown congenital disorders (with different features in each case), and performed next-generation sequencing to try to identify the cause. The analysis strategy is predicated on the assumption that the causal variants are either *de novo* or recessive. The proband and both parents are initially whole-exome sequenced and variants are selected if they are absent in both parents and 900 controls (putative de novos), or if they are homozygous in the proband, heterozygous in both parents and rare or absent in the controls (recessives). The raw alignments of all putative de novos are then inspected, and the majority are eliminated because of low coverage or evidence of the variant in the parents that was not sufficient to be called. All genes affected by remaining variants are then inspected for candidacy based on known human conditions and animal phenotypes. Of the 9 trios analyzed to date, 2 potentially causal variants have been identified, and these are undergoing functional follow up. The remaining 7 presented no good candidates and are now undergoing whole-genome sequencing to investigate copy number variants and uncaptured exons. This approach represents a new tool in clinical genetics and could significantly increase the proportion of families that receive explanations for their child's disability.

91

Pre-phasing: a computationally efficient approach for imputing from new reference panels in genomewide association studies. *C. Fuchsberger[1], B. Howie[2], M. Stephens[2,3], G. Abecasis[1], J. Marchini[4].* 1) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, University of Chicago; 3) Department of Statistics, University of Chicago; 4) Department of Statistics, University of Oxford.

Genotype imputation is a key step in the analysis of genomewide association studies (GWAS). The approach works by identifying haplotypes shared between study samples, which are typically genotyped on commercial SNP arrays, and more extensively characterized individuals, such as those examined by The International HapMap Project or The 1000 Genomes Project. Imputation methods can accurately assign genotypes (or genotype probabilities) at markers that have not been directly observed, thereby increasing GWAS power while facilitating fine-mapping efforts and meta-analyses. While there are several effective approaches for performing imputation, technological improvements are rapidly increasing the numbers of SNPs and individuals in GWAS and reference data sets, which increases the computational demands on existing methods. We describe and evaluate an efficient approach for addressing this problem. Our approach works in two steps. First, haplotypes are estimated for each of the GWAS individuals. Second, the estimated haplotypes are used directly for imputation. This approach reduces the computational cost of imputation in two ways: (i) the GWAS samples must be phased only once, whereas standard methods would implicitly re-phase with each reference panel update; (ii) computationally, it is much faster to match a phased GWAS haplotype to one reference haplotype than to match unphased GWAS genotypes to a pair of reference haplotypes. For example, when we imputed the GAIN Psoriasis data set using a CEU panel from the 1000 Genomes Pilot Project (120 haplotypes typed at 6.8M variants), our cluster required ~550 CPU hours to run a conventional imputation approach like MACH or IMPUTE v1. By contrast, an initial implementation of our pre-phasing approach required <25 CPU hours to impute into the same data set. Using a conventional approach, the squared correlation (r2) between imputed array genotypes and those determined using Affymetrix arrays was 0.91 for SNPs with MAF >5%, 0.76 for SNPs with MAF of 3-5% and 0.67 for SNPs with MAF 1-3%. Despite its much lower computational cost, the pre-phasing approach actually improved accuracy to 0.92, 0.77 and 0.69 for these three MAF bins. Our methods are implemented in C++, run on Windows, Mac and Linux and are available at http://genome.sph.umich.edu/wiki/minimac (minimac) and http://mathgen.stats.ox.ac.uk/impute/impute_v2.html (IMPUTE 2.0).

92

Phasing of Many Thousands of Genotyped Samples. *A. Williams[1,2], N. Patterson[2], D. Reich[1,2].* 1) Harvard Medical School, Boston, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA.

We introduce an efficient method for phasing large genotype datasets of unrelated individuals applicable to up to 100,000 samples or more. Our method is practically relevant because the accuracy of existing methods improves with larger sample sizes and because genotype datasets continue to grow in size. We evaluate our method on a merged dataset with 16,265 samples from the Wellcome Trust Case Control Consortium (WTCCC) and 88 HapMap CEU individuals over 386,353 SNPs. The 88 HapMap CEU samples are parents of trios and extremely accurate haplotypes are available; we compare to these haplotypes to compute switch error rates. We verified that existing methods increase in accuracy as sample size grows by separately merging 1000, 3000, 5000, and 16,265 WTCCC samples with the 88 HapMap samples and running Beagle. Compared to running with 1088 samples, Beagle's switch error rate improves by 17.8%, 23.2%, and 35.5% for the datasets with 3088, 5088, and 16,265 samples, respectively. We use an iterative approach that builds a hidden Markov model (HMM) based on estimated haplotypes, infers the maximum likelihood estimated haplotypes from this HMM for each sample, and then repeats the process by rebuilding an HMM from the updated haplotypes. To construct an HMM, the method partitions the genome into windows of length L and produces states corresponding to each unique haplotype within a window. Rather than emitting a single allele, states emit the entire corresponding haplotype: a string of alleles of length L. This provides a complete model of the linkage disequilibrium (LD) within each window and improves runtime by only modeling state transitions at window boundaries. Our iterative method works by first initializing haplotypes randomly, then constructing an HMM over a short window of 4 SNPs. Subsequent iterations operate over increasingly larger windows (8, 15, 21, 33, and 39 SNPs), enabling increasingly longer modeling of LD. When analyzing the 16,353 WTCCC+HapMap samples, our method runs 9.24 times faster than Beagle, with a switch error rate of 3.53% compared to Beagle's 3.10%. Additionally we found that while our runtime is super-linear, it scales better than Beagle for which runtime over 16,353 samples is 130 times slower than for 1088 samples, while our comparable slowdown is only 51.5 times. Our method's speed and scalability will enable improved accuracy by leveraging very large sample sets that will soon be available.

93

Sequencing Genes in-silico Using Single Nucleotide Polymorphisms from Genome-Wide Association Studies. *X.C. Zhang[1], B. Zhang[1], S.S. Li[1], X. Huang[1], J.A. Hansen[2,3], L.P. Zhao[1].* 1) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, Seattle, WA 98109, USA; 2) Division of Clinical Research, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, Seattle, WA 98109, USA; 3) School of Medicine, University of Washington, 1959 NE Pacific Street, Seattle, WA 98195, USA.

The advent of high throughput sequencing technology has enabled the 1000 Genomes Project Pilot 3 to generate complete sequence data for more than 906 genes and 8,140 exons representing 697 subjects. The 1000 Genomes database provides a critical opportunity for further interpreting disease associations with single nucleotide polymorphisms (SNPs) discovered from genome-wide association studies (GWAS). Currently, direct sequencing candidate genes or regions on a large number of subjects remains both cost and time prohibitive. To accelerate the transition from discovery to functional studies, we propose an in-silico gene sequencing method, which predicts phased sequences of intragenic regions, using SNPs available from GWAS or imputable using the HapMap project. The key underlying idea of our method is to infer diploidic sequence-based multiallelic genotypes at every functional locus utilizing the deep sequencing data from the 1000 Genomes Project, and to build prediction models using flanking SNPs. Using this method, we have developed a database of predictive models for 611 known genes. Sequence prediction accuracy for these genes is 96.26% on average (ranges 79% - 100%). This database of predictive models can be enhanced and scaled up to include new genes as the 1000 Genomes Project sequences additional genes on additional individuals. Applying our predictive model for the KCNJ11 gene to the Wellcome Trust Case Control Consortium (WTCCC) T2D cohort, we demonstrate how the prediction of phased sequences inferred from GWAS SNP genotype data can be used facilitate interpretation and identify probable functional mechanism such as protein changes.

94

The Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging: Ethnic Diversity, Genetic Structure, Family Relatedness and Power of a GWAS in a Cohort of 100,000. *N. Risch[1,2], M. Kvale[1], T. Hoffmann[1], S. Hesselson[1], B. Dispensa[1], S. Rowell[2], L. Walter[2], C. Somkin[2], S. VandenEeden[2], C. Quesenberry[2], L. Croen[2], L. Kushi[2], R. Whitmer[2], C. Iribarren[2], M. Sadler[2], D. Ranatunga[2], L. Shen[2], M. Henderson[2], D. Smethurst[2], D. Ludwig[2], S. Sciortino[2], D. Olberg[2], A. Finn[3], P.-Y. Kwok[1], C. Schaefer[2].* 1) Inst Human Gen, Univ California, San Francisco, San Francisco, CA; 2) Kaiser Permanente Northern California Division of Research, Oakland, CA; 3) Affymetrix, Inc, Santa Clara, CA.

The Kaiser Permanente Northern California Research Program on Genes, Environment and Health (RPGEH) has as a goal the construction of a cohort of 500,000 individuals with extensive electronic health, environmental exposure and behavioral data, along with biospecimens, to enable large scale genetic epidemiology research. To facilitate the genomics aspect of the research, in 2009 we received an ARRA grand opportunity award to perform genome-wide genotyping and telomere length analysis of 100,000 RPGEH participants, which we have now successfully completed. The cohort is highly diverse in ethnicity, including 1/4 minority that encompasses substantial numbers of individuals with African, West Asian, South Asian, East Asian, Pacific Island and Native American ancestry. ARRA conditions allowed for 14 months for genotyping, for which we chose the high-throughput Affymetrix Axiom system. To accommodate the ancestral diversity in our sample, we developed 4 novel ethnic-specific arrays that contained between 675K SNPs (for European ancestry) and 894K SNPs (for African ancestry), designed to provide coverage of common and low frequency variants in the various groups. To enable GWAS and other genomic analyses, preliminary analyses of the genotype data include extensive QC, characterization of population structure, identification of related individuals, imputation of SNPs not on the array, and identification of CNVs. Preliminary results indicate between 10,000 and 20,000 first, second and third degree relative pairs in this sample, and extensive genetic stratification due to ancestral diversity and ancestry-related assortative mating. Inclusion of family relationships in analysis allows for the simultaneous estimation of individual SNP contributions and overall heritability. The ethnic diversity allows for characterization and dissection of ethnicity-related health disparities. The Axiom arrays can easily detect large CNVs, which we have identified in a substantial number of individuals. Preliminary analysis of the microarray SNP contents suggests the ability to impute many millions of SNPs with high reliability, down to modest allele frequencies, provided adequate reference samples are available. In summary, the genomics brought to this large and unique resource is providing unprecedented opportunities to unravel the genetic and environmental architecture underlying complex longitudinal aspects of age-related disease and the processes of aging.

95

Genome-wide study of autoimmune hypothyroidism using existing genomic data and electronic medical records. *D.C. Crawford[1,2], J.C. Denny[3,4], M.D. Ritchie[1,2], M.A. Basford[5], Y. Bradford[1], H.S. Chai[6], R.L. Zuvich[1], L. Bastarache[3], P. Peissig[7], D. Carrell[8], J. Pathak[6], R.A. Wilke[4], L. Rasmussen[6], X. Wang[5], J.A. Pacheco[12], A. Kho[9], N. Weston[8], M. Matsumoto[6], K.M. Newton[7], R. Li[10], I.J. Kullo[11], C. Chute[6], R.L. Chisholm[12], E.B. Larson[8], C.A. McCarty[13], D.R. Masys[3], D.M. Roden[4, 14], S.J. Bielinski[15], M. de Andrade[6].* 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN; 3) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 4) Department of Medicine, Vanderbilt University, Nashville, TN; 5) Office of Research, Vanderbilt University, Nashville, TN; 6) Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 7) Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield WI; 8) Group Health Research Institute, Seattle, WA; 9) Department of Medicine, Northwestern University, Chicago, IL; 10) Office of Population Genomics, National Human Genome Research Institute, Bethesda, MD; 11) Division of Cardiovascular Diseases, Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 12) Center for Genetic Medicine, Northwestern University, Chicago, IL; 13) Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfiled, WI; 14) Department of Pharmacology, Vanderbilt University, Nashville, TN; 15) Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, MN.

Autoimmune hypothyroidism (AH) is the most common thyroid disease, though its genetic risk factors are unknown. Large-scale DNA databanks linked to electronic medical records (EMRs) may enable “reuse” of existing genomic and clinical data to discover and replicate genotype-phenotype associations. The Electronic Medical Records and Genomics (eMERGE) Network, established in 2007, is a collaboration of five EMR-linked biobanks and includes Group Health Cooperative/University of Washington, Marshfield Clinic, Mayo Clinic, Northwestern University, and Vanderbilt University. The eMERGE Network uses phenotype data derived from EMRs to perform genomic association studies. For each participating study site, GWAS using Illumina BeadChips (660W-Quad or 1M) were conducted for dementia, cataracts, peripheral arterial disease, type 2 diabetes, and cardiac conduction. Because these ~17,000 DNA samples are linked to EMRs, the network is able to “reuse” the genotype data to conduct network-wide GWAS for additional diseases and traits such as AH. Electronic selection logic using billing codes, laboratory values, text queries, and medication records was used to identify cases and controls at each site, and validated by partial manual review. The electronic selection logic identified cases and controls with positive predictive values of 92.4% and 98.5%, respectively. Single SNP tests of association using logistic regression and assuming an additive genetic model were performed. Analyses were performed matched and unmatched, and models were adjusted for birth decade, sex, and site of ascertainment (± principal components). Among European American cases (n=1,317) and controls (n=5,053), common variants at 9q22 near thyroid transcription factor 2 (*FOXE1*) were associated with AH at genome-wide significance (rs7850258, odds ratio 0.74, p=3.96x10-9). We sought replication in an independent dataset of European Americans, and this locus was statistically significant with 263 cases and 1616 controls (odds ratio 0.60, p=1.15x10-6). This is the first report of a genome-wide significant association for AH. Overall, the electronic selection logic was portable among five institutions with five different EMR systems. Future use of EMR-linked genomic data may allow discovery of new genotype-phenotype associations with little or no additional genotyping costs.

96

Epidemiologic Architecture for Genes Linked to Environment: serum vitamins A and E modify HDL-C, LDL-C, and triglyceride GWAS-identified associations in the National Health and Nutrition Examination Surveys. *L. Dumitrescu[1], R. Goodloe[1], K. Brown-Gentry[1], D.C. Crawford[1,2].* 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN 37232; 2) Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN 37232.

Despite many successes, lipid-associated variants discovered through genome-wide association studies (GWAS) do not account for the majority of heritability estimated for these traits. Epidemiological studies have long indicated that certain environmental factors are capable of shaping lipid distributions in the population. However, environmental modifiers of known genotype-phenotype associations are just recently emerging in the literature. We, as part of the Population Architecture using Genomics and Epidemiology (PAGE), have genotyped GWAS-identified variants in samples collected by the Centers for Disease Control and Prevention for the National Health and Nutrition Examination Surveys (NHANES). NHANES is a cross-sectional survey of Americans representing three major groups: non-Hispanic ***w***hites (n=2,435), non-Hispanic ***b***lacks (n=1,407), and ***M***exican ***A***mericans (n=1,734). Genetic NHANES contains DNA samples from three surveys: NHANES III, 1999-2000, and 2001-2002. Along with lipid levels, NHANES contains an abundance of environmental variables, including serum vitamin A and E levels, both of which are antioxidants that may play a role in lipid metabolism. Using linear regression assuming an additive genetic model, gene-environment interactions were modeled with a multiplicative interaction action term between either vitamin A or ln(vitamin E) and 23 GWAS-identified lipid-associated variants for HDL-C, LDL-C, and ln(TG) levels. In models adjusted for age, sex, and marginal effects, three SNPxvitamin A and six SNPxvitamin E interactions were identified at a significance threshold of $p<2.2x10^{-3}$. A majority of the significant interactions were associated with ln(TG) (4/9) and were among W (6/9). The most significant interaction was APOB rs693xvitamin E ($p=8.9x10^{-7}$) for LDL-C levels among MA, and this same interaction was significant in W ($p=2.67x10^{-4}$) but not B (p=0.11). Other interactions with this APOB variant (rs693xvitamin A and rs693xvitamin E) were significantly associated with ln(TG) levels among W at $p=2.2x10^{-3}$ and $4.7x10^{-5}$, respectively. The nine significant interaction models individually explained 0.35-1.28% of the variation in one of the lipid traits. Our results suggest that the antioxidants vitamins A and E impact GWAS-identified associations for lipid traits; however, these significant interactions account for only a fraction of the overall variability observed for HDL-C, LDL-C, and TG levels in the general population.

97

A genome-wide association study meta-analysis reveals multiple loci implicated in sex hormone regulation. *J.R.B. Perry[1,2], R. Haring[3], A. Koster[4], V. Lagou[2,18], L. Lyytikäinen[5], M. Mangino[6], A. Petersen[7], L. Stolk[8,17], D. Vaidya[9], L. Vandenput[10], V. Aalto[11], M.F. Wellons[12], C. He[13], N.C. Onland-Moret[14], C. Ohlsson[15], K. Lunetta[16] on behalf of the CHARGE+ Sex Hormone Consortium.* 1) Genetics of Complex Traits, Peninsula Medical School, University of Exeter, UK; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 3) Institute of Clinical Chemistry and Laboratory Medicine, University of Greifswald, Germany; 4) Laboratory for Epidemiology, Demography and Biometry, National Institute on Aging, Bethesda, Maryland; 5) Department of Clinical Chemistry, University of Tampere and Tampere University Hospital, Tampere, Finland; 6) Department of Twin Research, Kings College London, UK; 7) Institute of Genetic Epidemiology, Helmholtz Zentrum Munich, Neuherberg, Germany; 8) Dept of Internal Medicine, Erasmus MC Rotterdam, the Netherlands; 9) Department of Medicine, John Hopkins University, USA; 10) Center for Bone and Arthritis Research, Department of Internal Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; 11) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Finland; 12) Department of Obstetrics and Gynecology, Division of Reproductive Endocrinology and Infertility, The University of Alabama at Birmingham, USA; 13) Department of Public Health, Indiana University School of Medicine, Indiana, USA; 14) Julius Center for Health Sciences and Primary Care, UMC Utrecht, The Netherlands; 15) Centre for Bone and Arthritis Research, Department of Internal Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; 16) Boston University School of Public Health Department of Biostatistics, Boston, MA, USA; 17) Netherlands Consortium of Healthy Ageing, Rotterdam, the Netherlands; 18) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK.

Sex Hormone Binding Globulin (SHBG) is a glycoprotein responsible for transporting sex hormones to target tissues. Produced mostly by the liver, circulating SHBG is thought to regulate the biologically available fraction of estrogens and androgens. The possible function of this protein outside of hormone regulation is debated, but recent evidence suggests that it may have a causal role in type 2 diabetes pathophysiology. SHBG levels remain relatively stable over time in individuals, yet there is significant variation between individuals. Twin studies have suggested a heritable component to this variability, although to date only a single common variant at the SHBG gene locus has been robustly associated with SHBG protein levels. As part of the CHARGE+ consortium, we designed a large-scale GWAS meta-analysis to reveal the genetic architecture of this trait. We performed a 12 cohort GWAS meta-analysis for variants associated with SHBG, which encompassed 22,737 individuals (9809 females, 12,928 males). Replication was performed in an additional set of up to 5274 individuals (1840 females, 3434 males) across four cohorts. A total of 16 independent signals reached genome-wide significance ($P < 5x10^{-8}$), highlighting genes implicated in a variety of biological processes. The highest ranking signal was at the SHBG gene locus, rs12150660 ($P = 1x10^{-110}$). Conditional analysis revealed evidence for an additional five independent signals in this gene region, including a low-frequency missense variant with an almost 1 SD effect on levels (rs6258, $P = 1x10^{-59}$). Other implicated gene regions include SLCO1B1 (previously linked to response to statin treatment) and GCKR (linked to a variety of metabolic traits including type 2 diabetes). We also noted significant differences in effect estimates between sexes for some of these signals, notably at 4q13.2 (Male $P = 2x10^{-8}$, Female P = 0.56, Heterogeneity P = 0.008) and Xq22.3 (Male $P = 2x10^{-20}$, female P = 0.05, Het P = 0.006). The variants identified explain approximately 14% of the heritable component of this trait. This study represents the largest GWAS performed on sex hormone factors, highlighting a diverse range of pathways involved in SHBG regulation. It also demonstrates the importance of considering sex-specific effects, and secondary signals in gene regions.

98

MultiPhen: Joint model of multiple phenotypes increases discovery in GWAS. *P.F. O'Reilly[1], C.J. Hoggart[1], Y. Pomyen[2], P. Elliott[1], M.R. Jarvelin[1], L.J. Coin[1].* 1) Epidemiology & Biostatistics, Imperial College London, London, United Kingdom; 2) Dept of Cancer and Surgery, Imperial College London, London, United Kingdom.

The genome-wide association study (GWAS) approach has discovered hundreds of genetic variants associated with diseases and quantitative traits. However, despite clinical and statistical association between many phenotypes, GWAS are generally performed one-phenotype-at-a-time. Here we introduce a new method and software, MultiMap, which models multiple phenotypes simultaneously in a fast and interpretable way. MultiMap tests the linear combination of the phenotypes most associated with genotype at each SNP, thus potentially capturing effects hidden to single phenotype GWAS. We demonstrate via simulation that this approach provides a dramatic increase in power in many scenarios. There is a boost in power for both pleiotropic variants and those that affect only one phenotype. We applied MultiMap to lipid traits in the Northern Finland Birth Cohort 1966 (NFBC1966). In these data MultiMap discovers 23% more SNPs with known associations than the standard GWAS approach, while applying MultiMap in addition to the standard approach provides 39% increased discovery.

99

Type 2 Diabetes risk alleles show extreme directional differentiation among human populations, compared to hundreds of other diseases. *R. Chen[1,4], E. Corona[1,2,4], M. Sikora[3], J.T. Dudley[1,2,4], A. Moreno-Estrada[3], A.A. Morgan[1,2,4], G.B. Nilsen[5], S.E. Lincoln[5], C.D. Bustamante[3], A.J. Butte[1,2,4].* 1) Division of Systems Medicine, Department of Pediatrics, Stanford University, CA; 2) Program in Biomedical Informatics, Stanford University, CA; 3) Department of Genetics, Stanford University, CA; 4) Lucile Packard Children's Hospital, 725 Welch Road, Palo Alto, CA; 5) Complete Genomics, Inc., 2071 Stierlin Court, Mountain View, CA.

Understanding the role human population substructure plays in patterning the distribution of genetic variation relevant to complex disease is critical to both population and medical genetics. We curated multi-ethnic disease association data from 4,106 human genetic studies representing 770 diseases, recording the precise risk alleles and their measured population frequencies, as well as statistical metrics of association and estimated effect size for disease-SNP associations. Using this database, called VariMed, we compared the distribution of risk allele frequencies and predicted genetic risk across 1,397 individuals from 11 HapMap populations, 1,064 individuals from 53 HGDP-CEPH populations, and 49 individuals with high-depth whole genome sequences drawn from 10 distinct populations. We find strikingly similar patterns of population differentiation across Type 2 diabetes (T2D) susceptibility risk alleles, which are shown to be statistically significant comparing to null distributions of frequency-matched European genomic alleles and other disease risk alleles. Specifically, 12 T2D risk alleles showed significantly higher frequencies in populations from Sub-Saharan Africa and lower frequencies in East Asia populations. Furthermore, T2D shows the most significant differentiation in individual risk across population groups compared to 48 other diseases having 3 or more independent risk loci. Our results suggest that these differences in risk associated allele frequencies may play a role in the observed disparity of T2D incidence rates across ethnic groups.

100

Demographic histories of African hunting-gathering populations inferred from genome-wide SNP variation. *S. Soi[1], L. Scheinfeldt[1], C. Lambert[2], J. Hirbo[1], A. Ranciaro[1], S. Thompson[1], J.M. Bodo[3], A. Froment[4], M. Ibrahim[5], A. Juma[5], T. Nyambo[6], S. Omar[7], C. Wambebe[8], D. Meskel[9], G. Belay[9], S.A. Tishkoff[1].* 1) Genetics and Biology, University of Pennsylvania, Philadelphia, PA; 2) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; 3) Ministère de la Recherche Scientifique et de l'Innovation, BP 1457, Yaoundé, Cameroon; 4) UMR 208, IRD-MNHN, Musée de l'Homme, 75116 Paris, France; 5) Department of Molecular Biology, Institute of Endemic Diseases, University of Khartoum, 15-13 Khartoum, Sudan; 6) Department of Biochemistry, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania; 7) Kenya Medical Research Institute, Center for Biotechnology Research and Development, 54840-00200 Nairobi, Kenya; 8) International Biomedical Research in Africa, Abuja, Nigeria; 9) Department of Biology, Addis Ababa University, Addis Ababa, Ethiopia.

Africa is the geographic origin of anatomically modern humans; it is also home to a third of all modern languages, including four major language families: Niger-Kordofanian, Afro-Asiatic, Nilo-Saharan, and Khoesan. Despite the importance of African populations for studying human origins and the complexity of demographic and linguistic relationships among African populations, genome-wide analyses of sub-Saharan variation have been sparse. To address this deficiency, we used Illumina 1M-Duo SNP arrays to genotype samples (N=697) from 44 sub-Saharan populations, which we supplemented with published data sets. Principal components analysis (PCA) and linear regression were used to assess the statistical effect of geography and linguistics on the partitioning of genetic variation. As ascertainment bias can distort the allele frequency spectrum, we examined patterns of linkage disequilibrium (LD), haplotype sharing, and identity by descent (IBD) to understand the demographic relationship among populations. To affirm that LD-based analyses were robust to ascertainment bias, we assessed the rank correlation of estimates of effective population size from the rate of LD decay within populations and estimates of population size based on the variance of microsatellite repeat lengths from previously published data (Spearman's $\hat{\rho}$=0.782, p=0.011). Additionally, the presence of long IBD tracts between individuals indicates recent common ancestry. Thus, we used the GERMLINE algorithm to infer IBD tracts between individuals in hunting-gathering populations and neighboring agriculturalist and pastoralist populations. To infer the time to most recent common ancestor and test demographic models while accounting for the confounding effects of migration and changes in population sizes, we employed Approximate Bayesian Computation (ABC) using summaries of haplotype frequency, diversity and sharing within and between populations. We report, for the first time, evidence for recent common ancestry of Ethiopian hunter-gatherers and the Kenyan Sanye/Dahalo, who speak a language with remnant clicks, with click-speaking eastern African Khoesan populations. This work supports archaeological and linguistic studies that indicate that the distribution of Khoesan speaking populations may have extended as far north as Ethiopia.

101

Genome-wide comparison of African-ancestry populations from CARe and other cohorts reveals signals of natural selection. *G. Bhatia[1], N. Patterson[2], B. Pasaniuc[3], N. Zaitlen[3], G. Genovese[4], S. Pollack[3], S. Mallick[5], S. Myers[6], A. Tandon[5], C. Spencer[7], C.D. Palmer[8], A.A. Adeyemo[9], M. Fornage[10], W.H.L. Kao[11], A. Ogunniyi[12], G. Papanicolaou[13], C.N. Rotimi[14], J.I. Rotter[15], B. Salako[12], B.O. Tayo[16], S. McCarroll[5], P. Sabeti[17], G. Lettre[18], P.D. Jager[19], J. Hirschhorn[8], X. Zhu[20], R. Cooper[16], D. Reich[5], J.G. Wilson[21], A.L. Price[3] for the CARe Analysis Core.* 1) Harvard-MIT Division of Health, Science and Technology; 2) Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, USA; 3) Epidemiology, Harvard School of Public Health, Boston, MA; 4) Division of Nephrology, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA 02215, USA; 5) Department of Genetics, Harvard Medical School, Boston, MA, USA; 6) Department of Statistics, University of Oxford, Oxford, United Kingdom; 7) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 8) Divisions of Genetics and Endocrinology and Program in Genomics, Children's Hospital Boston, Boston, MA, USA; 9) NIH Intramural Center for Research on Genomics and Global Health, National Human Genome Research Institute, Bethesda, MD, USA; 10) Institute of Molecular Medicine and Division of Epidemiology School of Public Health, University of Texas Health Sciences Center at Houston, Houston, TX, 77030, USA; 11) Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland, United States of America; 12) Department of Medicine, University of Ibadan, Ibadan, Nigeria; 13) National Heart, Lung, and Blood Institute (NHLBI), Division of Cardiovascular Sciences, NIH, Bethesda, MD 20892, USA; 14) Center for Research on Genomics and Global Health, National Human Genome Research Institute, Bethesda, MD, USA; 15) Cedars-Sinai Medical Center, Medical Genetics Institute, Los Angeles, CA, USA; 16) Department of Preventive Medicine and Epidemiology, Loyola University Chicago Stritch School of Medicine, Maywood, IL, USA; 17) Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA; 18) Département de Médecine, Université de Montréal, C.P. 6128, succursale CentrePville, Montréal, Québec, Canada; 19) Division of Molecular Immunology, Center for Neurologic Diseases, Brigham and Women's Hospital and Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115; 20) Department of Epidemiology and Biostatistics, School of Medicine, Case Western Reserve University, Cleveland, USA; 21) University of Mississippi Medical Center, Jackson, Mississippi, USA.

The study of recent natural selection in humans has important applications to human population history and medicine. Signals of selection fit into three main categories: unusually long, recent haplotypes; deviations from the expected allele frequency spectrum; and unusual population differentiation. Signals of the first two types are only expected in the case that a novel or very rare variant is subject to selection and then “sweeps” to high frequency. If, however, selection acts on a common or “standing” variant, as has been suggested in recent studies, these tests would be unlikely to uncover a signal. Therefore, a key advantage of the unusual population differentiation approach is the ability to detect selection on standing variation. Studies of population differentiation to detect selection are maximally powered when comparing closely related populations that have large effective population size, with data from a large number of individuals (>> $1/F_{ST}$). This approach has been applied genome-wide to populations within Europe and within East Asia, and to candidate loci of populations within Africa. Our study is the first to apply this approach genome-wide to closely related populations with majority African ancestry. These populations: African Americans, Nigerians, and Gambians, are genetically close (F_{ST} < 0.01 for all pairs) so that our large sample size (N=12,485) allows us to detect selection with moderate changes in allele frequency. To accomplish this analysis, we developed a new tree-based method, incorporating information from all 3 populations in order to increase power to detect selection and enable resolution of the population subject to selection. In addition, we have developed methods to increase power by accounting for substantial European-related admixture in the populations. Our genome-wide significant results corroborate loci previously reported to be under selection in Africans including HBB, and CD36. At the HLA locus on chromosome 6, results suggest the existence of multiple, independent targets of population-specific selective pressure. In addition, we report a novel, genome-wide significant ($P=1.36\times10^{-11}$) signal of selection in the Prostate Stem Cell Antigen (PSCA) gene. The most significantly differentiated marker in our analysis, rs2920283 is highly differentiated in both Africa and East Asia and is tightly linked to a known coding variant with prior genome-wide significant associations to bladder and gastric cancers.

102

The Evolution of *SRGAP2*: a gene specifically duplicated in the human lineage and associated with cortical development of the brain. *M.Y. Dennis[1], X. Nuttle[1], P.H. Sudmant[1], F. Antonacci[1], S. Sajjadian[1], L. Vives[1], T.A. Graves[2], R.K. Wilson[2], F. Polleux[3], E.E. Eichler[1,4].* 1) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA; 2) The Genome Institute at Washington University School of Medicine, Washington University School of Medicine, Saint Louis, MO, USA; 3) Dorris Neuroscience Center, The Scripps Research Institute, La Jolla, CA, USA; 4) Howard Hughes Medical Institute, Seattle, WA, USA.

Major changes in behavior, physiology, and anatomy occurred during the evolution of our species that ultimately contributed to the traits that make us distinctly human. An important source for phenotypic change and adaptive evolution is the deletion and duplication of genes. We recently identified 23 genes that duplicated specifically in the human lineage, many of which have been implicated in neuronal function and brain development. One gene in particular, *SRGAP2*, has recently been shown to function in human cortex development by inducing neurite outgrowth and branching and regulating neuronal migration and morphogenesis. In order to assess the functional impact and reconstruct the evolutionary history, we generated complete finished genome sequence for each of the three copies using clones from a hydatidiform mole bacterial artificial clone (BAC) library. We resolved the genomic context of each of the three duplicate copies (1.8 Mbp total), discovered 469 kbp of missing euchromatic sequence from the human reference genome (build 37) and determined that each of the three copies is fixed for copy number (n=155 sequenced human genomes). Phylogenetic and comparative sequence analyses suggest that the original *SRGAP2* gene-containing segment (240 kbp) was duplicated from the ancestral locus (1q32.1) to chromosome 1q21.1 approximately 3.4 million years ago; this event was followed by a secondary duplication (>555 kbp) creating a third copy at chromosome 1p12.1 approximately 2.4 million years ago. All three copies are predicted to encode functional proteins with multiple fixed non-synonymous changes distinguishing the duplicated genes. We show that the three copies of *SRGAP2* are expressed in both fetal and adult brain, although full-length transcripts of the duplicated copies have altered carboxy termini that may antagonize the function of the parental copy. Designing PCR assays to distinguish the various paralogs, we have begun a screen of patients with autism and intellectual disability identifying individuals with large-scale deletions and duplications that include *SRGAP2* genes--events not observed in the general population. In total, these results make this gene and its duplicates compelling candidates for understanding the evolution and cortical expansion of the human brain.

103

The remarkable evolution of ABO in primates. *L. Segurel[1], E.E. Thompson[1], J. Lovstad[1], A. Venkat[1], S. Margulis[2], J. Moyse[2], S. Ross[2], K. Gamble[2], M. Przeworski[1,3], C. Ober[1].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Lincoln Park Zoo, Chicago, IL; 3) Ecology and Evolution, University of Chicago, Chicago, IL.

The ABO blood group system was the first polymorphism discovered in humans, and remains to date one of very few examples of a long-term balanced polymorphism. Amazingly, the same two polymorphisms that determine A and B specificity in humans are present in at least nine other primate species. Whether the occurrence of A and B antigens in primates is the result of an ancient trans-species polymorphism or due to numerous instances of convergent evolution has been debated for over twenty years, with examples of trans-species polymorphism in macaques and gibbons, but data in other species, including humans, interpreted as support for convergent evolution of genotypes. Thus, although ABO is a textbook case of balancing selection, the nature of the selective pressures and the evolutionary mechanism underlying the convergent phenotypes remain unresolved. Here, we present for the first time sequencing and phenotypic data on large samples of apes that help to disentangle previous hypotheses. When focusing on the functionally important region, we identify shared polymorphisms among apes and estimate a divergence time between human A and B alleles on par with the divergence time between human B and gorilla B alleles. These data challenge the previous understanding of the evolution of ABO in primates, pointing to a trans-species polymorphism within apes - the only known example outside the MHC.

104

Genomic Reconstruction of an Extinct Population from Next-Generation Sequence Data - Insights from the Taìno Genome Project. *J.K. Byrnes[1], J.L. Rodríguez-Flores[2], A. Moreno-Estrada[1], C.R. Gignoux[3], S. Gravel[1], W. Guiblet[4], F. Zakharia[1], J. Dutil[5], E.G. Buchard[3], T.K. Oleksyk[4], J.C. Martínez-Cruzado[4], C.D. Bustamante[1], The 1000 Genomes Project Consortium.* 1) Department of Genetics, School of Medicine, Stanford University, California 94305, USA; 2) Department of Biological Statistics and Computer Science, Cornell Univeristy, New York 14853, USA; 3) Institute for Human Genetics, University of California San Francisco, California, USA; 4) Department of Biology, University of Puerto Rico at Mayagüez, Puerto Rico 00680; 5) Ponce School of Medicine, Puerto Rico 00732.

The first Native American people encountered by Europeans across the Caribbean were given the collective name “Taìnos” by the arriving Spaniards. One hundred years after this initial contact, the Taìnos were effectively extinct due to war, slavery, suicide, hunger, and disease. Today, the ancestral legacy of the Taìnos is found in traces of their genomes still present in the inhabitants of the islands. An incredible opportunity to recover these ancestral segments came when 35 Puerto Rican trios were included in the sampling plan of the 1,000 Genomes Project (TGP). The TGP is producing low-coverage, whole-genome sequence, high-coverage exome sequence for the 70 unrelated parents and genotyping chip data on all individuals in this sample. Due to the complex history of Puerto Rico, the genomes of its inhabitants show three-way admixture including ancestry from European, African, and Taìno origins. Using trio-phased OMNI array data, we applied a principal components based method to infer the ancestral origin of each locus on the genome for all individuals. Although the Taìno admixture proportion is small (0.09 ± 0.01 S.E.) relative to the African (0.13 ± 0.03 S.E.) and European (0.77 ± 0.03 S.E.) proportions, greater than 85% of the genome is covered by five or more chromosomes of Taìno origin. Looking at the ancestry tract length distribution, we can infer various aspects of the demographic history. For example, Taìno ancestry tracts follow an exponential distribution suggesting a single pulse of indigenous Taìno ancestry incorporation consistent with historical records of rapid extinction of the Taìnos. Given our ancestry inference, we can use the high-throughput sequencing data to measure heterozygosity, estimate time to most recent common ancestor between maternal and paternal lineages, and construct the site frequency spectrum in an ancestry-specific way. This provides further information on demographic history including effective population size estimates of the source populations contributing to the admixture event. Finally, we identify Taìno specific genomic variation cataloging what remains of this lost ancestral lineage. The data will be housed and annotated in an online genome browser based at the University of Puerto Rico at Mayagüez as a platform for undergraduate and graduate education in genomics and bioinformatics. This is the first known reconstruction of the genomic variation of an extinct human population using modern data.

105

Development of a Panel of Ancestry Informative Markers for Latin Americans from Genomewide Data. *J.M. Galanter[1], J.C. Fernandez[2], C.R Gignoux[1], J. Barnholtz-Sloan[3], C. Fernandez-Rozadilla[4], A. Hidalgo-Miranda[2], P. Raska[4], C. Ruiz Ponte[4], Y. Ruiz[4], P. Taboada[4], L. Porras[4], A. Salas[4], I.S. Zolezzi[2], A. Bigham[5], G. Gutierrez[6], T. Grutsaert[7], F. Leon Velarde[8], L. Moore[9], E. Vargas[10], M. Cruz[11], J. Escobedo[12], C. Bustamante[13], M. Shriver[5], E. Ziv[1], E. Gonzalez Burchard[1], R. Haille[14], E. Parra[15], A. Carracedo[4], LACE Consortium.* 1) University of California, San Francisco, San Francisco, CA; 2) Instituto Nacional de Medicina Genómica; Mexico City, DF, Mexico; 3) Case Western Reserve University School of Medicine, Cleveland, OH; 4) Universida de de Santiago de Compostela, Lugo, Spain; 5) Penn State University, Happy Valley, PA; 6) University of Colorado, Boulder, CO; 7) University of Syracuse, Syracuse, NY; 8) Universidad Peruana Cayetano Heredia, Lima, Peru; 9) Wake Forest University, Winston-Salem, NC; 10) Universidad Mayor de San Andre's, La Paz, Bolivia; 11) Centro Medico Nacional Siglo XXI, Mexico City, Mexico; 12) Hospital General Regional 1, Mexico City, Mexico; 13) Stanford University, Stanford, CA; 14) University of Southern California, Los Angeles, CA; 15) University of Toronto, Mississauga, Ontario, Canada.

Latin Americans are descendants of Native American, European, and African ancestors. Complex historical factors have resulted in varying proportions of ancestral contributions between individuals within each ethnic group and between ethnic groups. Using clustering methods, it is possible to estimate individual ancestral proportions from genotypes of SNPs whose allele frequencies differ between ancestral groups. Ancestry Informative Markers (AIMs) are polymorphisms selected in small panels to exhibit substantially different allele frequencies between populations from different geographic regions that can be used for this purpose.

In a two-stage project, we developed a panel of 400 AIMs optimized to characterize Latin American populations. In the first stage, we used genomewide data from two African populations, three European populations, and six Native American populations to select AIMs. Markers were chosen on the basis of locus specific branch length for the I_N statistic for each of the three main continental populations that are the basis of modern populations in the Americas. AIMs were selected to be informative, evenly distributed throughout the genome, capable of being genotyped on a widely available commercial platform (Sequenom), and universal, having little within-continent heterogeneity. We validated the panel of AIMs by comparing it to ancestry estimates using GWAS data. The panel of markers provided balanced discriminatory power between the three ancestral populations, and could accurately estimate ancestry of admixed populations compared to ancestry determined with genomewide data.

In the second stage, we genotyped 18 populations from throughout the Americas using these AIMs and estimated ancestry differences within and between these populations. We used the data to estimate the number of generations since admixture and found populations with high African ancestry have undergone 5 - 7 generations of admixture while populations with low African ancestry had values between 8 and 12 generations, corresponding well with the historical record. This panel is intended to be a resource for the community and we provide anonymized ancestral African, European, and Native American genotype information. We hope that investigators can use the selected panel of AIMs, which can be easily genotyped on readily available platforms, as a cost-effective tool to estimate continental ancestry in modern populations of the Americas.

106

Dating ancient admixture: the date of gene flow from Neandertals into modern Humans. *S. Sankararaman[1,2], N. Patterson[2], S. Paabo[3], D. Reich[1,2], Neandertal Genome Analysis Consortium.* 1) Department of Genetics, Harvard Medical School, Boston, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology,Leipzig.

Comparisons of DNA sequences between Neandertals and present-day humans have shown that Neandertals share more genetic variants with non-Africans than with Africans. This could be due to interbreeding between Neandertals and modern humans when the two groups met subsequent to the emergence of modern humans outside Africa. However, it could also be due to population structure that antedates the origin of Neandertal ancestors in Africa. To distinguish the African substructure scenario from the recent gene flow scenario we try to estimate a date for when Neandertals and modern humans last exchanged genes. The decay of linkage disequilibrium between pairs of SNPs shared with Neandertals in modern human populations can be informative of the time of gene flow. Estimates of such ancient events using LD-based methods are seriously biased by demographic events since gene flow and by inaccuracies in existing genetic maps.

To estimate the date of gene flow, we computed the average signed LD between pairs of SNPs separated by a range of genomic distances and measured the rate of decay of this statistic. We also construct a statistical model that permits us to estimate and correct for the uncertainty of a given genetic map. We show both in theory and using coalescent simulations that this statistic, coupled with a specific ascertainment scheme and estimates of the map uncertainty, provides accurate estimates of the date of ancient gene flow under a range of demographic scenarios and can distinguish recent gene flow from ancient population structure.

We applied our statistic to SNP data in the 1000 genomes project. We observe that the empirical extent of LD is a factor of 3.5-times greater in non-Africans than in Africans, consistent with Neandertal sharing more alleles with non-Africans than with West Africans. Applying our correction for map uncertainty and taking into account biases in our statistic, we conclude that the true date of gene flow was 37000-86000 years BP, suggesting that it occurred when modern Humans encountered Neandertals, presumably in western Eurasia.

107

Introducing array comparative genomic hybridization into routine prenatal diagnosis practice: a prospective study on 1166 consecutive clinical cases. *F. Fiorentino[1], F. Caiazzo[1], S. Napoletano[1], L. Spizzichino[1], S. Bono[1], M. Sessa[1], A. Nuccitelli[1], A. Biricik[1], A. Gordon[2], G. Rizzo[1], M. Baldi[1].* 1) "GENOMA" - Molecular Genetics Laboratory, Rome, Italy; 2) Bluegnome Ltd, Cambridge CB22 5LD, UK.

Objective: Experience with array-based comparative genomic hybridization (aCGH) use for clinical prenatal diagnosis is still relatively limited. Prospective trials on a large sample size are necessary before aCGH can be recommended for routine clinical use in prenatal diagnosis. To assess the feasibility of offering aCGH for prenatal diagnosis as a first-line test, a prospective study was performed on a cohort of 1166 consecutive prenatal samples, comparing the results achieved from aCGH with those obtained from a conventional karyotype. Methods: Women undergoing amniocentesis or chorionic villus sampling (CVS) for standard karyotype, between 1 October 2010 and 31 May 2011, were offered aCGH analysis. A total of 1166 prenatal samples were processed in parallel using both aCGH, performed on DNA isolated from amniotic fluid (88.7%) or CVS (9.5%) and cultured amniocytes (1.8%), and G-banding for standard karyotyping. Results: Clinically significant chromosome abnormalities were identified in 37(3.2%) samples, 27(73.0%) of which were also detected by conventional karyotyping. In 10(27.0%) samples, aCGH identified pathogenic copy number variations (CNVs) that would not have been found if only a standard karyotype had been performed. Six were de novo, not recorded as benign CNVs, 4 resulted inherited. Eight of the above CNVs were concerning well-established syndromes. Benign CNVs were identified in 152(13.0%) samples. Following parental studies, no findings of unclear significance remained. aCGH was also able to detect chromosomal mosaicism as lower as 10% level. There was a complete concordance between the conventional karyotyping and aCGH results, except for 2 cases, that were correctly diagnosed by aCGH. Conclusions: This study demonstrates that aCGH represents an improved diagnostic tool for prenatal detection of chromosomal abnormalities, allowing identification of submicroscopic clinically significant imbalances that are not detectable by conventional karyotyping. The results of the study indicated that the aCGH approach was robust, with no false positive or false negative findings, suggesting that the technique has the potential to replace the traditional cytogenetic analysis without missing significant results. Our findings provide a further evidence on the feasibility of introducing aCGH into routine prenatal diagnosis practice as first-line diagnostic test to detect chromosomal abnormalities in prenatal samples.

108

Validation of single cell Whole Genome Amplification (WGA) for Preimplantation Genetic Haplotyping (PGH) and its application using DNA microarrays. *G. Altarescu, H. El Harar, S. Zeligson, S. Perlberg, R. Beeri, D. Zeevi, T. Eldar-Geva, I. Varshaver, E. Margalioth, E. Levy-Lahad, P. Renbaum.* Zohar PGD Lab & IVF Unit, Shaare Zedek Medical Ctr, Jerusalem, Israel.

Background: Preimplantation genetic diagnosis (PGD) for molecular disorders involves development of disease and family specific protocols that allow simultaneous amplification of the mutation with multiple polymorphic markers in single cells. The development of PGD tests for single-gene disorders is challenging and only a limited number of amplified markers can be analyzed. An alternative and universal approach, PGH relies on whole genome amplification from a single cell to give microgram quantities of DNA, which allows the testing of multiple loci using PCR based protocols. Aims: To validate a WGA protocol for single and multiple cell amplification, including assessment of amplification and allele drop-out (ADO) rates, using both standard PCR protocols and DNA microarray chips. Methods: Single cells from embryos that reached the 6-8 cell stage, along with 2-5 cells from blastocyst biopsies were used for PGH analysis using the Qiagen RepliG multiple strand displacement amplification kit. Multiple loci on the genome from our PGD database were tested. Marker informativity was assessed using family haplotypes for each biopsied embryo for the calculation of ADO rates. In addition, WGA samples from a single cell and two-cell biopsy were analyzed using an Affymetrix microarray SNP chip. Results: Seven samples of blastocyst biopsy and nine single cell blastomeres were amplified using WGA. Fifty three loci on four different chromosomes were analyzed from the WGA samples obtained from single blastomeres and blastocyst biopsied cells. Using standard PCR protocols the amplification rate for blastocysts was 99.5% and for singles blastomeres 87.5%. ADO rates ranged between 8% for blastocyst biopsied cells to 16.3% for blastomeres. The results were in concordance to the parental haplotypes built from genomic DNA. Analysis using DNA microarrays yielded 90-92% call rates with ADO rates between 26-29%. An unbalanced translocation, confirmed by FISH could also be detected. Conclusions: WGA appears to be an efficient and accurate method for obtaining significant amounts of DNA from single cells PGH and can also be used for microarray anlysis. PCR based ADO rates are similar to those obtained by conventional PGD methods. The WGA approach incorporates information from multiple polymorphic markers and thus identifies high-risk haplotypes, creating a universal generic test for any single-gene disorder and chromosomal imbalances.

109

Are heterozygous carriers of cohesin mutations at an increased risk of aneuploidy? *B.M. Murdoch, N. Owen, S.I. Nagaoka, T.J. Hassold, P.A. Hunt.* Washington State University, Pullman, WA.

Cohesion proteins (cohesins) tether sister chromatids prior to cell division, and thereby play a critical role in chromosome segregation. Not surprisingly, mutations in genes encoding cohesin proteins have profound effects on genome stability: in somatic tissues, defects in cohesins have been linked to a number of malignancies and, in germ cells, to infertility or to an increase in aneuploid gametes. However, the germ cell studies have been based on studies of model organisms and, in particular, on mice homozygous for null mutations encoding the meiosis-specific cohesins REC8 or SMC1B. Since more subtle allelic variation - a likely scenario in the human population - has not been examined, we were interested in asking whether "milder" genotypes might also have deleterious effects on meiosis. Accordingly, we assessed the consequences of haploinsufficeny for either *Rec8* or *Smc1*B on meiotic chromosome dynamics in female mice. We generated surface spread preparations from fetal ovarian samples to examine the formation of the synaptonemal complex (SC), the incidence of synaptic errors and the level of meiotic recombination. Surprisingly, for each of these variables - and for both *Rec8* or *Smc1*B - we identified significant differences between heterozygous females and their wildtype sibling controls. For example, we found that haploinsufficency resulted in shorter and less compact SCs and an approximate two-fold increase in asynapsis and other synaptic defects. Further, in both *Rec8* and *Smc1*B heterozygotes we observed significant 5-10% decreases in the genome-wide number of MLH1 foci, a marker of meiotic crossovers. The increase in meiotic defects in prophase oocytes prompted us to examine conventional cytogenetic preparations in metaphase II (MII) eggs from adult females, asking whether chromosome abnormalities were increased in heterozygotes. Indeed, carriers of each mutation exhibited an approximate two-fold increase in chromosome errors, including both numerical and structural abnormalities. Thus, our results indicate that haploinsufficency for meiotic cohesins increases the likelihood of meiotic errors, suggesting that women who are asymptomatic carriers of a cohesin mutation may have an increased risk of a chromosomally abnormal pregnancy.

110

Human Embryos with Aneuploid Cells Documented at the Cleavage Stage Undergo Genetic Correction During Differentiation to the Blastocyst Stage. *P.R. Brezina[1], A. Benner[2], R. Ross[3], A. Barker[4], K. Richter[5], G.R. Cutting[6], W.G. Kearns[7].* 1) GYN/OB; Johns Hopkins Medical Institutions, Baltimore, MD. U.S.A; 2) Genetics, The Center for Pre-Implantation Genetics, LabCorp: Rockville, MD, U.S.A; 3) IVF Lab, La Jolla IVF; La Jolla, CA, U.S.A; 4) IFV Lab, Arizona Center for Fertility Studies, Phoenix, AZ, U.S.A; 5) Research, Shady Grove Fertility Reproductive Science Center, Rockville, MD. U.S.A; 6) The McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins Medical Institutions, Baltimore, MD, U.S.A; 7) Genetics and GYN/OB, The Center for Preimplantation Genetics Labcorp and Johns Hopkins Medical Institutions, Rockville and Baltimore, MD.

Background: We determined if day 3 aneuploid embryos can undergo genetic normalization during differentiation to the blastocyst stage. **Materials and Methods:** Patients underwent in vitro fertilization (IVF) and preimplantation genetic screening (PGS) due to repeat pregnancy loss (RPL). Single blastomeres taken from Day-3 embryos underwent two rounds of DNA amplification followed by a 23-chromosome SNP microarray analysis using HumanCytoSNP-12 DNA beadchips and GenomeStudio and KaryoStudio software. Following Day-3 biopsy, all embryos remained in a standard commercially available media until Day-5 post oocyte fertilization. Day-5 blastocysts with euploid Day-3 results either underwent uterine transfer or were cryopreserved. Day-5 blastocysts grown from embryos with aneuploid Day-3 results underwent surgery to separate the inner cell mass (ICM) from the trophectoderm (TE). An average of 100 TE cells and a range of 40 ICM cells to the entire ICM cell population were obtained from each embryo and immunocytochemistry was performed using anti-oct3/4 to confirm the ICM and anti-cdx2 to identify the TE cells. DNA amplification and microarray analyses were then performed as described above. Binomial confidence intervals for proportions were calculated. **Results:** 12 patients were enrolled. 126 embryos were subjected to Day-3 biopsy. Of cleavage stage embryos with euploid karyotypes, 43 (69.4%) developed to the blastocyst stage. In contrast, only 25 (39.1%) of the embryos with an aneuploid karyotype at the cleavage stage progressed to the blastocyst stage. A euploid karyotype obtained at the cleavage stage was predictive of progression to the blastocyst stage with a positive predictive value of 69.4% [95% CI: 57-79%] and a negative predictive value of 60.9% [95% CI: 49-72%]. Evaluation of the 25 blastocysts that developed from aneuploid day 3 embryos revealed that 68% (17/25) [95% CI: 48-83%] possessed a euploid ICM and 76% (19/25) [95% CI: 56-89%] possessed a euploid TE with 64% (16/25) [95% CI: 44-80%] having both a euploid ICM and TE. No mosaicism, detected at a level of 5%, was observed. **Conclusions:** The genetic normalization observed in this study has significant implications in numerous scientific fields. An understanding of such in vitro reparative mechanisms could further gene repair and stem cell transplant therapy and could impact the management of patients undergoing infertility care by reassessing the disposition of abnormal Day-3 aneuploid embryos.

111

Mosaicism do not affect accuracy of 24 chromosomes preimplantation genetic screening on cleavage stage embryos. *A. Biricik[1], F. Fiorentino[1], G. Kokkali[2], L. Rienzi[3], L. Spizzichino[1], S. Bono[1], A. Gordon[4], F.M. Ubaldi[3], K. Pantos[2].* 1) "GENOMA" - Molecular Genetics Laboratory, Rome, Italy; 2) Centre for Human Reproduction, Genesis Athens Hospital, Athens, Greece; 3) G.EN.E.R.A Centre for Reproductive Medicine, Clinica Valle Giulia, Rome, Italy; 4) Bluegnome Ltd, Cambridge CB22 5LD, UK.

Introduction: Different randomized clinical trials (RCTs) have shown that preimplantation genetic screening (PGS), as it was practiced, has not provided the expected benefits. The possible explanation for this poor clinical performance has been mainly attributed to the embryonic chromosomal mosaicism, that is present on day-3 of development, which may decrease the chances of a live birth by prematurely labeling an embryo as abnormal. In this study we aimed to evaluate the accuracy of the 24-chromosomes PGS performed on cleavage stage embryos, in order to ascertain if the tested blastomeres were representative for the whole embryo. Methods: embryos biopsy was carried out at day-3. Single cell DNA was amplified by whole genome amplification (WGA) and processed by Array-CGH according to the 24sure protocol, BlueGnome. Euploid embryos were then selected for transfer on day-5 or day-6 of the same cycle. In order to verify the results obtained following day-3 PGS, chromosomally abnormal embryos that developed to blastocyst stage were re-biopsied on day-6 and reanalyzed. Results: 111 PGS cycles were carried out for 104 couples. The mean maternal age was 39.0±3.7 years. A total of 838 embryos were biopsied on day 3. Overall, 768 (91.6%) embryos were successfully diagnosed, 553 (72.0%) of which resulted aneuploid. Embryos suitable for transfer where identified in 75 cycles (67.6%). Following transfer of 124 embryos, 50 women (mean maternal age 38.1±3.2 years) had a clinical pregnancy (66.7% pregnancy rate/ET). A total of 63 embryos implanted (50.8% implantation rate/ET), for 60 of which heart beat was also detected. After the clinical cases, 218 non-transferred embryos from 64 PGS cycles were successfully reanalyzed. Aneuploidy mosaicism was detected in 100/218 (45.9%) embryos. Despite high levels of mosaicism found, all day-3 aneuploid embryos followed-up were again diagnosed as abnormal after re-analysis on trofectoderm cells, confirming at the end the previous results regardless of the actual abnormal genotype. Conclusions: Post-zygotic errors leading to mosaicism were common. However, mosaic embryos were confirmed as chromosomally abnormal after re-analysis at blastocyst stage. Although a larger follow-up study are required in order to confirm the above findings, array-CGH analysis on single blastomeres has demonstrated an accurate aneuploidy detection tool and may assist in identifying abnormal embryos at cleavage stage.

112

The transcriptome of a human polar body accurately reflects its sibling oocyte. *A. Reich[1], P. Klatsky[2], S. Carson[2], G. Wessel[1].* 1) Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI 02912, USA; 2) Division of Reproductive Endocrinology and Infertility, Women and Infants Hospital, Alpert School of Medicine, Brown University, Providence, RI 02905, USA.

Improved methods are needed to reliably and accurately evaluate the developmental potential of human embryos created through in vitro fertilization (IVF). Currently embryos are prioritized by morphology, an imperfect criterion, and this inaccuracy necessitates the creation and transfer of multiple embryos to produce a successful live birth. The first polar body (PB) is extruded from the oocyte before fertilization and can be biopsied without damaging the oocyte and because much of early development is driven by maternally deposited mRNA in the oocyte, we tested the hypothesis that the PB transcriptome is representative of that of the oocyte. Polar body biopsy was performed on metaphase II (MII) oocytes followed by single-cell transcriptome analysis of the oocyte and its sibling PB. Over 12,700 unique mRNAs and miRNAs from the oocyte samples were compared to the 5,431 mRNAs recovered from the sibling PBs (5,256 shared mRNAs or 97%, including miRNAs). Analysis of the mRNA expression levels of transcripts expressed in oocytes and PBs reveals that no genes are significantly differentially expressed between the two populations, indicating that the human PB transcriptome reflects that of the oocyte in both the expressed genes as well as the abundance levels of those gene products. This analysis was able to detect up to a four orders of magnitude difference in transcript abundance between individually sequenced oocytes; the same order of magnitude difference was also detected in the individually sequenced sibling PBs. These results suggest that this single cell analysis is accurate and sensitive over the wide dynamic range of mRNA transcript abundance that is generally found within transcriptomes. Furthermore we identified a conservative list of 215 genes that could be used as potential candidates for future clinical applications. Refinement of this approach could lead to the first molecular diagnostic of gene expression in MII oocytes, possibly allowing for both oocyte ranking and embryo preferences in IVF.

113

Fetal brain-specific transcripts are reproducibly found in amniotic fluid. *L. Hui[1, 2, 4], K.L. Johnson[1, 2], D.K. Slonim[3], D.W. Bianchi[1,2].* 1) Dept of Pediatrics, Division of Genetics, Tufts Medical Center, Boston, MA., USA; 2) Mother Infant Research Institute, Tufts Medical Center, Boston, MA., USA; 3) Dept of Computer Science, Tufts University, Medford, MA., USA; 4) Discipline of Obstetrics, Gynaecology, and Neonatology, University of Sydney, Sydney, NSW, Australia.

Background: In prior work we developed the amniotic fluid core transcriptome (AFCT) by mining data from cell-free fetal (cff) RNA in 12 euploid midtrimester amniotic fluid (AF) supernatants (Gene Expression Omnibus datasets GSE16176 and GSE25634). The AFCT was defined as the 470 well-annotated genes present in all 12 samples ($P < 0.04$). Preliminary analyses suggest that the AFCT derives from multiple tissues. **Aim:** To explore whether fetal transcripts associated with nervous system development are present in the AFCT. **Methods:** We examined the AFCT for nervous system transcripts using 3 resources: (i) Ingenuity Pathway Analysis™ (IPA); (ii) the GNF Gene Expression Atlas (we defined genes as highly nervous system-specific if the gene had an expression value > 30 MoM for a nervous system structure, and no unrelated tissue expression > 10 MoM); and (iii) manual examination of the AFCT gene list for relevant Entrez Gene names. The functions of the identified genes were then assessed using Entrez Gene and UniProtKB knowledge bases.**Results:** Of the 4129 total nervous system-associated genes in the IPA gene list, 59 were present in the AFCT. A further 6 candidate genes were identified by their GNF Gene Expression Atlas profiles, and 4 genes were identified by their Entrez Gene names alone. Of the 69 total candidate genes, 19 had specific roles in nervous system development or function. Genes in the AFCT with highly fetal brain-specific expression patterns and known roles in brain function included neuron navigator 2 (*NAV2*), calcium/calmodulin-dependent protein kinase II inhibitor 1 (*CAMK2N1*), brain neuron cytoplasmic protein 1 (*D4S234E*), zinc finger protein 238 (*ZNF238*), and cysteine and glycine-rich protein 2 (*CSRP2*). Synaptobrevin 2 (*VAMP2*), a human brain biomarker previously identified using RNA expression analysis, was identified in the AFCT from the IPA gene list. The AFCT also contained transcripts expressed by extracranial neural structures such as spinal cord, dorsal root and superior cervical ganglia, and the atrioventricular node. **Conclusion:** Gene transcripts that are highly expressed in the normal developing nervous system are consistently detectable in midtrimester AF supernatant. AF cff RNA may be a source of potential biomarkers of nervous system development in live ongoing pregnancies. Our findings may have many future applications in the study of congenital disorders of the nervous system using material that is routinely discarded.

114

The effects of advanced paternal age on genetic risks are mediated through dysregulation of HRAS signalling in the testis. *A. Goriely[1], S.J. McGowan[1], S. Pfeifer[2], A. Itani[3], G.A.T McVean[2], A.O.M. Wilkie[1].* 1) WIMM, University of Oxford, Oxford, OX39DS, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 3) Institute of Reproductive Sciences, University of Oxford, Oxford, UK.

Although advanced paternal age is a well known genetic risk factor, the mechanisms that mediate this effect are poorly understood. A small group of disorders, including Apert syndrome (FGFR2 mutations), achondroplasia and thanatophoric dysplasia (FGFR3), and Costello syndrome (HRAS), that we collectively call `paternal age-effect' (PAE) disorders, provides a good model to study the biological basis of the PAEs. We previously analysed spontaneous FGFR2/3-associated mutations directly in human sperm and showed that they accumulate in the sperm of ageing men. We also identified somatic FGFR3 and HRAS mutations in testicular tumours. The combined evidence suggests that PAE mutations, although occurring rarely, provide a selective advantage to spermatogonial stem cells (SSC), resulting in their clonal expansion and relative enrichment in sperm over time, accounting for the PAE observed in the associated syndromes. To further characterise the mechanisms associated with the PAE, we have quantified the Costello syndrome mutations at codon G12 of HRAS directly in 91 human sperm samples, using massively parallel sequencing. These mutations were detected at significant levels in the sperm of most men and increased significantly with the donor's age. The relative prevalence of different mutations was strikingly different to that observed in most cancers. We found that G12S (34G>A at a CpG dinucleotide) is the most common HRAS mutation in sperm (average ~1:45,000), followed by G12D (35G>A) (~1:125,000), while G12C (34G>T) and the strongly activating G12V (35G>T) mutations were observed at lower levels (~1:300,000-1:400,000). Unexpectedly, we also identified many instances of tandem nucleotide substitutions in sperm, suggesting an unusual mechanism of mutagenesis within the testis. The relative abundance of HRAS mutations observed in sperm correlates well with the documented population prevalence of Costello alleles. These results suggest that the PAE is mediated through activation of the growth factor receptor-RAS pathway, which is a key determinant of SSC self-renewal. As RAS is required in many cellular contexts, dysregulation of this pathway is expected to be relevant to the pathology of other disorders. We speculate that the mechanisms described here are likely to contribute to the mutational burden associated with complex disorders for which PAEs have been documented, including some cancers and neurocognitive disorders such as autism and schizophrenia.

115

Interpreting disease-associated SNPs using epigenomics and comparative genomics signatures. *M. Kellis.* MIT / Harvard / Broad Institute, Cambridge, MA.

The large number of single-nucleotide polymorphisms (SNP) from genome-wide association studies (GWAS) that implicate non-coding regions in various diseases poses the important challenge of interpreting their likely molecular mechanisms of action, needed for inferring drug targets and therapeutics. In this abstract, we describe our use of epigenomics and comparative resources from the ENCODE project and the comparative sequencing of 29 mammals to help prioritize disease-associated SNPs in non-coding regions. (1) In the context of the ENCODE project, we have used chromatin maps of nine histone modifications in nine human cell types to systematically characterize cis-regulatory elements, their cell type-specificities, and their functional interactions, dramatically expanding the functional annotation of non-coding regions. Our annotations reveal active, poised and inactive promoters; strong, weak, and poised enhancers; insulators; heterochromatic, repressed and repetitive regions. We use the dynamics of these chromatin states across multiple cell types to predict both the transcription factors that activate or repress enhancer regions, and the target genes that these enhancers likely regulate, based on their correlated patterns of activity. (2) In the context of the comparative analysis of 29 mammalian genomes, we have defined high-resolution maps of evolutionary constraint across the human genome, and used them to prioritize disease-associated SNPs that disrupt evolutionarily-conserved non-coding elements. These include individual binding sites for transcription factors involved in regulatory processes relevant to the disease phenotype, non-coding RNAs with conserved secondary structures pinpointing potential functional domains, and synonymous SNPs within protein-coding exons that we predict disrupt overlapping functional elements under excess constraint. (3) These epigenomics and comparative genomics maps have enabled us to revisit disease-associated SNPs in non-coding regions and reveal their predicted upstream regulators and downstream target genes. In some cases, the SNPs specifically disrupt or create evolutionarily-conserved regulatory motif instances for predicted causal regulators, providing a potential mechanistic explanation for the observed SNP-disease association. Our results suggest a general framework for using genomic and epigenomic datasets to interpret cis-regulatory connections and their role in health and disease.

116

Near complete characterization of genetic variation in the Kuusamo population isolate. *K. Palin[1], K. Rehnström[1], O. Pietiläinen[2,3], J. Suvisaari[3], S. Ripatti[2,3], J. Lönnqvist[3], M. Perola[2,3,4], V. Salomaa[3], A. Palotie[1,2], R. Durbin[1].* 1) The Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 2) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland; 3) National Institute for Health and Welfare (THL), Helsinki, Finland; 4) University of Tartu, Estonian Genome Center, Tartu, Estonia.

The increasing output of the new generation sequencing machines have enabled dissection of human genetic variability in astonishing detail. The 1000 genomes project is on track to uncover most variants down to 1% allele frequency in world wide outbred populations. Due to the large size and rapid growth of the human population, this approach will miss a lot of variants with allele frequency lower than 1%. In outbred populations such rare variants are abundant due to the long time to the Most Recent Common Ancestor (MRCA), whereas they might have significant frequencies in population isolates where time to the MRCA is much shorter. In population isolates, most people are expected to share a recent common ancestor at most loci with many other individuals from the same isolate. Here we describe results of whole genome sequencing of 200 individuals from the isolated founder population of Kuusamo in north eastern Finland. This approach reveals variants that are ultra-rare outside the isolate but were carried by one of the ~100 founders and have survived or even drifted to higher frequency in the current population of ~20 000. We will show how we can use the abundance of recent common ancestry within the isolate to improve variant calling from low coverage sequencing and to impute >73% of haplotypes present in Kuusamo to very high accuracy. The accuracy is a direct effect of the isolation: most individuals share long segments Identical By Descent (IBD) from recent common ancestors at any locus making it possible to transfer information between the individuals at inferred IBD segments. In addition, we estimate to discover 75-90% of the variants that were present in the original founders and have survived the past ~12 generations of isolation. As we are more likely to sample high frequency variants, these are expected to cover over 90% of the haplotypes in the present day population. To demonstrate how the sequence level phased haplotypes can be used for disease gene mapping, we are imputing the sequence to a collection of 140 genotyped schizophrenia patients (about half of the cases in the region) and search for rare variants that may have drifted up in frequency and contribute to the observed excess in disease burden.

117

Disease-gene identification and sex-specific mutation rate estimation from whole-genome sequence data on 21 individuals in a five-generation heart disease pedigree. *C.D. Huff[1], H. Hu[1], J.C. Roach[2], J. Xing[1], A.F.A. Smit[2], G. Glusman[2], A.K. Holloway[3], V. Garg[4], B. Moore[1], R. Hubley[2], H. Li[2], S.Z. Montsaroff[2], D.E. Abbott[2], L.E. Hood[2,5], K.S. Pollard[3], L.B. Jorde[1], M. Yandell[1], D.J. Galas[2,5], D. Srivastava[6].* 1) Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah, Salt Lake City, Utah, USA; 2) Institute for Systems Biology, Seattle, WA, USA; 3) Gladstone Institute of Cardiovascular Disease and University of California, San Francisco, San Francisco, CA, USA; 4) Department of Pediatrics,The Ohio State University and Center for Cardiovascular and Pulmonary Research, Research Institute at Nationwide Children's Hospital; 5) Luxembourg Center for Systems Biomedicine, University of Luxembourg, Luxembourg; 6) Gladstone Institute of Cardiovascular Disease and Departments of Pediatrics and Biochemistry, University of California, San Francisco, San Francisco, CA, USA.

Whole-genome sequences from large pedigrees allow direct estimates of sex-specific mutation rates and present new opportunities and challenges for disease-gene discovery. We present a pedigree-based, disease-gene identification method that builds on the approach implemented in the VAAST software package. This tool, implemented in VAAST version 1.1, incorporates pedigree information and the observed inheritance pattern with information about amino acid substitution severity and genetic variation in a control population under a unified framework. The method allows researchers to combine whole-genome data from one or more pedigrees and sporadic cases to increase power while accurately assessing statistical significance in a manner that accounts for the pedigree relationships and the number of tests performed. We applied this tool to whole-genome data from a five-generation pedigree to identify genes associated with congenital heart defects. Of the 21 individuals in the pedigree, 11 are affected with cardiac septal defects. The variant responsible for septal defects in the family is a known missense mutation in the GATA4 gene that had been previously identified through traditional linkage analysis. In our whole-genome analysis of the septal defect phenotype, the GATA4 gene was highly significant (Bonferroni corrected p-value = 4.2×10^{-9} x 21,000 genes = 8.9×10^{-5}), and no other gene reached statistical significance. We were able to detect GATA4 with genome-wide significance when as few as 5 individuals were included in the analysis. Three of the individuals with septal defects are also affected with additional cardiac abnormalities, including cardiomyopathy. In our analysis of this phenotype, the highest-ranked candidate gene in the genome was statistically significant when the analysis was restricted to known cardiomyopathy genes (Bonferroni corrected p-value = 0.01). To estimate the male-specific intergenerational mutation rate, we identified novel single nucleotide variants (SNVs) that were absent in a father but were present on the paternal chromosomes of one of the father's offspring. We identified 11 mutations in approximately 700 Mb of sequence data, with estimated false-positive and false-negative rates of less than 1×10^{-3}, for a male-specific, intergenerational mutation rate of 1.6×10^{-8} per nucleotide. Comparing this result to previous sex-averaged mutation rate estimates, we obtain an estimate of 2.5 for the ratio of male-to-female mutation rates.

118

Comprehensive Characterization of Human Genome Variation by High Coverage Whole-genome Sequencing of Forty Four Caucasians. *J.G. Zhang[1,2], H. Shen[1,2], J. Li[1,2], C. Xu[1,2,3], Y. Jiang[1,2,3], Z.K. Wu[1,2,3], F.P. Zhao[1,2], L. Liao[1,2], Q. Tian[1,2], C. Papasian[2], H.W. Deng[1,2,3,4].* 1) Center for Bioinformatics and Genomics, Department of Biostatistics and Bioinformatics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA 70112, USA; 2) School of Medicine, University of Missouri-Kansas City, Kansas City, MO 64108, USA; 3) Center of System Biomedical Sciences, University of Shanghai for Science and Technology, Shanghai 200093, P. R. China; 4) College of Life Sciences and Bioengineering, Beijing Jiaotong University, Beijing 100044, P. R. China.

Whole genome sequencing studies are essential to obtain a comprehensive understanding of the vast pattern of human genomic variations. Here we report the results of a high-coverage whole genome sequencing study for 44 unrelated healthy Caucasian adults. For each individual, 147.7-229.6 gigabases (Gb) of sequence were generated and mapped to the NCBI human reference genome (build 37.1), resulting in an average 65.8-fold (range: 51.7-80.3x) genomic coverage. We identified approximately 11 million single nucleotide polymorphisms (SNPs), 2.8 million short insertions and deletions, and over 500,000 block substitutions. We showed that, although previous studies, including the pilot phase of the 1000 Genome Project, have catalogued the vast majority of common SNPs, many of the low-frequency and rare variants remain undiscovered. For instance, of all the SNPs (n = 10,862,507) mapped to autosomes and X chromosome, approximately 2.1 and 3.2 million SNPs were not reported in the 1000 Genome Project data set and dbSNP respectively, and about 1.1 million SNPs were novel to both data sets. On average, each individual genome carried approximately 330 loss-of-function variants that resulted in protein truncation, frameshift change, or the loss of a stop codon. At the population level, 103 of these variants occurred in a homozygous state in an average individual genome, which would completely “knock out” the annotated genes. While about 40% of these genes were “knocked out” in just one or two individuals, 111 genes were “knocked out” in over 30% of our samples, suggesting that a number of genes are commonly “knocked-out” in general populations. Gene ontology analysis suggested these commonly “knocked-out” genes were significantly enriched biological processes related to antigen processing and immune responses. This study represents the first high-coverage analyses of multiple genomes for apparently healthy human subjects in a single population of the same ethnicity. Our results contribute towards a more comprehensive characterization of human genomic variation, especially for less-common and rare variants, and provide a valuable resource for future genetic studies of human variation and diseases.

119

Mining the Transcriptomes of Malaria-infected Children using joint genotypic and expression analysis. *Y. Idaghdour[1], J. Quinlan[1], E. Gbeha[1], C. Rahimy[2], A. Sanni[3], P. Awadalla[1].* 1) Sainte-Justine Research Center, University of Montreal, Montreal, QC; 2) Université d'Abomey-Calavi, National Sickle Cell Disease Center, Cotonou, RP, Benin; 3) Université d'Abomey-Calavi, Biochimie et Biologie Moléculaire, Cotonou, RP, Benin.

In the era of genome and systems biology, joint analysis of phenotypic, genotypic and gene expression data emerged as a promising approach to map complex diseases. An increasing number of studies are showing strong correlations between exposures and transcriptional signatures. Here we address the sources of variability of gene expression in 93 malarial African children having various levels of parasitemia and 64 age-matched controls from West Africa using gene expression profiling and genome-wide genotyping. We defined the range of transcriptional states and evaluated association between transcription profiles and various phenotypes (parasitemia, complete blood count, hemoglobin levels, location, age, gender and ethnicity) using a combination of supervised and unsupervised statistical models. Over a third of the variation in the data is contained in the first three principal components, two thirds of which is explained almost exclusively by parasitemia status after adjusting for all other variables. Results of an analysis of covariance are consistent with the observation that parasitemia status has a heaviest and major influence on the leukocytes of children affected with malaria. Gene enrichment, MMC modeling and principal component analysis pinpoint key networks that are affected including those enriched for genes involved in key immune response pathways. Furthermore, we have conducted, to our knowledge, the first genome scan for eQTLs in malaria and identified several hundred variants associated with transcript abundance. We confirmed that associations between genotype and expression are orders of magnitude stronger than with disease, with one of the more prominent cis-regulatory association regions being the MHC. Insight gained from this study highlights the power of joint genotypic and expression analysis in the context of infectious diseases.

120

Whole Exome Sequencing of 258 Individuals Not Selected for Cancer History Identifies High-Penetrance Cancer Susceptibility Gene Mutations: A Genomics-First Screen. *W.S. Rubinstein[1,2,3], J.J. Johnston[1], F.M. Facio[1], D. Ng[1], L.N. Singh[1], J.K. Teer[1,4], J.C. Mullikin[4,5], L.G. Biesecker[1,4].* 1) Genetic Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Dept. of Medicine, NorthShore Univ HealthSystem, Evanston, IL; 3) Dept. of Medicine, University of Chicago Pritzker School of Medicine, Chicago, IL; 4) NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 5) Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

BACKGROUND Whole genome and exome sequencing costs continue to fall and more individuals are undergoing these assessments. We evaluated the potential of exome sequencing to detect mutations that cause cancer susceptibility syndromes in an adult cohort not ascertained for cancer. METHODS Whole exome sequencing was performed on 258 subjects of the ClinSeq project. Variants in 37 genes known to cause cancer susceptibility syndromes were evaluated using an algorithm that filtered for mutation type, quality metrics, frequency, and information in mutation databases, using established causation categories (1-5, with 5 indicating >99% confidence of pathogenicity, 4 indicating 95-99% confidence, 3 indicating 5-95%, and 2 and 1 the inverse of 4 and 5; Plon et al. 2008). RESULTS We identified 251 sequence variants that differed from the reference. Exclusions were made on sequence quality (n=24) and high frequency in the cohort (n=61) or dbSNP (n=11), leaving 155 variants of potential clinical importance. Six variants scored 4 or 5. Three of these six variants were in MUTYH, in a heterozygous state. One was in APC, p.Ile1307Lys, which is a low penetrance founder mutation in the Ashkenazim. Two male participants had deleterious BRCA2 mutations. The first (c.6174del p.Ser1982ArgfsX22) was in a Jewish family with multiple cases of breast and ovarian cancer. The second (c.8297del p.Thr2766AsnfsX11) did not meet family history-based referral criteria. We further evaluated several class 3 variants. Clinical characterization of a novel NF2 missense variant showed it was benign. We determined that a CDH1 missense variant reported as pathogenic is more likely benign. CONCLUSIONS Whole exome cancer annotation detected clinically important mutations in cancer susceptibility genes, including one in an individual without a significant family history. We suggest that a `genomics-first' sequencing approach is the only way this diagnosis could have been made presymptomatically. However, the current state of informatics, mutation databases, and the literature pose obstacles to adopting exome sequencing into clinical practice. These results show that clinically important results can be gleaned by exome annotation, while providing an early indication of the relative proportions of incidental findings and inaccuracies in genetic resources. Genomic screening is potentially transformative, but much work remains to be accomplished before it can be widely adopted as a clinical tool.

121

Empirical assessment of imputation of coding variants using next-generation sequencing. *N. Soranzo[1], Y. Memari[1,2], B. Zhang[3], J. Harris[2], G. Guo[4], S-Y. Shin[1], G. Clement[2], H. He[4], V. Anttila[1], W. Jia[4], A. Valdes[2], B. Dougherty[3], FM. Williams[2], T. Jiang[4], S. John[3], T. Spector[2].* 1) Human Genetics, Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 2) Department of Twin Research & Genetic Epidemiology, King's College London, London SE1 7EH; 3) Pfizer Global Research & Development, Groton, CT, USA; 4) BGI-Shenzhen, Shenzhen, China.

The decreasing cost of next generation sequencing technology will soon enable comprehensive analyses of intermediate to low-frequency variants in genetic association studies. The relative performance of whole-genome or whole-exome sequencing versus the relatively cheaper imputation from dense haplotype maps such as the 1000 Genomes Project (1KGP) remains to be established. We sequenced the exomes of 240 individuals from a UK cohort at high coverage and we compared this exhaustive set of coding variants to 1KGP-based imputation of genome-wide SNP data to address the following three questions: (i) how well does 1000G imputation captures coding variants; (ii) what is the gain in complementing the 1KGP panel with exomes as additional reference data for imputation; (iii) how these gains scale with increasing numbers of sequenced exomes. We found that on average 50% of the variants detected from sequencing exomes are not recovered by imputation from 1KGP, with coverage and accuracy of imputation being strongly dependent on allele frequency. Namely, through 1KGP imputation we recover over 90% of common variants (minor allele frequency (MAF) of 5% of more) found in the UK cohort, but this fraction decreases to 50% for variants with MAF≤1% and to less than 10% for singletons. We further show that the addition of the exome variants to the 1KGP imputation reference panel informs the discovery of population-specific coding variants. We show that this design enables the imputation of approximately half of the genetic variants which are unique to the exome reference panel, although with a lower imputation accuracy compared to when the variant exists in 1KGP. We suggest that imputation from 1KGP fails to directly capture many population-specific variants, and that the addition of sequenced samples from populations can greatly enhance the value of these public resources.

122

Using whole-genome data to reveal distant relationships between individuals by maximum-likelihood analysis of genomic segments shared identically by descent (IBD). *D.J. Witherspoon, C.D. Huff, J. Xing, L.B. Jorde.* Dept Human Genetics, Univ Utah, Salt Lake City, UT.

The locations and extents of chromosomal segments shared by two individuals due to identity by descent (IBD) from a recent shared ancestor can be estimated from whole-genome data (resequencing or high-density SNP data.) This IBD segment information allows the detection and accurate estimation of genetic relationships between pairs of individuals as distant as 5th cousins, with detection power approaching the maximum theoretical limit. We implemented this method in the software ERSA (Estimation of Recent Shared Ancestry; Huff, Witherspoon et al. Genome Research, 2011.) We applied ERSA to 1000 Genomes sequence data and Human Genome Diversity Project (HGDP) SNP data to identify relationships among the sampled individuals. Many hundreds of previously unknown relationships were identified with high confidence in the 1000 Genomes samples, and several thousand were detected in the HGDP samples. First- and second-cousin pairs were identified in nearly all 1000G population samples, as well as in all HGDP samples with at least 20 individuals. Numerous more distant relationships were observed in all samples. In some HGDP samples (e.g. the Amazonian Surui and Karitiana, Mexican Pima and Maya, Siberian Yakut, Pakistani Kalash, and the Melanesian sample), all pairs of individuals were at least 3rd cousins or more closely related, suggesting the possibility of estimating pedigrees for the entire samples. In the Brahui and Burusho samples from Pakistan, the Bedouin, Druze, Palestinian and Mozabite from the middle east, and in the Basque HGDP sample, more than half of pairwise relationships were 3rd-cousin or closer. Another set of HGDP samples showed lower levels of relatedness, such that only 5-30% of pairwise relationships were at the 3rd-cousin level or closer (the African Mandenka and Biaka Pygmy, the Makrani, Hazara and Balochi of Pakistan, and Italian Sardinians.) In the majority of 1000 Genomes samples, and in the HGDP Han Chinese and Pakistani Sindhi and Pathan samples, fewer than 1% of pairs of individuals were 3rd cousins or closer. Relationships between individuals from different population samples are uncommon and nearly always distant. Thus the detected relationships reflect what is known about recent demographics and mating patterns in these populations. With such precise and detailed estimation of relationships, it will be possible to draw stronger and more exact inferences about recent effective population sizes and mating patterns.

123

Breakpoint signatures and complex chromosomal `shattering' from apparently balanced germline karyotypes revealed by sequencing 82 rearrangement breakpoints. *C. Chiang[1], A. Heilbut[1,2], C.A. Hanscom[1], C. Ernst[1,3], A.M. Lindgren[4], C.C. Morton[4,5], J.F. Gusella[1,3,5], M.E. Talkowski[1,3,5].* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Bioinformatics Program, Boston University, Boston, MA; 3) Department of Genetics and Neurology, Harvard Medical School, Boston, MA; 4) Departments of Obstetrics and Gynecology, and Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 5) Program in Medical and Population Genetics, Broad Institute of Harvard and M.I.T., Cambridge, MA.

Until recently the resolution of cytogenetic methods to delineate balanced chromosomal rearrangements (BCRs) has been low and the breakpoint features of pathogenic rearrangements have not been well characterized. Here we used optimizations in next-generation sequencing to detail 82 breakpoints at base pair resolution from 43 individuals with karyotypically balanced rearrangements and abnormal developmental phenotypes, the largest collection of its kind surveyed to date. At the nucleotide level, translocation breakpoints were found to differ from copy number variations (CNVs) previously described. Short sequence homology (2-20 bp) was observed in 27/82 (33%) of breakpoints, blunt end joining (0-1 bp homology) in 36/82 (44%) of breakpoints, and inserted sequence in 20/82 (24%) of breakpoints. The marked contrast in the frequency of blunt end joining, compared to 8-13% of CNVs, suggests a more prominent role for non-homologous end joining in translocations compared to common insertions, deletions and inversions. Notably, non-allelic homologous recombination was virtually absent from the translocation breakpoints; only 2/82 (2.4%) of breakpoints exhibiting greater than 80 bp of homology, potentially revealing a unique signature of translocation breakpoints or denoting limitations in sequence based methods. The frequency of genic regions disrupted, repeat-masked elements, and abnormal secondary structure was not significantly different than random simulations of breakpoints against the genome. However, at the sequence level, 8/43 (19%) of translocation cases exhibited complex rearrangements at one or both breakpoints, including three instances of apparent shattering and aberrant resolution of chromosomal alterations that we have been detailing to base pair resolution, similar to what has recently been described as “chromothripsis” in more complex rearrangements of cancer genomes. These three cases contained over 10 breakpoints each, and reinserted shattered fragments megabases away from their origin. Our analyses reveal a profile of microhomology and complex series of rearrangements which may indicate that rare, pathogenic BCRs undergo a different mutational etiology from that seen in population-based rearrangements described to date.

124

An evidence-based approach to guide the development of content on chromosomal microarrays and to support interpretation of clinically significant copy number variation. *E. Thorland[1], E. Rooney Riggs[2], S.T. South[3], H.M. Kearney[4], E. Kaminsky[2], E.S. Williams[2], V. Horner[2], K. Hanson[5], R.M. Kuhn[6], D.M. Church[7], S. Aradhya[8], D.H. Ledbetter[9], C. Lese Martin[2], The International Standards for Cytogenomic Arrays Consortium.* 1) Department of Laboratory Medicine & Pathology, Mayo Clinic, Rochester, MN; 2) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 3) Department of Pediatrics and Pathology, ARUP Laboratories, University of Utah, Salt Lake City, UT; 4) Fullerton Genetics Center, Mission Health System, Asheville, NC; 5) Department of Obstetrics & Gynecology, Columbia University, New York, NY; 6) Department of Biomolecular Engineering, Center for Biomolecular Science and Engineering, University of California at Santa Cruz, Santa Cruz, CA; 7) National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD; 8) GeneDx, Gaithersburg, MD; 9) Geisinger Health System, Danville, PA.

Utilizing multiple lines of evidence from clinical practice and published literature, the evidence-based review process has been widely used to develop standards for medical decision-making and to explore complex clinical questions. This approach can similarly guide the development of technical parameters and interpretation guidelines for genetic tests, such as those based on cytogenomic microarrays (CMA). The International Standards for Cytogenomic Arrays (ISCA) Consortium is a group of researchers, clinicians, and clinical laboratory geneticists dedicated to improving patient care by promoting high standards for CMA testing. An integral step toward this goal is to ensure that clinical array design is based on evidence that genomic regions/genes with high density coverage are associated with relevant clinical disorders. In order to address this issue, an Evidence-Based Review committee composed of ISCA members with laboratory, clinical, and bioinformatics experience has been established. This committee was charged with evaluating the quality and variety of evidence needed to determine if a particular gene or region should be targeted on a whole-genome CMA design. The committee has proposed a ranking system that considers the following: relevance of the human phenotype associated with mutations of the gene or region; mode of inheritance for that phenotype; number of individual causative mutations and probands reported; mutational mechanism underlying the phenotype; data for the gene or region in public genome variation databases; and expert consensus opinion. This system will initially be used to evaluate genes currently targeted with high density probe coverage on the ISCA 180K array, with the intent of removing and adding genes according to the evidence supporting their clinical importance. The process will be expanded to include evaluation of recurrent and non-recurrent copy number variant regions based on genomic and phenotypic data from the ISCA Consortium database. Given the dynamic nature of this process, genes will be re-evaluated on a periodic basis to incorporate emerging evidence. Though initially developed for the evaluation of array design, the results of the evidence review can also be used to support the interpretation of copy number variation at the targeted regions. A small pilot project is underway to assess the effectiveness of this system, with the intent of soliciting and incorporating community input in the future.

125

A large autism cohort reveals genomic imbalances associated with many known syndromic deletions and also those linked to potentially novel neurodevelopmental genes. *S. Aradhya, D.M. Riethmaier, A. Fuller, B. Boggs, G. Richard, J.M. Meck.* GeneDx, Gaithersburg, MD.

Among a large number of patients referred for whole genome oligonucleotide array CGH testing, 862 were submitted with a specific indication of autism. In this group of patients, we identified 175 (20.3%) copy number changes that were considered to have a role in the autism phenotype. Among the 175 events, 109 were known clinically significant genomic changes and the remaining 66 were variants of unknown significance. Parental studies were performed in 102 out of the 175 cases and showed that half of the known clinically significant deletions or duplications were de novo events. A variety of classic microdeletion/duplication syndromes were found in the tested cohort, including four cases of the classic 16p11.2 deletion associated with autism susceptibility, ten duplications mapping to 7q11.23 ,17p11.2, or 22q11.2 and respectively constituting reciprocal rearrangements of the Williams, Smith-Magenis, and DiGeorge/VCFS deletion syndromes. There were also five cases of 15q11-q13 duplication also known to be associated with autism. 1q21.1 duplications and 15q13.3 deletions were typically inherited from reportedly unaffected parents. In addition, we also found single-gene deletions involving NRXN1 and AUTS2, which play a role in neurological development and are suspected to be involved in autism. Interestingly, this cohort included two cases of a deletion involving the DMD and three involving deletion of the STS gene; both of these genes have been recently implicated in neurologic phenotypes. Lastly, five cases involved complex intrachromosomal rearrangements or unbalanced translocations. The remaining cases, especially the de novo events, provide an opportunity to define novel autism-related loci. To that effect, twelve de novo events that have not been described previously involve genes with neurologic functions. Comparison to copy number data from clinically normal populations showed that the data described here do not represent benign variation. The data from a large cohort presented here offer valuable insight into the types of genomic changes observed in autism and point to novel loci that may play a role in its pathogenesis.

126

Sequencing chromosomal rearrangements in autism and other developmental disorders reveals a complex genetic architecture across diagnostic boundaries. *M.E. Talkowski[1,2,10], A. Heilbut[1], J. Rosenfeld[7], C. Hanscom[1], C. Chiang[1], C. Enrst[1], A. Lindgren[4], S. Ahsan[4], A. Kirby[1,10], D. Harris[4], B. Soloman[8], A. Gropman[9], D. Lucente[1], K. Sims[1], T.K. Ohsumi[5], M.L. Borowsky[5], J. Miles[6], B-L. Wu[1], Y. Shen[1], L.G. Shaffer[7], M.J. Daly[1,10], C.C. Morton[3,4,10], J.F. Gusella[1,2,10].* 1) Center for Human Genetic Res, Massachusetts General Hospital, Boston, MA; 2) Departments of Genetics and Neurology, Harvard Medical School, Boston, MA; 3) Departments of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital, Boston, MA; 4) Department of Pathology, Harvard Medical School, Boston, MA; 5) Department of Molecular Biology, Massachusetts General Hospital, Boston, MA; 6) Departments of Pediatrics, Medical Genetics & Pathology, The Thompson Center for Autism & Neurodevelopmental Disorders, Columbia, MO; 7) Signature Genomic Laboratories, LLC, Spokane, WA; 8) Medical Genetics Branch, National Human Genome Research Institute, Bethesda, MD, USA; 9) Department of Neurology, Children's National Medical Center and the George Washington University of the Health Sciences, Washington, DC, USA; 10) Program in Medical and Population Genetics, Broad Institute of Harvard and M.I.T., Cambridge, MA.

Delineation of balanced chromosomal rearrangements (BCRs) offers a powerful route to discovery of genes underlying complex disorders, but remains a relatively untapped reservoir of high penetrance alterations because cytogenetic resolution is low and BCRs are not detected by genome-wide microarrays. Here, we used optimizations in paired-end sequencing to precisely delineate BCRs to base pair resolution in 25 patients with autism, autism spectrum disorder, and other neurodevelopmental disorders (NDD). In three subjects, the breakpoints disrupted single genes involved in epigenetic and methylation patterning within defined microdeletion syndrome regions (2q33.1, SATB2; 2q23, MBD5; 9q34, EHMT1). In others, BCRs disrupted genes previously associated with a spectrum of childhood and adult onset neuropsychiatric disorders (TCF4, ZNF804A, PDE10A, AUTS2, CDKL5, C18orf1, GRIN2B). We also discovered novel genes not previously recognized as contributing to NDD, including CHD8, UTRN, KIRREL3, NLRP1, GNA14, SMG6, FAM107B, ZNHIT6, and PON3, further implicating epigenetic mechanisms as well as previously unrecognized pathways in NDD. A survey of over 37,800 individuals tested for copy number alterations independently supported these findings, including small, gene-specific deletions in independent NDD cases altering many of these novel loci and previously interpreted as having “unknown clinical significance”. Sequencing also uncovered extensive rearrangement complexity, including a high proportion of breakpoint-associated inversions with potential clinical and mechanistic significance, as well extensive exchanges of material from multiple breakpoints in single individuals. Our findings provide strong evidence for a contribution of these genes to NDD and support the view that for many, disruption can result in variable expression and pleiotropic effects ranging from autism to schizophrenia, challenging the notion of distinct genetic etiology for these disorders. These findings significantly expand the set of known genes and pathways responsible for NDD and will have profound implications for autism research, mechanistic studies of genomic rearrangements, and future interpretation in clinical diagnostics.

127

High resolution prenatal array CGH: improved detection, increased complexity. *A. Breman, S. Darilek, A. Pursley, P. Ward, P. Hixon, W. Bi, C. Shaw, P. Stankiewicz, C. Bacino, C. Eng, A. Patel, J.R. Lupski, A. Beaudet, S.W. Cheung.* Molecular & Human Genetics, Baylor College Med, Houston, TX.

Objectives: Baylor College of Medicine has spent nearly a decade clinically implementing high resolution human genome analyses that use microarray technology to detect relevant genomic abnormalities. In addition to our experience with over 35,000 postnatal patient samples, we have now analyzed approximately 1000 clinical prenatal samples by array CGH. Notably, more than half of these were analyzed on our custom high resolution clinical array which utilizes 105,000 oligonucleotides to interrogate regions of the human genome associated with over 150 known genetic disorders with a backbone coverage on average of one oligo every ~30 kb. Methods: Array CGH analysis was performed on DNA extracted directly from amniotic fluid (AF) in 494 (49%) cases or chorionic villus sampling (CVS) in 150 (15%) cases and from cultured cells (amniocytes/CVS) in 339 (34%) cases. For the remaining 24 cases, the source of fetal DNA was tissue, fibroblasts, or fetal blood. Results: In total, copy number changes were detected in 211 (21%) cases. Of these, 74 (7.4%) cases had findings that were interpreted as clinically significant and 129 (12.8%) were interpreted as likely benign. The remaining 8 (0.8%) cases had results of uncertain clinical significance. Of the cases in which clinically significant genomic imbalances were detected, 24 (2.4% of total cases or 32% of abnormal cases) would not have been detected by conventional G-banded prenatal chromosome studies alone. From these data we have observed a >2-fold higher detection rate for submicroscopic copy number changes using our custom high resolution array. In addition, we have identified 8 cases in which⁄ 2 CNVs were inherited in a bi-parental fashion. In many of these cases follow-up information is lacking, and therefore it remains to be determined whether particular CNV combinations, i.e. bi-locus or digenic CNV combinations, will be clinically significant in the child. Conclusion: Evolving array coverage has increased our detection rates for clinically significant copy number changes in prenatal cases, while also adding complexity to the interpretation process when multiple copy number changes are identified. With the number of cases having two or more inherited CNVs likely to increase, findings involving biparental inheritance are likely to have implications for interpretation in prenatal diagnosis, and additional family studies may be warranted.

128

Patterns of Recombination Along Nondisjoined and Normally Disjoined Chromosomes 21 that Exhibit Multiple Recombinant Events on 21q. *T. Oliver[1], S.W. Tinker[1], E. Graves Allen[1], N. Hollis[1], A.E. Locke[1], L.J.H. Bean[1], R. Chowdhury[2], F. Begum[3,4], M. Marazita[3,5], V. Cheung[6,7,8], E. Feingold[3,4], S.L. Sherman[1].* 1) Dept Human Genetics, Emory University, Atlanta, GA; 2) Department of Pediatrics, University of Pennsylvania, Philadelphia, PA; 3) Department of Human Genetics, Graduate School of Public Health University of Pittsburgh, Pittsburgh, PA; 4) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 5) Center for Craniofacial and Dental Genetics, Division of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA; 6) Departments of Pediatrics and Genetics, University of Pennsylvania, Philadelphia, PA; 7) Howard Hughes Medical Institute, University of Pennsylvania, Philadelphia, PA; 8) Department of Genetics, University of Pennsylvania, Philadelphia, PA.

We previously examined characteristics of maternal chromosomes 21 that exhibited a single recombination on 21q and proposed that certain recombination configurations are risk factors for either meiosis I (MI) or meiosis II (MII) nondisjunction. The primary goal of this analysis was to examine characteristics of maternal chromosomes 21 that exhibited multiple recombinations on 21q to determine whether any different risk factors or mechanisms are suggested. In order to identify the origin (maternal or paternal) and stage (MI or MII) of the meiotic errors, as well as placement of recombination, we genotyped over 1500 SNPs on 21q. Our analyses included 785 maternal MI errors, 87 of which exhibited two recombinations on 21q, and 283 maternal MII errors, 81 of which exhibited two recombinations on 21q. Our analyses of the multiple recombinations on 21q, the first of its type, showed that among MI cases, that the average location of the distal recombination was proximal to that of normally segregating chromosomes 21 (35.28 vs 38.86 Mb), a different pattern than that seen for single events and one that suggests an association with genomic features or decreased interference. For MII errors, the most proximal recombination was closer to the centromere than that on normally segregating chromosomes 21 and this proximity was associated with increasing maternal age. This pattern is consistent with that seen among MII errors that exhibit only one recombination, thus potentially the same mechanism that causes MII errors when there is either 1 or 2 recombinants on 21q. These findings provide insight on mechanisms underlying meiotic chromosome mal-segregation.

129

Somatic mosaicism of large chromosomal anomalies in blood cells of normal adults. *C.C. Laurie[1], C.A. Laurie[1], K. Doheny[2], L. Zelnick[1], C. McHugh[1], J. Shen[1], X. Zheng[1], H. Ling[2], K. Hetrick[2], E. Pugh[2], D. Mirel[3], C. Amos[4], T. Beaty[5], L. Bierut[6], N. Caporaso[7], N. Freedman[7], E. Feingold[8], J. Li[9], C. Haiman[10], J. Heit[11], W. Lowe[12], T. Manolio[13], M. Marazita[14], J. Murray[15], L. Pasquale[16], G. Jarvik[17], I. Ruczinski[18], V. Seshan[19], B. Weir[1], GENEVA Consortium.* 1) Dept of Biostatistics, Univ Washington, Seattle, WA; 2) Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD; 3) Broad Institute (MIT/Harvard), Cambridge, MA; 4) MD Anderson Cancer Center, University of Texas, Houston, TX; 5) Department of Epidemiology, School of Public Health, Johns Hopkins University, Baltimore, MD; 6) Department of Psychiatry, Washington University School of Medicine, St Louis, Missouri; 7) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland; 8) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 9) Department of Human Genetics, University of Michigan, Ann Arbor , MI; 10) Department of Preventive Medicine, University of Southern California, Los Angeles, CA; 11) Department of Internal Medicine, Mayo Clinic, Rochester, MN; 12) Division of Endocrinology, Metabolism, and Molecular Medicine, Northwestern University, Chicago, IL; 13) Office of Population Genomics, National Human Genome Research Institute at National Institutes of Health, Bethesda, MD; 14) Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA; 15) Department of Epidemiology in Public Health, University of Iowa, Iowa City, IA; 16) Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Harvard University, Boston, MA; 17) Division of Medical Genetics, University of Washington, Seattle, WA; 18) Department of Biostatistics, School of Public Health, Johns Hopkins University, Baltimore, MD; 19) Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY.

Somatic mosaicism for large chromosomal anomalies is common in children with developmental abnormalities, but has not been well studied in normal adults. We developed an automated method to detect large (>50 kb) chromosomal anomalies from SNP microarray data (B Allele Frequency Log R Ratio) and to classify them as constitutional or mosaic. This method was applied to more than 50,000 subjects from the `Gene-Environment Association Studies' (GENEVA) consortium of case-control, cohort and family studies, which focus on common diseases, such as alcoholism, lung cancer and cardiovascular disease. A majority of the participants in these studies were either healthy adult controls or adults affected by late-onset diseases. DNA samples were from blood (91%) or saliva/buccal cells (9%). Our method detects constitutional and mosaic copy number variants (CNV), as well as mosaic copy-neutral loss of heterozygosity (i.e. segmental uniparental disomy, UPD). Anomalies transmitted from parent to offspring in family studies provide a standard for characterizing the BAF/LRR properties of constitutional anomalies with discrete copy numbers. Intermediate BAF/LRR properties identify somatic mosaics. The following results are for autosomal mosaics. The probability of having one or more mosaics is very highly associated with age (p<1e-16) but not with sex, ethnicity, smoking status or DNA source. The frequency of mosaics in subjects under 50 years old is very low (< 0.5%), but rises rapidly thereafter, reaching a frequency of 2-3% in adults over 70 years old. The mosaic segments are often large (median 15.5 Mb), sometimes spanning an entire chromosome or chromosomal arm. Disomic/trisomic mosaics of most autosomes were detected (2,3,8,9,12-19,21,22). One third of mosaic anomalies detected involve acquired UPD of terminal segments, consistent with their origin through mitotic recombination. Two thirds of the mosaics are CNVs, of which 78% involve loss and 22% gain of copy number. There is an excess of multiple mosaics per subject compared with Poisson expectation. Some observed anomalies are characteristic of leukemia/lymphoma, but most occur in subjects with no known history of hematological cancer. The age-dependence of mosaic anomaly frequency suggests that most are acquired late in life. The health consequences are unknown, but mosaic anomalies may represent pre-leukemic conditions and/or contribute to age-related decline in cellular function.

130

Identification of single gene alterations detected by whole-genome array comparative genomic hybridization. *N.J. Neill, B.C. Ballif, J.B. Ravnan, B.S. Torchia, R.A. Schultz, J. Ellison, L.G. Shaffer.* Signature Genomic Laboratories, PerkinElmer Inc., Spokane, WA.

The application of array comparative genomic hybridization (aCGH) to clinical diagnostics in the past decade has led to the discovery of numerous novel microdeletion/duplication syndromes and dramatically changed our understanding of structural rearrangement within the genome. In our laboratory, approximately 23,726 patients have been tested using oligonucleotide-based, whole-genome aCGH since 2008. A total of 8,944 copy-number abnormalities in 7,318 patients were identified and 1,316 (14.7%) of these abnormalities (ranging in size from 10 kb to 2.71 Mb) were found to affect only a portion of a single gene, including 712 purely intragenic alterations. Of these, 218 alterations had known inheritance, and 16 (7.3%) were found to be *de novo*, all of which were deletions. There was a strong correlation between an increase in the number of coding exons affected by a deletion and the likelihood the deletion was *de novo*. Significant enrichment of single gene alterations was observed in genes such as *PARK2*, *AUTS2*, *RBFOX1*, *CSMD1*, and *CNTN4*. Many of these alterations were inherited rather than *de novo*, suggesting that these regions represent hotspots of mutation predisposed to the formation of rearrangements. Additional, *de novo* single gene alterations were identified, including deletions of *CREBBP*, *TCF4*, *WHSC1*, *NPAS3*, and *SMC1A*. As a whole, this data demonstrates the frequency of single gene alterations detected by high-resolution aCGH, illustrates the distribution of these alterations throughout the genome, and establishes the utility of aCGH in increasing our appreciation of the relevance of small, submicroscopic alterations which may go undetected by traditional methods.

131

Two novel loci for melanoma susceptibility on chromosome bands 1q42.12 and 1q21.3. *S. Macgregor[1], M. Law[1], G.W. Montgomery[1], D.T. Bishop[4], C.I. Amos[5], Q. Wei[5], K.M. Brown[3], N.G. Martin[1], G.J. Mann[2], N.K. Hayward[1].* 1) Queensland Inst Med Res, Brisbane, Australia; 2) Westmead Institute of Cancer Research, University of Sydney at Westmead Millennium Institute and Melanoma Institute Australia, Westmead, NSW 2145 Australia; 3) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA; 4) Section of Epidemiology and Biostatistics, Leeds Institute of Molecular Medicine, Cancer Research UK Clinical Centre at Leeds, St James's University Hospital, Leeds, UK; 5) Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA.

We performed a genome-wide association study of melanoma in a discovery cohort of 2,168 Australian melanoma cases and 4,387 controls, confirming several previously characterised melanoma-associated loci and identifying two novel susceptibility loci on chromosome 1. The most significant genotyped SNPs in the novel loci were at 1q21.3 (best genotyped SNP rs7412746, P=2.5 x 10^{-7}, OR=0.82) and at 1q42.12 (best genotyped SNP rs3219090, P=9.5 x 10^{-7}, OR=0.82). Both new findings were replicated in three independent case-control studies (Europe: 2,804 cases, 7,618 controls; USA 1: 1,804 cases, 1,026 controls; USA 2: 585 cases, 6,500 controls). Estimates of the ORs in the combined replication cohorts were 0.89 for rs7412746 (P=1.5x10^{-5}) and 0.91 for rs3219090 (P=3.4 x 10^{-3}). The meta-analysis P-values for all case-control studies combined were P=9.0 x 10^{-11} for rs7412746 and P=9.3 x 10^{-8} for rs3219090. At 1q42, rs3219090 and all other SNPs in high linkage disequilibrium lie in or near to PARP1, a DNA repair gene. Imputation in the discovery cohort using 1000 Genomes data revealed most evidence for assocation at rs2695238 (P=3.8 x 10^{-7}) 40kb from rs3219090. Linkage disequilibrium at the 1q21 locus covers several genes, including SETDB1, ARNT, LASS2, ANXA9, CTSK and CTSS. In publicly available lymphoblastoid cell line data, rs7412746 (and several correlated SNPs) is an eQTL which explains a substantial proportion (15%) of the variance in expression of ARNT and CTSS. Expression of ARNT is correlated with a number of genes in the region, including CTSK and to a lesser extent SETDB1. Based on recently published work in zebrafish, SETDB1 is an outstanding candidate because it was shown to accelerate melanoma formation. We further examined the relationship between rs7412746 genotype and expression in two sets of melanoma cell lines (N=39 and N=38 hybrized to Affymetrix and Illumina expression arrays, respectively). In the Affymetrix data, rs7412746 explained 8% and 2% of the variance in expression of ARNT and SETDB1, respectively. In the Illumina data, rs7412746 explained 7% and 9% of the variance in expression of ARNT and SETDB1, respectively. In summary, we have found two new loci underlying melanoma susceptibility. At the 1q42 locus, genetic variation in a DNA repair gene influences disease risk. At the 1q21 locus, the lead SNP cis-regulates gene expression of several genes, including one reported in a model organism to accelerate melanoma formation.

132

Pathway analysis of malignant melanoma identifies 30 pathways associated with melanoma cancer risk. *C.E. Amos[1], L. Wang[1], W.V. Chen[1], S. Fang[1], J.E. Lee[2], Q. Wei[1].* 1) Dept Epidemiology, MD Anderson Cancer Ctr, Houston, TX; 2) Dept Surgery, MD Anderson Cancer Ctr, Houston, TX.

Malignant melanoma causes 75% of skin-cancer related death. and siblings of cases have a 2-4 fold higher melanoma risk, suggesting the importance of genetic factors in its etiology. We therefore conducted a genome-wide association analysis of melanoma using samples and data from 1804 melanoma cases from M.D. Anderson Cancer Center and 1026 age and sex-matched healthy controls. Samples were analyzed using an Omni 1M Quad V1-0_B SNP chip, and standard quality control filters were applied with assistance from the GENEVA coordinating center, including removing SNPs and samples with low call rates (<95% and < 90% respectively) and removing non-Caucasians and genetic outliers. After being filtered, 818237 SNPs had minor allele frequency >0.01, passed the Hardy-Weinberg Disequilibrium test (p> 10-5) and were retained for analyses. Pathway-based analyses were conducted on these data using the quadratic test developed by Li et al. (Eur J Hum Genet, 2011), which includes all of the data from individual association tests within 50Kb of each gene and adjusts for associations among the markers. Results identified 30 pathways that showed nominally significant associations with melanoma risk. The most significant of these, showing p values less than 5 x 10-7, were Neuroactive Ligand-Receptor Interaction (NLRI), Cytokine-Cytokine receptor interactions (C-C), MAPK Signaling Pathway, Olfactory Transduction, WNT Signaling Pathway, Tight Junction, Cell Communication, and Melanogenesis. The architecture of these pathways in predicting risk was quite different with the NLRI pathway showing no very significant genetic effects, but a preponderance of small contributions, while the C-C had very significant signals from selected Interferon genes and Interleukin 5. Further studies to assess the p-values through permutation analyses are underway, and replication studies will be conducted to identify which of these findings are verified by additional GWAS studies. In addition, we will assess the association of summary scores from these pathways with survival outcomes.

133

Genome sequencing of melanoma families identifies a novel recurrent mutation in MITF predisposing to familial and sporadic melanoma. *K.M. Brown[1], S. Yokoyama[2], S. MacGregor[3], A.E. Cust[4], J. Taylor[5], P.D. Pharoah[6], D.F. Easton[6], A.M. Dunning[6], J.A. Newton-Bishop[5], G.W. Montgomery[3], N.G. Martin[3], G.J. Mann[7], D.T. Bishop[5], H. Tsao[8], J.M. Trent[9], D.E. Fisher[2], N.K. Hayward[3].* 1) DCEG-LTG, National Cancer Institute, Bethesda, MD, USA; 2) Department of Dermatology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; 3) Queensland Institute of Medical Research, Brisbane, QLD, Australia; 4) Cancer Epidemiology and Services Research, Sydney School of Public Health, Sydney Medical School, University of Sydney, Australia; 5) Section of Epidemiology and Biostatistics, Leeds Institute of Molecular Medicine, Cancer Research UK Clinical Centre at Leeds, St James's University Hospital, Leeds, England, UK; 6) University of Cambridge, Cambridge, England, UK; 7) Westmead Institute of Cancer Research, University of Sydney at Westmead Millennium Institute and Melanoma Institute Australia, Westmead, NSW, Australia; 8) Department of Dermatology, Wellman Center for Photomedicine, Massachusetts General Hospital, Harvard Medical School, Boston MA, USA; 9) Translational Genomics Research Institute, Phoenix, AZ, USA.

The two known high penetrance melanoma predisposition genes, *CDKN2A* and *CDK4*, account for less than 50% of the high density familial clustering of melanoma. To identify other familial melanoma genes, we conducted whole-genome sequencing of DNA from probands of several melanoma families. In one affected individual we identified a novel germline variant (c.G1075A, NM_000248.3; p.E318K, NP_000239.1) in the melanoma lineage-specific oncogene *MITF*. The variant cosegregated with melanoma in some but not all cases in the family, indicating a possible intermediate risk variant. Consistent with this, the variant was significantly associated with melanoma in a large Australian case-control sample, giving a moderately large odds ratio (OR) of 2.28, 95% CI 1.15-4.77 (carrier frequency for cases and controls, 0.0168 and 0.0074, respectively; *P*=0.008). A similar effect was seen in a case-control analysis of individuals from the United Kingdom (UK *P*=0.007; combined *P*=0.0002, OR 2.17, 95% CI 1.38-3.35). In the Australian sample the variant allele was also significantly over-represented in cases with a family history of melanoma (OR 2.99, 95% CI 1.21-7.29), multiple primary melanomas (OR 4.33, 95% CI 1.54-11.41) or both (OR 8.97, 95% CI 2.73-26.01). MITF E318K cosegregates with melanoma in minimally 28% of families carrying the variant and is present in 7 of 205 (3.4%) melanoma cell lines tested, providing further support for an association with melanoma. The variant allele was also associated with increased nevus count (combined *P*=0.0017, OR 2.53, 95% CI 1.42-4.52) and non-blue eye colour (combined *P*=0.013, OR 2.06, 95% CI 1.14-3.90). Functional analysis of E318K showed that MITF encoded by the variant allele had impaired sumoylation and increased transcriptional activity for several MITF targets. These data suggest that *MITF* is a melanoma susceptibility gene and highlights the utility of whole-genome sequencing to identify novel rare variants associated with disease susceptibility.

134

An oncogenic MITF germline substitution p.E318K impairs sumoylation and predisposes to melanoma and renal carcinoma. *B. Bressac-de Paillerets[1,2], F. Lesueur[1,3], S. Giuliano[4], M. de Lichy[1], K. Bille[4], B. d'Hayer[1], H. Mohamdi[2,5,6], A. Remenieras[1, 16], E. Maubec[2,7], P. Vabres[8], L. Thomas[9], D. Zelenika[10], P. Galan[11], V. Chaudru[2,6,12], S. Richard[13], G. Lenoir[1], M. Lathrop[6,10], M-F. Avril[14], F. Demenais[2,5,6], R. Ballotti[4,15], C. Bertolotto[4,15], The French familial melanoma study group.* 1) Biopathology, Institut de Cancérologie Gustave Roussy, Villejuif, France; 2) INSERM, U946, Bâtiment IGM, Fondation Jean Dausset, 27 Rue Juliette Dodu, 75010 PARIS, France; 3) Genetic Cancer Susceptibility group, International Agency for Research on Cancer, 15 Cours Albert Thomas, 69372 Lyon Cedex 08, France; 4) INSERM, U895 (équipe 1), Equipe labélisée Ligue Contre le Cancer, C3M, 151 route Saint-Antoine de Ginestiere, BP2, 06204 Nice Cedex 3, France; 5) Université Paris Diderot, Paris 7, Institut Universitaire d'Hématologie, 75010 Paris, France; 6) Fondation Jean Dausset-CEPH, 75010 Paris, France; 7) AP-HP, Hôpital Bichat, Service de Dermatologie, Faculté Paris Diderot, 46 rue Henri-Huchard, 75018 Paris; 8) Department of Dermatology, Centre Hospitalier Universitaire, BP 77908, 21079 Dijon Cedex, France; 9) Lyon 1 University and Centre Hospitalier Lyon Sud, Department of Dermatology, 69495 Pierre Bénite CEDEX, France; 10) Commissariat à l'Energie Atomique, Institut de Génomique, Centre National de Génotypage, 2 rue Gaston Crémieux, CP5721, 91057 Evry CEDEX, France; 11) UMR557 INSERM, U1125 INRA, CNAM, Paris 13, CRNH Idf, Bobigny, France; 12) Université d'Evry Val d'Essonne, Evry, France; 13) Génétique Oncologique EPHE-INSERM U753, Faculté de Médecine, Université Paris-Sud 11 and Institut de Cancérologie Gustave Roussy, 114 rue Edouard Vaillant, 94805 Villejuif Cedex, France; 14) AP-HP, Hôpital Cochin -Tarnier, Service de Dermatologie et Faculté Paris Descartes, 89 rue d'Assas, 75006, Paris, France; 15) Department of Dermatology, Centre Hospitalier Universitaire, Hôpital Archet 2, BP3079, 06200 Nice Cedex 3, France; 16) Present address: Genetic Department, Institut Paoli Calmettes, 232 Bd St Marguerite, 13273 Marseille Cedex 9, France.

Patients affected with melanoma or renal cell carcinoma (RCC) show an excess of second primary malignancies, including RCC for melanoma patients and melanoma for RCC patients. A recent study of the characteristics of coexisting melanoma and RCC in the same patients supports a genetic predisposition underlying the association between these two cancers. The microphthalmia associated transcription factor (MITF) acts as a melanoma oncogene; it also stimulates the transcription of hypoxia inducible factor (HIF1A), whose pathway is targeted by kidney cancer susceptibility genes. By sequencing the MITF gene, we detected a germ line missense substitution (p.E318K) in 5 of 62 patients affected with melanoma and RCC. When compared with 1,659 controls, this MITF substitution occurred at a significantly higher frequency in melanoma +RCC patients (p = 1.3x10-4), melanoma -only patients (p = 7.8 x10-5), and RCC-only patients (p = 0.008). Overall, p.E318K substitution carriers had a higher than fivefold increased risk of developing melanoma, RCC or both cancers (odds-ratio = 5.55, 95% confidence interval = 2.59 to 12.91). Codon 318 falls in the second MITF-conserved, small-ubiquitin like modifier (SUMO)-1 consensus binding site (? KXE). The p.E318K substitution severely impairs binding of small-ubiquitin like modifier (SUMO) to MITF, resulting in increased transcription of HIF1A, and leading to constitutive signalling. This substitution is more potent than wild type MITF in promoting melanocytic and renal cell growth, consistent with a gain-of-function role in tumorigenesis. Our data provide novel insights on the link between sumoylation and cancer, possibly mediated by oxidative stress, and opens novel strategies for cancer prevention.

135

Melanoma exome sequencing identifies MEK1/2 as additional driver genes and potential predisposing variants in genes of DNA repair pathways. *S. Nikolaev[1], D. Rimoldi[2], C. Iseli[2,3], A. Valsesia[2,3,4], C. Gehrig[1], K. Harshman[5], M. Guipponi[1], O. Bukach[6], V. Zoete[3], O. Michielin[3], K. Muehlethaler[2], D. Speiser[2], D. Robyr[1], J.S. Beckmann[7,8], I. Xenarios[3], T.D. Halazonetis[9], C.V. Jongeneel[2,3,10], B.J. Stevenson[2,3], S.E. Antonarakis[1].* 1) GeDev, University of Geneva, Geneva, Switzerland; 2) Ludwig Institute for Cancer Research, University of Lausanne, Lausanne, Switzerland; 3) Swiss Institute of Bioinformatics, CH-1015, Lausanne, Switzerland; 4) Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland; 5) Lausanne Genomic Technologies Facility, Center for Integrative Genomics, CH-1015, Lausanne, Switzerland; 6) Department of Microbiology and Molecular Medicine, University of Geneva, Geneva, Switzerland; 7) Department of Medical Genetics, University of Lausanne, Switzerland; 8) Service of Medical Genetics, CHUV, Lausanne, Switzerland; 9) Department of Molecular Biology, University of Geneva, Geneva, Switzerland; 10) National Center for Supercomputing Applications and Institute for Genomic Biology, University of Illinois, Urbana, IL.

Exome sequencing was performed to detect somatic mutations in protein-coding regions of seven melanoma cell lines and their corresponding germline DNA. All samples exhibited high levels of somatic mutations with a hallmark of exposure to UV light, which was missing amongst clone-specific mutations in two metastases derived from the same patient. Two melanomas with non-canonical BRAF mutations harbored gain of function MEK1/2 mutations. Screening of a larger cohort of melanoma patients revealed the presence of recurring MEK1/2 mutations with an overall frequency of 8%. Missense and nonsense somatic mutations were recurrently found in three novel candidate melanoma genes (FAT4, LRP1B and DCS1). Remarkably, all melanoma patients had potentially damaging rare germline variants in genes involved in DNA repair pathways.

136

Genome-Wide Association Study of Melanoma Progression and Blood Biomarkers. *S. Fang[1], L. Wang[1], J. Gershenwald[2], W. Chen[1], S. Vattathil[1], C. Schacherer[2], J. Gardner[2], Y. Wang[2], T. Bishop[3], J. Barrett[3], E. Grimm[4], C. McHugh[5], C. Laurie[5], K. Doheny[6], E. Pugh[6], Q. Wei[1], C. Amos[1], J. Lee[2].* 1) Epidemiology, UT M.D Anderson Cancer Cente, Houston, TX. 1515 Holcome Blvd,Houston, Tx, 77030, USA; 2) Surgical Oncology,UT M.D Anderson Cancer Cente, Houston, TX. 1515 Holcome Blvd,Houston, Tx, 77030, USA; 3) Section of Epidemiology & Biostatistics Leeds Institute of Molecular Medicine, University of Leeds; 4) Experimental Therapeutics, The University of Texas M. D. Anderson Cancer Center, 1515 Holcome Blvd,Houston, Tx, 77030, USA; 5) 5Department of Biostatistics, University of Washington, Seattle, WA; 6) 6Center for Inherited Disease Research, School of Medicine, Johns Hopkins University, Baltimore, MD.

Melanoma is a rare, albeit potentially highly aggressive malignant tumor of melanocytes, accounting for the majority of deaths from skin cancer. Standard clinicopathological features (eg, primary tumor thickness, ulceration, sentinel lymph node status) cannot completely predict which patients will recur. For those who do recur, current therapies are effective for only a minority of patients. Therefore, further identification of the role of tumor-associated biomarkers and their genetic variation in melanoma progression is needed. In order to accomplish this, we conducted a genome-wide association analysis of melanoma prognosis using samples and data from 1804 melanoma cases from M.D. Anderson Cancer Center and 1026 age and sex-matched healthy controls. We evaluated the influence of genetic factors on Breslow tumor thickness and plasma levels of the inflammatory biomarkers interleukin (IL)-12p40, and the relationship of tumor thickness and IL-12p40 levels to disease progression. Tumor thickness was found to be associated with both disease-free survival and overall survival. Elevated IL-12p40 was discovered to predict poorer overall survival. We identified association of MUC2(rs12365253, P=5.62×10-8) and CALR/RAD23A(rs1049481, P=5.83×10-7) genes with tumor thickness, and strong association of EBF1(rs6895454, P=7.65×10-11) and IL12B (SNP5-158735664, P=1.49×10-19) genes with IL-12p40 level. Our results further demonstrated evidence associating these genes with disease progression. These findings will help to elucidate the mechanism of melanoma progression, define high-risk groups for adjuvant therapy, and provide new clinical indicators for response to therapy over time.

137

Discovery of novel recurrent mutations in childhood early T-cell precursor acute lymphoblastic leukaemia by whole genome sequencing. *J. Zhang[1], L. Ding[15,16], L. Holmfeldt[2], G. Wu[1], S.L. Heatley[2], D. Payne-Turner[2], J. Easton[3], X. Chen[1], J. Wang[4], M. Rusch[1], C. Lu[15,16], J.R. Collins-Underwood[2], J. Ma[2], S.B. Pounds[5], M. Kleppe[8], J. Cools[8], M.L. Hermiston[9], K.A. Shimano[9], G. Basso[10], S.P. Hunger[11], M.L. Loh[9], B. Wood[12], S. Winter[13], K.P. Dunsmore[14], R.S. Fulton[15,16], L. Fulton[15,16], C.-H. Pui[6], R.K. Wilson[15,16,17], J.R. Downing[2], C.G. Mullighan[2], St Jude Children's Research Hospital - Washington University Pediatric Cancer Genome Project.* 1) Computational Biology, St Jude Children's Research Hospital, Memphis, TN; 2) Department of Pathology, St Jude Children's Research Hospital, Memphis, TN; 3) Pediatric Cancer Genome Project, St Jude Children's Research Hospital, Memphis, TN; 4) Department of Information Sciences, St Jude Children's Research Hospital, Memphis, TN; 5) Department of Biostatistics, St Jude Children's Research Hospital, Memphis, TN; 6) Department of Oncology, St Jude Children's Research Hospital, Memphis, TN; 7) Department of Pharmaceutical Sciences, St Jude Children's Research Hospital, Memphis, TN; 8) Department of Molecular and Developmental Genetics, Center for Human Genetics, VIB, K.U.Leuven, Leuven, Belgium; 9) Department of Pediatrics, University of California School of Medicine, San Francisco, CA, US; 10) Onco-Hematology Laboratory, Department of Pediatrics, University of Padua, Italy; 11) Section of Pediatric Hematology/Oncology/Bone Marrow Transplantation and Center for Cancer and Blood Disorders, University of Colorado Denver School of Medicine, The Children's Hospital, Aurora, CO, US; 12) Department of Laboratory Medicine, Seattle Children's Hospital, Seattle, WA, US; 13) University of New Mexico, Albuquerque, NM, US; 14) Pediatric Hematology Oncology, University of Virginia, Charlottesville, VA, US; 15) The Genome Institute at Washington University, St Louis, MO, US; 16) Department of Genetics, Washington University School of Medicine, St Louis, MO, US; 17) Siteman Cancer Center, Washington University School of Medicine, St Louis, MO, US.

Acute lymphoblastic leukaemia (ALL) with early T-cell precursor (ETP) immunophenotype is a highly aggressive subtype of childhood leukaemia. To gain insights into the genetic basis of this disease, we performed whole genome sequencing of tumour and normal DNA from 12 children with ETP ALL and assessed the frequency of the identified non-silent somatic mutations in a separate cohort of 52 ETP and 42 non-ETP T-ALL samples. This identified a high frequency of activating mutations in genes regulating cytokine receptor and RAS signalling (NRAS, KRAS, FLT3, IL7R, JAK3, JAK1 and BRAF), and inactivating mutations disrupting haemopoietic and lymphoid development in ETP ALL (EP300, ETV6, GATA2, GATA3, IKZF1, RUNX1). The IL7R mutations were located in the transmembrane of the receptor, and are activating in vitro. GATA3 alterations were commonly biallelic and observed exclusively in ETP ALL. We also identified recurring mutations in genes not previously implicated in leukaemogenesis, including DNM2, ECT2L and RELN. Ten of 12 ETP ALL cases harboured chromosomal rearrangements, several of which involved complex multichromosomal translocations and resulted in the expression of chimeric in-frame novel fusion genes disrupting haemopoietic regulators, including ETV6-INO80D, NAP1L1-MLLT10 and RUNX1-EVX1. These results indicate that mutations that drive proliferation and impair differentiation are hallmarks of this disease. This mutational spectrum recapitulates that commonly observed in acute myeloid leukaemia, and suggests that additional myeloid-directed therapies might improve the currently poor outcome of ETP ALL.

138

GATA2 is a New Predisposition Gene for Familial Myelodysplastic Syndrome (MDS) and Acute Myeloid Leukemia (AML). *H.S. Scott[1,2,3,4], C.N. Hahn[1,2,3], C.E. Chong[1,2,3], C.L. Carmichael[4], E.J. Wilkins[4,5], P.J. Brautigan[1,2], X.C. Li[1,2], M. Stankovic[1,2], M. Lin[1,2], A. Carmagnac[4], Y.K. Lee[1,2], C.H. Kok[6,7], L. Gagliardi[1], K.L. Friend[8], P.G. Ekert[9], C.M. Butcher[6,7], A.L. Brown[2,7], I.D. Lewis[2,3,7], L.B. To[2,3,7], A.E. Timms[10], J. Storek[11], S. Moore[1], M. Altree[12], R. Escher[4,13], P.G. Bardy[7], G.K. Suthers[12,14], R.J. D'Andrea[2,3,6,7], M.S. Horwitz[10].* 1) Dept of Molecular Pathology, SA Pathology (IMVS), Adelaide, SA, Australia; 2) Centre for Cancer Biology, Adelaide, SA, Australia; 3) School of Medicine, University of Adelaide, SA, Australia; 4) Molecular Medicine Division, Walter and Eliza Hall Institute of Medical Research, Parkville, Vic, Australia; 5) Neurogenetics Laboratory, Howard Florey Institute, Parkville, VIC, Australia; 6) Department of Haematology and Oncology, The Queen Elizabeth Hospital, Woodville, SA, Australia; 7) Department of Haematology, SA Pathology, Adelaide, SA, Australia; 8) Department of Paediatric and Reproductive Genetics, SA Pathology, Adelaide, SA, Australia; 9) Cell Biology Laboratory, Children's Cancer Centre, Murdoch Children's Research Institute, Parkville, VIC, Australia; 10) Department of Pathology, University of Washington School of Medicine, Seattle, WA, USA; 11) Department of Medicine, University of Calgary, Calgary, Alberta, Canada; 12) SA Clinical Genetics Service, SA Pathology, Adelaide, SA, Australia; 13) Medical Clinic, Regional Hospital Emmental, Burgdorf, Switzerland; 14) Department of Paediatrics, University of Adelaide, Adelaide, SA, Australia.

Autosomal dominant familial predisposition to MDS/AML can be mediated by mutations in the transcription factor genes, RUNX1 and CEBPA. Using candidate gene sequencing, we have identified novel domain specific heterozygous mutations in the GATA2 gene in 4 MDS/AML families. An identical novel T354M mutation was observed in 3 families and a 355delT mutation in 1 family, all of which segregated with multigenerational transmission of MDS and/or MDS/AML. Haplotype mapping suggested that the T354M mutation has multiple ancestral origins. The 2 mutated adjacent Thr residues are in 5 consecutive highly conserved Thr residues at the DNA-binding, protein-protein interacting second zinc finger (ZF2) of GATA2. Pathogenic GATA2 coding mutations were not seen in 695 normal individuals nor were they found in 268 sporadic MDS/AML patient samples or germline samples from 35 other families predisposed to AML and various other hematological malignancies. Structural modeling demonstrated that replacement of the Thr for Met at residue 354, while not contacting DNA, potentially disturbs the integrity of ZF2. In contrast, 355delT shortens the conserved Thr string, likely impacting on the orientation and position of L359 which directly contacts DNA. Consistent with this, DNA binding was reduced for T354M and almost completely abolished for 355delT. Reporter assays indicated that T354M and 355delT had greatly reduced transactivation ability on response elements of GATA2 target genes. Of note, T354M and 355delT showed a markedly reduced synergistic effect with PU.1 on the CSF1R promoter. Competition assays demonstrated that these mutations may be acting in a dominant negative fashion in some contexts. In stable promyelocytic HL-60 cell lines expressing regulatable GATA2 (WT, T354M or 355delT), T354M alone inhibited differentiation and apoptosis while allowing proliferation to proceed even under the powerful stimulus to differentiate with all-trans retinoic acid. Analysis of microarray expression studies implicates perturbation of key hematopoietic transcription factor signatures for T354M and 355delT. Discovery of GATA2 mutants in MDS/AML predisposed families provides new tools for probing the mechanism of GATA2-induced leukemogenesis, and possibly for clarifying its role in “stemness” maintenance of haematopoietic stem cells. It may also facilitate more effective diagnosis, prognosis, counselling, selection of related bone marrow transplant donors, and development of therapies.

139

Treating Stiff Skin Syndrome: Study Of A Rare Mendelian Disorder Reveals Novel Therapeutic Strategies for Complex Acquired Scleroderma. *E.E. Gerber[1], D. Huso[2], B. Loeys[3], E. Davis[4], F. Wigley[5], H.C. Dietz[1, 6].* 1) Molecular Biology & Genetics, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine, Baltimore, MD, USA; 3) Center for Medical Genetics, Ghent University Hospital, Ghent 9000, Belgium; 4) Department of Anatomy and Cell Biology, McGill University, Montreal, Quebec, Canada H3A 2B2; 5) Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; 6) Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

To gain insight into the pathogenesis of a common but complex presentation of scleroderma, Systemic Sclerosis (SSc), we studied a rare inherited form called stiff skin syndrome (SSS) that shows childhood onset of diffuse skin fibrosis. We previously reported that all families with SSS harbor heterozygous missense mutations in the 4th 8-cysteine (8-cys) domain of fibrillin-1; mutations elsewhere in fibrillin-1 (including the identical mutations in other 8-cys domains) cause Marfan syndrome (MFS), a condition with no phenotypic overlap with SSS. Given that the 4th 8-cys domain harbors the only Arg-Gly-Asp (RGD) motif through which fibrillin-1 binds cell-surface receptors called integrins, we hypothesized that profibrotic programs may be initiated by failure of integrin mediated events. In keeping with this hypothesis, we now show full recapitulation of dense dermal fibrosis by 3 months of age in mouse lines harboring either a SSS-associated mutation (W1570C) or one that substitutes RGE for RGD (leading to an obligate loss of integrin binding). Dermal fibrosis associates with increased deposition of fibrillin-1 and increased concentration of the profibrotic cytokine TGF in both people and mice with SSS; identical findings are also observed in biopsies from patients with acquired SSc. Cells and tissues in both conditions show increased surface expression of 5 1 and v 3 integrins, suggesting an attempt at compensation in cells that can no longer sense matrix ligands (i.e. fibrillin-1) via integrins. Both of these integrins have been reported to activate TGF . To determine if failed integrin signaling is indeed the critical initiating event, we treated both W1570C and RGE mice with a 1-integrin activating antibody and observed full prevention of skin fibrosis in association with normalization of integrin expression. This treatment also normalized the fibrotic synthetic repertoire of SSc fibroblasts. In keeping with a contribution of increased TGF signaling, both W1570C and RGE mice showed full reversal of established skin fibrosis upon treatment with a panspecific TGF neutralizing antibody. These data are consistent with a model in which perturbation of integrin-matrix ligand interaction initiates skin fibrosis, perhaps in an attempt of cells to reconstitute a matrix that they can no longer sense. Elucidation of the pathogenesis of a rare Mendelian disorder (SSS) has revealed two novel therapeutic strategies for a more common but complex disease (SSc).

140

A liver-specific transgenic mouse model identifies new disease-associated biomarkers and establishes antioxidants as an ameliorative treatment for the renal disease of methylmalonic acidemia (MMA). *E. Manoli[1], J.R. Sysol[1], L. Li[2], C. Garone[3], S. Young[4], J.L. Sloan[1], R.J. Chandler[1], V. Hoffmann[5], P. Zerfas[5], S. DiMauro[3], J. Schnermann[2], C.P. Venditti[1].* 1) Genetics and Molecular Biology Branch, NHGRI, NIH, Bethesda, MD; 2) Kidney Disease Branch, NIDDK, NIH, Bethesda, MD; 3) Columbia University Medical Center, College of Physicians & Surgeons, New York, NY; 4) Division of Medical Genetics, Duke University Medical Center, Durham, NC; 5) Division of Veterinary Resources, ORS, NIH, Bethesda, MD.

We generated mice that express methylmalonyl-CoA mutase (*Mut*) cDNA under the control of a liver-specific promoter on a knockout background (*Mut*$^{-/-}$;Tg$^{INS\text{-}Alb\text{-}}$*Mut*) to model extrahepatic manifestations, study pathophysiology and examine therapeutic interventions. Low-level hepatic *Mut* expression conferred complete rescue from the neonatal lethality displayed by *Mut*$^{-/-}$ mice and allowed disease-associated renal pathology to be induced with a high-protein diet. Ingestion of a high-protein chow for 2 months resulted in elevated plasma methylmalonic acid levels (μM) in the *Mut*$^{-/-}$;Tg$^{INS\text{-}Alb\text{-}}$*Mut* mice (1500±620) compared to similarly-treated *Mut*$^{+/-}$ littermates (7.4±0.6), growth failure and increased mortality. *Mut*$^{-/-}$;Tg$^{INS\text{-}Alb\text{-}}$*Mut* mice developed tubulointerstitial nephritis associated with a decreased glomerular filtration rate (GFR) [37.6±3.9% of *Mut*$^{+/-}$ GFR, p<0.0001] and elevated creatinine levels. Mitochondria of proximal tubular epithelial cells were enlarged and had shortened cristae; kidney immunohistochemistry showed increased succinate dehydrogenase and decreased cytochrome c oxidase staining. Expression analysis using whole kidney RNA from the protein-challenged *Mut*$^{-/-}$;Tg$^{INS\text{-}Alb\text{-}}$*Mut* mice compared to age, sex and diet matched littermates revealed differentially expressed mRNAs from several pathways including immune response, lipid metabolism, ketone biosynthesis and cell survival. One significantly up-regulated gene encoding a secreted glycoprotein was also increased in the plasma and urine of *Mut*$^{-/-}$;Tg$^{INS\text{-}Alb\text{-}}$*Mut* mice and the concentration correlated inversely with GFR (r=-0.45; p<0.01). This protein was further studied in the plasma of 46 *mut* MMA patients (NCT00078078); levels were strongly associated with renal indices (patient plasma creatinine and cystatin-C values; p<0.01 for both). The inclusion of ubiquinol and vitamin E in the high protein diet for 2 months ameliorated the loss of GFR in the *Mut*$^{-/-}$;Tg$^{INS\text{-}Alb\text{-}}$*Mut* mice (37.6±3.9% [pre-] vs. 60±4.8% of *Mut*$^{+/-}$ GFR [post-treatment], p<0.01) and normalized urinary and plasma disease-associated biomarkers despite persisting metabolite elevations induced by the dietary challenge. This novel mouse model has allowed identification of new biomarkers using genomic approaches and provided evidence for a therapeutic effect of antioxidants on the renal disease of MMA. These results should have broad extension to other metabolic disorders manifesting mitochondrial dysfunction.

141

AAV-mediated gene transfer to skeletal muscle results in sustained reduction of hyperbilirubinemia in an animal model of Crigler-Najjar syndrome type 1. *N. Pastore[1], R. Sepe[1], E. Nusco[1], F. Vetrini[1], A. Auricchio[1,2], N. Brunetti-Pierri[1,2].* 1) Telethon Institute of Genetics and Medicine, Naples, Italy; 2) Dept. Pediatrics Federico II University of Naples, Italy.

Crigler-Najjar syndrome is an autosomal recessive disorder of bilirubin metabolism presenting with severe unconjugated hyperbilirubinemia due to deficiency of the liver-specific uridine diphospho-glucuronosyl transferase 1 A1 (UGT1A1). Current therapy relies on phototherapy to prevent life-threatening elevations of serum bilirubin levels but liver transplantation is the only permanent treatment. Given the mortality and morbidity related to the transplant procedures, there is high motivation at developing gene therapy for this disorder. Although correction of the deficient enzymatic activity in the affected organ, i.e. the liver, would be most straightforward, expression within an ectopic tissue to clear toxic metabolites from the circulation is very attractive. The muscle is the preferred tissue for this goal because of its simple and safe access through intramuscular injections (IM). Moreover, the IM route has been investigated extensively for gene therapy of various diseases and in human clinical trials as well. In this study, we have investigated the efficacy of muscle-directed gene therapy for Crigler-Najjar syndrome type 1 using Adeno Associated Viral (AAV) vectors. Serotypes 1 and 8 of an AAV vector expressing the UGT1A1 under the control of the muscle-specific creatine kinase (MCK) promoter were injected at the dose of 3x10e12 genome copies/kg into the muscles of one month-old Gunn rats, the animal model of Crigler-Najjar syndrome type 1. IM injections of both AAV vectors resulted in the expression of functionally active UGT1A1 enzyme in the muscle as demonstrated by Western blot analysis and enzyme assay on muscle tissues. AAV-injected Gunn rats showed a 50-75% reduction of baseline serum bilirubin levels by 3 weeks post-injection which were sustained for at least 1 year post-injection. Taken together, these data show that clinically relevant and sustained reduction of serum bilirubin levels can be achieved by simple and safe IM injections in the Gunn rats. AAV-mediated muscle directed gene therapy has potential for the treatment of patients with Crigler-Najjar syndrome type 1.

142

Antisense oligonucleotide-mediated exon skipping restores primary cilia assembly in fibroblasts harbouring the common LCA CEP290 c.2991+1655G>A mutation. *X. GERARD[1,2], I. PERRAULT[3,4], S. HANEIN[3,4], E. SILVA[5], K. BIGOT[6], S. DEFOORT-DELHEMMES[7], M. RIO[3,4], A. MUNNICH[3,4], D. SCHERMAN[2], J. KAPLAN[3,4], A. KICHLER[1,2], J.-M. ROZET[3,4].* 1) GENETHON, Evry, France; 2) CNRS UMR8151-INSERM U1022, Paris Descartes University, Paris, France; 3) Genetics and epigenetics of metabolic, sensorineural diseases and birth defetcs, INSERM U781-Paris Descartes University & Institute of genetic diseases IMAGINE, Necker - Enfants Malades University Hospital, Paris, France; 4) Genetics Department, Necker - Enfants-Malades University Hospital, Paris, France; 5) Centre for Herediatry Eye Diseases, Department of Ophthalmology, University Hospital of Coimbra, Portugal; 6) Centre d'Exploration et de Ressoures en Thérapeutique Ophtalmologique (CERTO), Necker - Enfants Malades University Hospital University Hospital, Paris, France; 7) Service d'Exploration de la Vision et Neuro-Ophtalmologie, University Hospital of Lilles, France.

Purpose: Leber congenital amaurosis (LCA) is a severe hereditary retinal dystrophy responsible for congenital or early-onset blindness. The most common disease-causing mutation (>10%) is located deep in intron 26 of the CEP290 gene (c.2991+1655 A>G) where it creates a strong splice donor site that leads to the insertion of a cryptic exon encoding a premature stop codon. The aim of this study was to assess the feasibility of an antisense oligonucleotide (AON)-mediated exon skipping strategy to correct this aberrant splicing. Material and Methods: Fibroblast cell lines of patients harbouring the c.2991+1655 A>G mutation (n=4, 3/4 homozygous) and controls (n=3) were transfected using antisense and sense 2'O-methyl phosphorothioate-modified oligonucleotides designed to target exon splicing enhancer (ESE) around the mutation. The skipping was optimized for oligonucleotide sequences and concentrations, transfection conditions and treatment time. The efficiency of skipping was assessed using qRT-PCR, Western blot analysis using a polyclonal antibody recognizing the C-terminus of the CEP290 protein and primary cilia counting. Results and discussion: The level of expression of CEP290 messengers was unchanged when control cell lines were transfected using the sense ONs or AONs ($p>0.05$). Likewise, no change in expression was noted when patient's cells were treated with the sense ONs ($p>0.05$). Conversely, a highly significant increase in expression of the wildtype CEP290 allele was obtained when cells were treated with AONs ($0.029<p<0.002$) with expression levels reaching that of controls. Western blot analysis evidenced increased levels of CEP290 in patients' cell lines treated with the AONs but not the sense ONs. Finally, following serum-starvation, primary cilia expression was significantly reduced in patient's fibroblasts compared to controls lines (48.6%± 6.5% vs 83.6%±3.2 %; $p=0.0097$). Upon transfection with the antisense ONs but not the sense versions, the proportion of ciliated cells increased significantly in patients, reaching levels similar to controls (75.3%±3.5% vs 78.3%±3.4%; $p=0.624$). This suggests that antisense ON-mediated exon skipping resulted in a significant improvement of cilia assembly and/or maintenance. CEP290 mutations are the most common cause of LCA. The results of this study show therapeutic potential of exon skipping for the treatment of the mutation c.2291+1655A>G which alone accounts for 10% of LCA cases.

143

Interference of myostatin and TGF-beta signaling by antisense-mediated exon skipping as an adjunctive therapy for Duchenne muscular dystrophy. *P.A. Hoen[1], D.U. Kemaladewi[1,2], W.M. Hoogaars[1], S.H. van Heiningen[1], A. Aartsma-Rus[1], P. ten Dijke[2], G.J. van Ommen[1].* 1) Human Genetics, Leiden University Medical Center, Leiden, Netherlands; 2) Molecular and Cellular Biology, Leiden University Medical Center, Leiden, Netherlands.

Duchenne muscular dystrophy (DMD) is a severe neuromuscular disorder, which is characterized by the loss of muscle fibers and replacement by fibrotic tissue due to the lack of the dystrophin protein. Myostatin and Transforming Growth Factor (TGF)-beta play important roles in regulation of muscle differentiation and fibrosis, and are emerging as attractive therapeutic targets for DMD treatment. The most promising therapy for DMD currently in clinical trials is antisense oligonucleotide (AON)-mediated exon skipping to restore the reading frame in the DMD gene and correct the primary genetic defect. We investigated if additional exon skipping to knockdown the myostatin/TGF-beta type 1 receptors Acvr1b (ALK4) and Tgfbr1 (ALK5) could be used to treat the secondary disease pathology. Transfection of ALK4 or ALK5 AONs to cultured muscle cells results in skipping of the targeted exons, ~50% downregulation of the full length transcript and enhanced myoblast differentiation. Local administration into the muscles of dystrophin-deficient mdx mice results in specific knockdown of full length ALK4 (~80%) or ALK5 (~50%) expression, leading to 1.5-fold increased expression in marker genes for muscle differentiation and 2.5-fold decrease in expression of fibrotic markers. Where it was previously thought that myostatin signals via both ALK4 and ALK5, we now demonstrate that myostatin uses exclusively ALK4 in muscle cells. Since ALK5 is the predominant TGF-beta receptor in muscle, we are able to specifically modulate myostatin or TGF-beta signaling in the muscle. We further evaluated the therapeutic potential of the ALK4 and ALK5 AONs by systemic administration in mdx mice. We show a decrease in creatine kinase activity in the serum, a biomarker for muscle integrity, and normalization of gene expression patterns, with most pronounced effects in mice treated with the combination of ALK4 and ALK5 AONs. Thus, ALK4 and ALK5 AONs show promise as therapeutic agents for DMD, preferably in combination with DMD AONs.

144

Novel Nitric Oxide-Directed Treatment of Urea Cycle Disorder for Morbidity Independent of Hyerammonemia. *A. Erez[1,2], S. Nagamani[1], M. Premkumar[1,2], P. Campeau[1], W. Mitch[1], L. Salviati[4], N. Bryan[3], B. Lee[1,2,5].* 1) Baylor College of Medicine, Houston, TX; 2) Texas Children Hospital, Houston, TX; 3) Brown Foundation Institute of Molecular Medicine, University of Texas - Houston Health Science Center; 4) Department of Pediatrics, University of Padova, Padova, Italy; 5) Howard Hughes Medical Institute, Houston, TX.

In humans, the urea cycle disorder argininosuccinic aciduria (ASA) caused by deficiency of argininosuccinic acid lyase (ASL), shows systemic features that in part reflect global dysregulation of nitric oxide (NO) homeostasis. In spite of early detection and optimal treatment, many patients develop long term complications and the currently available diagnostic assays cannot predict clinical outcome. One potential explanation for this may be our recent finding that ASL protein is essential outside its catalytic role, for a novel complex assembly necessary for utilizing both endogenous as well as exogenous arginine for NO synthesis. This led us to hypothesize that supplementing ASA patients with NO donors would correct the NO deficiency and reverse the morbidity caused by the NO deficit. We studied 3 ASA patients for their tissue specific response to NO donors by performing vascular reactivity studies. We found a dramatic decrease in the patients' response to flow mediated dilatation, a NO dependent process but no difference in their response to NO donors as compared to controls. These results imply that with ASL deficiency there is a tissue autonomous requirement for NO supplementation while the downstream NO response pathway is intact. To further evaluate this requirement and to understand the effect of genotype, we expressed catalytic site mutants in ASA null cells. Supporting our in vivo results, cells with catalytic site mutations in ASL, showed a significant increase in NO production with exogenous arginine as compared to null cells, supporting the distinct structural role of ASL in arginine utilization for NO synthesis. As a proof of principle, we treated a hypertensive ASA patient who was unresponsive to combination antihypertensive therapy, with Isosorbide dinitrate and then sodium nitrite, an NOS independent NO sources. The patient showed a dramatic clinical response with normalization of his blood pressure over a year treatment allowing weaning of all other antihypertensive medications. Our work dissects the requirement of ASL as a catalytic enzyme necessary for endogenous arginine production, from its requirement as a protein in the complex utilizing arginine for NO production. These data support prospective long term controlled studies of NO sources for the long term morbidity associated with ASA that is independent of hyperammonemia.

145

Correction of cystine and sialic acid storage in human cystinotic and ISSD fibroblasts by microvesicles derived from Sf9 cells infected with baculovirus containing the human sequence for cystinosin or sialin. *J. Thoene[1], M. Witcher[1], J. Mullet[1], P. Courtoy[2], F. N'Kuli[2], P. Van Der Smissen[2], S. Hahn[3], S. Kerfoot[3].* 1) Pediatric Genetics, Univ Michigan, Ann Arbor, MI; 2) Cell Biology Unit,de Duve Institute, UCL Brussels, BE; 3) Seattle Children's Hospital Research Institute Seattle, WA.

INTRODUCTION: Lysosomal transport disorders include cystinosis ,(defective cystinosin), and ISSD, (defective sialin). Both are transmembrane transport proteins, unstable without a lipid bilayer. Microvesicles can transfer transmembrane proteins between non-adjacent cells.METHODS:Sf9 cells were infected with baculovirus containing the human cystinosin (cBac) or sialin (sBac) sequence. Conditioned media from infected Sf9 cells was sedimented at 140,000 g for 3h, the pellet re-suspended in Ham's F12 , placed on cystinotic or ISSD cells, and at 96h the were cells analyzed for cystine and sialic acid. RESULTS: Ultracentrifugation yielded a pellet that on TEM displayed a collection of vesicles from 50 to 200 nm. Suspension of cBac-derived vesicles in Ham's F12 and addition to cultures of cystinotic fibroblasts, caused 57 ± 21 % cystine depletion in 96h, compared to 27±10%depletion from vesicle-free control Ham's F12 (p =0.03).The cBac vesicle fraction produced no sialic acid depletion in ISSD cells (increase of +19 ±9 %, vesicle free control +14 ± 1% p=0.17). Vesicles from sBac treated Sf9 cells produced 40± 6% depletion in sialic acid content of ISSD fibroblasts compared to an increase of 14 ± 8 % in vesicle-free controls (p=0.04). The sBac vesicles did not cause depletion of cholesterol in NPC 1 cells (16 ± 6%; controls, 21 ± 6%; p=0.3). Cystine depletion persists for at least 192h after a single application of vesicles and after a wash and change to vesicle-free media at 96 h.There is a non-linear dose effect: a 5-fold increase in vesicles yielded a 20% increase in cystine depletion. Vesicles did not accelerate efflux of preloaded 3[H]-mannitol , arguing against increased exocytosis. Human cystinosin was identified by LC/MSMS in the vesicles obtained from cBac infected Sf9 cells (1.19 pmol/mg protein), but not in vesicles from sBac-infected Sf9 cells. Density gradient centrifugation of vesicles showed most of the bioactivity at the top of the gradient (density~1.03); these fractions lack LBPA, a marker for exosomes. CONCLUSIONS:Bioactive sedimentable factors in the baculovirus-infected Sf9 conditioned medium are associated with small, very-low density vesicles, possibly microvesicles. Although horizontal transfer of mRNA cannot yet be excluded, we speculate that fusion of Sf9-derived vesicles bearing protein correcting factors with deficient human fibroblasts leads to lysosomal targeting and partial specific correction of the transport defect.

146

The pathogenesis of hypoglycemia and cardiac disease in long-chain acyl-CoA dehydrogenase KO mice. *S.M. Houten[1], A.J. Bakermans[2], H.J. Herrema[3], H. te Brinke[1], T.H. van Dijk[3], M. van Weeghel[1], D.-J. Reijngoud[3], J.J. Prompers[2], R.J. Wanders[1].* 1) Laboratory Genetic Metabolic Diseases, Depts of Pediatrics and Clinical Chemistry, Academic Medical Center, Amsterdam, The Netherlands; 2) Biomedical NMR, Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands; 3) Laboratory of Pediatrics, University Medical Center Groningen, Groningen, The Netherlands.

Mitochondrial fatty acid) -oxidation (FAO) is a crucial pathway for energy homeostasis in liver, muscle and heart and is of particular importance during fasting. Recessively inherited defects are known for most of the enzymes involved in FAO. The main pathological consequences associated with FAO defects are hypoketotic hypoglycemia, skeletal myopathy and cardiac disease. Mouse models are useful to study the pathologic mechanisms and potential therapeutic approaches. To elucidate the pathogenetic mechanisms in FAO defects, we have identified the major metabolic adaptations to fasting in the long-chain acyl-CoA dehydrogenase (LCAD) KO mouse. LCAD is crucial for long-chain FAO in mice. Mice were studied after short term and overnight fasting. Blood, plasma and organs were collected and used for targeted metabolomics, and protein and gene expression analysis. Hepatic and whole body glucose metabolism was assessed by infusing isotopes. Cardiac function was evaluated using cinematographic magnetic resonance imaging, and cardiac triglyceride (TG) content was measured using localized proton magnetic resonance spectroscopy. We show that the hypoglycemia induced by fasting in LCAD KO mice is caused by an increased glucose requirement in peripheral tissues leading to hepatic glycogen depletion. The fasting state is characterized by an increased metabolic clearance rate and a decreased endogenous production of glucose, leading to hypothermia in LCAD KO mice. Gluconeogenesis was unaffected and importantly could not compensate for the increased glucose demand. Targeted metabolomics illustrated that levels of all gluconeogenic precursors were decreased, suggesting limited release and a potential inhibition of autophagy. In addition, the LCAD KO mouse displays cardiac hypertrophy, which is non progressive. Expression markers for cardiomyopathy are not increased. Myocardial TG content was higher in non-fasted LCAD KO mice and further increased upon fasting. In addition, left ventricular ejection fraction and diastolic filling rate decreased after fasting, whereas these functional parameters did not change in fasted WT mice. The increase in myocardial TG coincided with elevated levels of the potentially more lipotoxic ceramide. We conclude that during fasting, FAO is not only glucose-sparing, as classically defined, but also indispensable for maintenance of energy homeostasis and the prevention of accumulation of potentially lipotoxic intermediates.

147

Pharmacogenomics of paclitaxel-induced cytotoxicity and apoptosis in lymphoblastoid cell lines and paclitaxel-induced peripheral neuropathy in patients from the Cancer and Leukemia Group B 40101 clinical trial. *H.E. Wheeler[1], E.R. Gamazon[2], U.O. Njiaju[1], L.K. Gorsic[1], R.M. Baldwin[3], K. Owzar[4], E. Winer[5], C. Hudis[6], L.N. Shulman[5], M.J. Ratain[1], D.L. Kroetz[3], N.J. Cox[2], M.E. Dolan[1].* 1) Section of Hematology/Oncology, Department of Medicine, University of Chicago, Chicago, IL; 2) Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL; 3) Department of Bioengineering and Therapeutic Sciences, School of Pharmacy and Medicine, University of California, San Francisco, San Francisco, CA; 4) Department of Biostatistics and Bioinformatics, Duke University, Durham, NC; 5) Dana-Farber Cancer Institute, Boston, MA; 6) Memorial Sloan-Kettering Cancer Center, New York, NY.

Paclitaxel is a microtubule-stabilizing chemotherapy drug often used in the treatment of breast, ovarian and non-small cell lung cancers, yet variable resistance and toxicities among individuals limit successful outcomes. We used lymphoblastoid cell lines (LCLs) derived from the HapMap YRI (Yoruba), ASW (African Americans) and CEU (Northern and Western European ancestry) populations to measure paclitaxel-induced cytotoxicity and apoptosis. We have performed genome-wide association (GWA) studies for these two phenotypes using genotype data from the HapMap project. At a threshold of $p < 10^{-3}$, there are 98 SNPs that overlap between cytotoxicity and apoptosis. This overlap is greater than expected by chance as we found an enrichment of apoptosis-associated SNPs in the cytotoxicity-associated SNPs (empirical $p < 10^{-3}$). To better understand the relevance of our LCL model in terms of clinical paclitaxel-induced toxicity, we compared our results to a GWA study of peripheral sensory neuropathy in breast cancer patients treated with paclitaxel from the CALGB 40101 clinical trial. We observed an enrichment of cytotoxicity-associated SNPs in the peripheral neuropathy-associated SNPs from the clinical trial (empirical $p = 4 \times 10^{-3}$). Of the 36 SNPs that overlap between the clinical trial ($p < 0.05$) and the cytotoxicity study ($p < 10^{-3}$), 29 of them are expression quantitative trait loci (eQTLs), which is a significant enrichment of this functional class (empirical $p = 8 \times 10^{-3}$). One of these eQTLs is located in *ABCC1*, which encodes multidrug resistance-associated protein 1, an ATP-binding cassette transporter (cytotoxicity $p = 6.5 \times 10^{-4}$, clinical trial $p = 1.3 \times 10^{-3}$). A second eQTL is located in *UNC13C*, an ortholog of the *C. elegans* gene *unc-13*, which is expressed in neurons throughout the body and regulates neurotransmitter release. The *UNC13C* SNP associated with paclitaxel-induced peripheral neuropathy ($p = 3.8 \times 10^{-2}$), cytotoxicity ($p=4.0 \times 10^{-5}$), and apoptosis ($p = 6.8 \times 10^{-6}$). Our approach identified genes to study in the LCL model system that have support for being important in patient response to paclitaxel. These results provide a framework for choosing which SNPs that come from LCL GWA studies to interrogate in future clinical studies (i.e. eQTLs) of chemotherapeutic-induced toxicities. If these particular findings are validated, the ability to identify cancer patients at risk of peripheral neuropathy could influence treatment choice and outcome.

148

Optimizing drug outcomes through pharmacogenetics: A case for preemptive genotyping. *J. Schildcrout[1,2], J. Denny[3,5], E. Bowton[7], W. Gregg[3,5], J. Pulley[7], M. Basford[7], J. Cowan[8], H. Xu[3], A. Ramirez[5], D. Crawford[4], M. Ritchie[3,4], J. Peterson[3,5,9], D. Masys[3,5], R. Wilke[5,6,7], D. Roden[5].* 1) Department of Biostatistics, Vanderbilt University, Nashville, TN; 2) Department of Anesthesiology, Vanderbilt University, Nashville, TN; 3) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 4) Department of Molecular Physiology & Biophysics and Center for Human Genetics Research, Vanderbilt University, TN; 5) Department of Medicine, Vanderbilt University, Nashville, TN; 6) Department of Pharmacology, Vanderbilt University, Nashville, TN; 7) Office of Personalized Medicine, Vanderbilt University, TN; 8) Vanderbilt Office of Research, TN; 9) VA Tennessee Valley Healthcare System, TN.

Current use of genomic information to guide drug prescribing uses a reactive model: genetic testing is ordered when a drug is prescribed, and medication adjustments, if needed, occur later. This approach misses the opportunity to optimize drug and dose selection when therapy is started (often the time of greatest pharmacogenomic value), and may require multiple encounters between patients and the healthcare system. An alternate model is preemptive, in which multiple genomic variants relevant to drug decision making are routinely integrated into healthcare records prior to and in anticipation of any relevant drug prescription. However, the frequency of opportunities for genetic data reuse on a single patient is unknown. To quantify patient exposure to medications with pharmacogenetic (PGx) associations, we identified a longitudinal cohort of 52,942 patients meeting a medical home definition (≥ 3 outpatient visits within two years following January 1, 2005) at Vanderbilt University Medical Center. We analyzed drug utilization in this patient population for a set of 57 drugs with established PGx associations, as documented primarily by Food and Drug Administration package inserts. Using established adverse event (AE) rates and frequencies of risk alleles influencing outcomes for six medications with severe AEs, we estimated the number of AEs that would have been prevented if genotypes had been available at the time of prescribing and had been used effectively to alter therapy. These six associations were warfarin-related bleeding, recurrent major cardiovascular events on clopidogrel, simvastatin-induced myopathy, breast cancer recurrence during tamoxifen therapy, abacavir hypersensitivity reactions, and azathioprine-induced myelosuppression. The five-year exposure rate to at least one medication with a known PGx AE was estimated to be 64.8% (95% confidence interval: 64.4%-65.2%), and the exposure rate to at least four PGx medications was greater than 10%. Among the six severe AEs studied, we estimated that 449 (95% CI: 275-637) might have been avoided with a preemptive genotyping program coupled to effective intervention strategies. These data support the development of a multiplexed, preemptive genotyping model as an efficient alternative to current methods. Implementation of this model will also facilitate evaluation of the clinical utility of pharmacogenomically-guided drug therapy.

149

A Nationwide Program Combining Test Facilitation for CYP2C19 Genotyping and Drug Utilization Review for Patients on Clopidogrel. *L.A. Castle, W.B. Dreitlein, H. Kourlas, M. Khalid, J.F. Barlow, R.S. Epstein.* Advanced Clinical Science & Research, Medco Health Solutions, Inc., Franklin Lakes, NJ.

BACKGROUND: Patients with certain genetic variations in CYP2C19 (2C19) metabolism or who are on concomitant medications that inhibit enzyme activity, such as proton pump inhibitors (PPIs), experience higher cardiovascular (CV) event rates on clopidogrel. Although a boxed warning in the clopidogrel label addresses testing for 2C19 polymorphisms, genotyping has not been routinely adopted. OBJECTIVE: Assess the uptake and impact of a personalized medicine program that informs physicians and patients about the value of 2C19 genotyping and facilitates access to testing. METHODS: A pharmacy benefit manager (PBM) claims database is used to identify patients prescribed clopidogrel. On physician authorization, patients receive education on 2C19 testing along with a DNA sample collection kit. Genotype (alleles *1-8, *17), metabolism phenotype, and a list of 2C19 inhibitors are reported to physicians. Physicians are notified of 2C19 drug-drug interactions through drug utilization review (DUR). Specially trained CV pharmacists contact physicians to discuss treatment options for patients with poor metabolism phenotypes. RESULTS: Physician characteristics were evaluated using the first 675 test opportunities; a longitudinal assessment of drug therapy by phenotype was completed for the first 100 patients tested between 12/1/10 - 4/30/11. Testing was offered most frequently to cardiologists (45.1%) and primary care physicians (PCPs) (34.6%). Neurologists had the highest acceptance rate (52.6%), followed by PCPs (41.7%) and cardiologists (30.2%). Metabolism phenotype distribution was 1% poor (PM), 29% intermediate (IM), 37% extensive (EM), and 33% ultrarapid (UM); 38% had at least one *17 allele. 37% were on a 2C19 inhibitor before testing (24% PPIs, 6% antidepressants, 7% both). The medication change rate after testing was 27% (22% either discontinuing PPIs or switching to pantoprazole, 5% discontinuing clopidogrel). No changes were made in antidepressant therapy or based on PM phenotype. CONCLUSION: PCPs, who care for a large proportion of patients on clopidogrel, generally supported 2C19 genotyping, while neurologists had the highest acceptance rate. In this analysis, 37% of patients were on a 2C19-inhibiting drug, which may increase the risk of CV events. The ability to combine genotyping with DUR messaging may not completely eliminate this risk, but raising awareness allows physicians to reassess therapy and could enhance care coordination across specialties.

150

Genome-wide genetic determinants of lipid response to rosuvastatin therapy. *D. Chasman[1], F. Giulianini[1], J. MacFadyen[1], B. Barratt[2], F. Nyberg[2,3], P. Ridker[1].* 1) Division of Preventive Medicine, Brigham & Women's Hospital, Boston, MA; 2) AstraZeneca Research and Development, Alderley Park, England and Mölndal, Sweden; 3) Unit of Occupational and Environmental Medicine, Department of Public Health and Community Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.

In statin trials, every 20 mg/dL reduction in cholesterol results in a 10 to 15 percent reduction in vascular event rates. However, the inter-individual variation in LDL-C response to statins is wide and may partially be determined on a genetic basis. A genome-wide association study of the statin response in LDL-C and other lipid fractions was performed among a total of 6,989 men and women of European ancestry who were randomly allocated to either rosuvastatin 20 mg daily or placebo in the JUPITER trial with genotyping from the Illumina Omni 1M platform. Both the absolute and fractional changes in LDL-C and other lipid fractions with treatment allocation were considered. Single nucleotide polymorphisms (SNPs) with genome-wide significant ($P<5x10^{-8}$) association with LDL-C reduction among 3,523 participants on rosuvastatin were identified at *ABCG2*, *LPA*, and *APOE*. A further association at *PCSK9* was genome-wide significant for baseline LDL-C and locus-wide significant for LDL-C reduction. No genome-wide significant associations were observed among 3,466 participants on placebo. The magnitudes of effects of the minor allele at the lead SNPs at the four loci ranged from 4.2-6.2 mg/dL or 4.5-6.8%. These effects increased absolute LDL-C reduction at *ABCG2* and decreased absolute LDL-C reduction at *PCSK9*, *LPA*, and *APOE*. Median absolute LDL-C reductions on rosuvastatin ranged from 40-64 mg/dL with increasing numbers of LDL-C-lowering alleles at the lead SNPs (P-trend = $6.2x10^{-20}$). A parallel genome-wide analysis of ApoB response to rosuvastatin revealed the same loci as were found for LDL-C response, while analysis of HDL-C, ApoA1, and triglyceride response did not reveal any genome-wide significant associations. At sub-genome-wide significance ($P<1x10^{-6}$), an additional association for LDL-C response was identified at *IDOL*, which mediates post-transcriptional regulation of the LDL receptor in response to intracellular cholesterol levels. Within the genome-wide genetic data, SNPs in the *SLCO1B1* and *LDLR* were confirmed as associated with LDL-C lowering, and a significant interaction was observed between SNPs in *PCSK9* and *LDLR*. Thus, common inherited polymorphisms that predominantly relate to statin pharmacokinetics and endocytosis of LDL particles by the LDL receptor influence individual patient response to rosuvastatin therapy. The candidate association at *IDOL* suggests the potential for novel therapeutic approaches to LDL-C lowering.

151

Predicting warfarin dosage in European and African Americans using DNA samples linked to an electronic health record. *A.H. Ramirez[1], Y. Shi[2], J.S. Schildcrout[2], J.T. Delaney[1], H. Xu[3], M.T. Oetjens[4], R.L. Zuvich[4], M.A. Basford[5], E. Bowton[5], M. Jiang[3], R. Zink[3], J. Cowan[6], J.M. Pulley[7], M.D. Ritchie[3,4,8], J.F. Peterson[1,3], D.R. Masys[3], D.M. Roden[1,9], D.C. Crawford[4,8], J.C. Denny[1,3].* 1) Department of Medicine, Vanderbilt University, Nashville, TN; 2) Department of Biostatistics, Vanderbilt University, Nashville, TN; 3) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 4) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 5) Office of Research, Vanderbilt University, Nashville, TN; 6) Institute for Clinical and Translational Research, Vanderbilt University, Nashville, TN; 7) Medical Administration, Vanderbilt University, Nashville, TN; 8) Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN; 9) Department of Pharmacology, Vanderbilt University, Nashville, TN.

Warfarin pharmacogenomic algorithms reduce dosing error, but perform poorly in non-European Americans and remain widely unimplemented. Electronic Health Record (EHR) systems linked to biobanks offer an infrastructure for discovery and implementation of pharmacogenomics, but have not yet been used for this purpose. We used BioVU, the Vanderbilt DNA repository linked to the EHR system, to identify Americans of European (EA) (n=1,022) and African (AA) (n=145) ancestries on stable warfarin therapy as defined by ⁄ 2 in-range INR values over 3 weeks with no out of range values and genotyped 15 SNPs in six genes: CYP2C9, VKORC1, EPHX1, GGCX, CYP4F2, and CALU. In EAs, robust associations were observed between warfarin dosage and CYP2C9*2 (p = 1.4x10-11), CYP2C9*3 (p = 8.7 x 10-25), CYP4F2 rs2108622 (p = 3.1 x 10-6), and three VKORC1 variants in strong LD ($r^2 > 0.99$) ($p < 5.3 \times 10\text{-}58$). However, among AAs, only CYP2C9*3 (p = 0.013) and two VKORC1 variants (p=0.001) replicated. In addition, there were associations in AAs with CYP2C9*6 (p = 0.001), CYP2C9*8 (p = 0.002), and CALU rs339097 (p = 0.02). Modeling the International Warfarin Pharmacogenomics Consortium (IWPC) algorithm and FDA recommendations in silico reduced error between predicted weekly dose and actual stable therapy weekly dose in EAs, but not meaningfully in AAs. The IWPC algorithm explained 55% of dose variation in EAs but only 34% in AAs, while a novel regression algorithm incorporating CYP2C9*6, CYP2C9*8, CYP4F2 rs2108622, and CALU rs339097 in addition to the IWPC variants maintained high performance in EAs explaining 58% of variation and explained over 49% of the variation in AAs. Additionally, in AAs the novel algorithm improved positive predictive value of low dose prediction from 27% to 75% and sensitivity of high dose prediction from 20% to 67% compared to the IWPC algorithm. Data such as these are necessary to successfully implement personalized warfarin dosing in a diverse population. A CLIA approved genotyping pathway with decision support for algorithm-dosed warfarin initiation integrated in the Vanderbilt EHR is under development and prospective outcomes are being followed. EHR-linked DNA biorepositories provide a real-world resource for validation of pharmacogenomic associations, generation of dosing algorithms, and a feasible platform for implementation of knowledge to improve patient safety.

152

Pharmacogenetic genome-wide association analysis in the NORDIL study identifies 9 putative SNPs associated with systolic and diastolic blood pressure response. *K.A. Pettigrew[1], O. Melander[2], C. Newton-Cheh[3], A.F. Dominiczak[1], S. Padmanabhan[1], Glasgow - Malmö Extreme BP Consortium.* 1) Institute of Cardiovascular and Medical Sciences, College of Medical Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom; 2) Department of Clinical Sciences, Hypertension and Cardiovascular Diseases, University Hospital Malmö, Lund University, Malmö, Sweden; 3) Center for Human Genetic Research and Cardiovascular Research Center, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, United States of America.

Background: Essential Hypertension is a heritable and modifiable determinant of the global cardiovascular health burden. Despite availability of many effective antihypertensive drugs, only about 40 percent of treated hypertensives have their blood pressure (BP) controlled. Initial therapy is often selected empirically, and low response rates to any particular antihypertensive drug suggest the current approach to therapy selection and hypertension management is not optimal. Methods: We performed a genome wide association study of antihypertensive drug response to two major classes of antihypertensive agents: beta-blockers (BB) and calcium channel blockers (CCB) in 4200 subjects from the NORdic DILtiazem (NORDIL) Study. This is a prospective, randomized, open-label study in subjects 50-74 years old, with diastolic BP⁄ 100 mmHg. Patients were assigned to two treatment arms: non-dihydropyridine calcium channel blocker, or beta-blockers ± diuretic. Genotyping was performed using the Illumina 610Quad chip and imputation done using NCBI build 36/HapMap r22. Association analysis using linear regression (under an additive genetic model) was performed on the percentage change in systolic and diastolic blood pressure monotherapy with CCB (n=1639) or BB (n=1070) at 6 months post-randomization. Results: Four independent loci reached a p<5x10-6 for BP response to CCB monotherapy - SBP response loci are 1q32 [beta(s.e): 1.99 (0.4); 2.9 x10-06], 2p23 [1.43 (0.3); 4 x10-07] and DBP response loci are 2p11.2 [1.47 (0.3); 4.4 x10-06] and 5p15.2 [2.13 (0.5); 3.9 x10-06]. There were 5 independent loci attaining p<1x10-5 for BP response to BB monotherapy - SBP response 8q11.2 [2.04 (0.4); p=4.6 x10-06], 13q22 [-4.31 (0.9); 2.5 x10-06], 13q22 [2.45 (0.5); 3 x10-07] and DBP response 3p24 [1.5 (0.3); 8.5 x10-06], 8q12 [1.49 (0.3); 6.2 x10-06]. None of these loci have previously been associated with hypertension or blood pressure. Conclusions: Antihypertensive response is a polygenic trait amenable to genetic dissection using GWAS. Validation of the top signals is underway and this will allow identification of novel pathways that can explain the inter-individual variability in BP response to treatment.

153

Personalized medicine study of Chinese population. *S. Qin[1], L. Shen[1], Y. Xiong[1], K. Tang[1], G. Feng[3], L. He[1,2].* 1) Bio-X Institutes, Shanghai Jiaotong University, Shanghai, China; 2) Institute for Nutritional Sciences, Shanghai Institute of Biological Sciences, Chinese Academy of Sciences, Shanghai, China; 3) Shanghai Institute of Mental Health, Shanghai, China.

Personalized medicine targets individualized treatment and care based on personal and genetic variation. It focuses on the best ways to develop new therapies and optimize prescribing by steering patients to the right drug at the right dose at the right time. However, in order to achieve the goals of personalized medicine, many obstacles must be resolved, which include scientific challenges, such as determining which genetic markers have the most clinical significance and conducting clinical studies to identify genetic variation that are correlated with a drug response. It is already known that genetic variation frequencies occur with varying prevalence in populations of different ethnic or geographic origins. Therefore, it is essential to conduct a genetic variants analysis in various populations. Few studies have focused on the Chinese Han population to date. In the present study, we have built sample bank for personalized medicine. Based on the sample bank, we systematically screen the genetic variation of four main DME genes (CYP2D-6,CYP2C19,CYP2C10,CYP2E1) in the four different geographic Chinese population covering the north, south, west and east of the Chinese mainland. We build a database for a normal Chinese population. Also, we found some allele and genotype frequencies demonstrated differences among different geographic populations. We identified many new alleles of the four genes in Chinese population. In order to investigate the phenotypes, we established the procedure for analyzing the functional consequence of these new alleles on enzyme activity using yeast as a heterologous expression system. The new CYP alleles we first identified in Chinese population are accepted by the CYP Allele Nomenclature Committee Also, we are recruiting samples related to drug efficiency and adverse drug reactions through cooperating with large hospitals in China. All the subjects are related to psychiatry disease, cardiovascular disease, cancer etc. We are now doing drug-response-related association study based on the sample bank. Also, we are now developing a SNP genotyping microarray for personalized medicine study of Chinese population, which include about ten thousand SNPs of drug metabolized genes, drug transporter genes and drug target genes. Based on the database we have built, we are developing the diagnostic microarray technology for comprehensive analysis of the four genes that can influence drug efficacy and adverse drug reaction.

154

A single common *TSPO* SNP alters receptor binding of a PET ligand: can genotyping identify patients most likely to benefit clinically from PET studies? *A. Yeo[1], D. Owen[2, 3], R. Gunn[3, 4, 5], K. Song[1], G. Wadsworth[2, 3], A. Lewis[2], C. Rhodes[2], D. Pulford[1], I. Bennacef[3], C. Parker[3, 4], P. StJean[1], L. Cardon[1], V. Mooser[1], P. Matthews[3, 4], E. Rabiner[3, 4], J. Rubio[1].* 1) Genetics, GlaxoSmithKline Research and Development, Stevenage, UK and King of Prussia, PA and Research Triangle Park, NC, USA; 2) Division of Experimental Medicine, Imperial College London, Hammersmith Hospital, London, UK; 3) GlaxoSmithKline Clinical Imaging Centre, Hammersmith Hospital, London, UK; 4) Centre for Neuroscience, Imperial College London, Hammersmith Hospital, London, UK; 5) Department of Engineering Science, University of Oxford, Oxford, UK.

Positron emission tomography (PET) imaging is an important tool frequently used to evaluate inflammation and degeneration in CNS diseases such multiple sclerosis and Parkinson's disease. Expression of the 18kDa Translocator Protein (TSPO) is used in PET studies to detect microglial activation, a hallmark of CNS inflammation. However, quantitative interpretations of signal are confounded by large inter-individual variability in binding affinity of second-generation TSPO radiotracers. Here we tested genetic markers in *TSPO* and genes encoding other proteins in the TSPO complex for association with binding affinity of [^{11}C]PBR28 in platelets isolated from 41 human subjects. Of 58 markers (SNPs & indels) evaluated, 20 markers were polymorphic in our samples and one (rs6971; Ala147Thr) alone was perfectly correlated as a Mendelian trait with binding affinity under a co-dominant model (adjusted p-value=5.7x10^{-10}). The reported observation that variation in PBR28 binding affinity appears consistent across all tissues within the same subject implies that our findings in platelets can likely be extrapolated to brain. PolyPhen analysis predicts that the Ala147Thr amino acid substitution in the 5th transmembrane domain of the TSPO protein results in a conformational change, which would likely affect the interaction between TSPO and other proteins in the TSPO complex. There is published evidence that Ala147Thr has an impact on biological functions of TSPO, such as pregnenolone production and plasma LDL-cholesterol levels. Ala147Thr has also been reported to be associated with separation anxiety. Here we will present the results of further genetic analyses of Ala147Thr in an expanded series of neurological and neuropsychiatric diseases. In conclusion, we have identified a common SNP that is unambiguously associated with PBR28-binding to TSPO in human platelets. This is an unusual example of target-receptor interaction in which a single polymorphism is associated with altered binding affinity. Extrapolating this intermediate phenotype to TSPO PET; screening of subjects for rs6971 may for the first time allow quantitative analysis and accurate interpretation of TSPO PET studies using PBR28 and other second-generation TSPO PET radioligands. Therefore, if replicated in PET imaging studies, these findings represent a significant advance in the field of TSPO PET neuroimaging by providing a genomic biomarker with the immediate potential for translation to the clinic.

155

Reticulon mutations in Hereditary Spastic Paraplegia. *A.P. Rebelo[1], G. Montenegro[1], J. Connell[2], R. Allison[2], R. Schüle[3], T. Deconinck[4], J. Huang[1], M. Pericak-Vance[1], P. Jonghe[4], L. Schöls[3], A. Orlacchio[5], E. Reid[2], S. Zuchner[1], C. Babalini[5], G. Bernardi[5].* 1) HIHG, University of Miami, Miami, FL., Select a Country; 2) Department of Medical Genetics, University of Canbridge, UK; 3) Hertie Institute for Clinical Brain Research and Center of Neurology, University of Tubingen, Germany; 4) Neurogenetics Group, University Hospital Antwerpen (UZA), Belgium; 5) (5) Laboratorio di Neurogenetica, Centro Europeo di Ricerca sul Cervello (CERC), Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Italy.

Hereditary spastic paraplegia (HSP) comprises a genetically heterogeneous neurodegenerative disorder affecting the long axons of the first motoneuron. Approximately 50 chromosomal loci have been mapped for HSP thus far explaining ~60% of all patients. Recent genetic and functional studies have highlighted a number of molecules constituting the machinery that regulates the tubular shape of the endoplasmic reticulum (ER). These proteins include Atlastin (SPG3A), Spastin (SPG4) and REEP1 (SPG31). Another group of proteins involved in ER tubulation are the reticulons. Here we show that a member of the reticulon gene family is a novel HSP gene. We have employed classic linkage analysis and large scale candidate gene screening to identify an identical single nucleotide insertion causing a protein truncation in two unrelated families, both exhibiting a LOD score >3. By screening 250 additional HSP index patients by sequencing and CNV analysis, we also found a complete deletion of the gene and a missense point mutation in this particular reticulon family protein. All changes were absent in more than 1,700 Caucasian control samples. Functional studies showed impaired subcellular localization of the truncated form of the reticulon. The missense mutant; however, co-localized with markers of the ER as did the wild-type protein. Co-immunoprecipitation assays revealed interactions between spastin and the reticulon protein, including the point mutant, supporting a possible role in shaping the ER network. Our data further underline the recently recognized importance of ER tubulation in the pathophysiology of HSP and possibly in other motor neuron diseases.

156

Mutation in C19orf12 causes hereditary spastic paraplegia type 43. *G. Landoure[1,2], B.G. Burnett[1], P-P. Zhu[1], J. Johnson[3], D. Hernandez[3], C. Rinaldi[1], M. Sangare[1], E. Rugarli[4], M. Traore[2], B. Traynor[3], C. Blackstone[1], K. Fischbeck[1].* 1) Neurogenetics Branch, NINDS/NIH, Bethesda, MD; 2) Service de Neuologie, Hopital du Point G, Bamako, Mali; 3) Laboratory of Neurogenetics, NIA/NIH, Bethesda, MD; 4) Biocenter University of Cologne, Germany.

Hereditary spastic paraplegias (HSPs) are inherited neurological disorders characterized by progressive spasticity and weakness with an estimated prevalence of 3-10/100,000. All modes of inheritance are seen, but dominant HSPs are most common. Over 20 recessive HSPs have been described; however, only nine disease genes have been identified to date. We previously reported a family from Mali (West Africa) with two sisters affected by spastic paraplegia with weakness of the lower limbs and marked atrophy of the distal upper extremity muscles. There was no known consanguinity, but the proportion of identity by descent by SNP analysis was higher than expected (0.67-0.77), consistent with parental inbreeding. Homozygosity mapping identified a region of extended homozygosity at chromosome 19p13.11-q12 shared by the affected sisters and not by the other unaffected family members or unrelated controls. Sequencing of candidate genes, including SPG20 and RAB3A, GDF1, FKBP8, and , COP in the homozygous region, was negative. We performed genome-wide exome sequencing in the two affected individuals, and found five non-synonymous homozygous single nucleotide variants in four genes within the previously identified locus. One sequence variant could not be confirmed by Sanger sequencing, and three were seen in controls. The remaining variant is located in the C19orf12 gene at position c.154G>C, predicting the amino change A52P. This residue is conserved across species, and the A52P homozygous variant was not found in about 300 ethnically matched controls. There was no difference in patient C19orf12 mRNA levels in lymphoblastoid cells compared to unaffected family members or normal controls by qRT-PCR. However, immunostaining of transfected cells showed that the mutant C19orf12 localized to the nucleus while the wild-type was present in the cytoplasm. C19orf12 is predicted to be a membrane protein, and the A52 residue is localized in a predicted membrane helix domain. The substitution of alanine to proline appears to disrupt C19orf12 a membrane targeting motif. Thus, we identified a novel gene for recessive HSP that adds to our understanding of the cellular biology and pathogenesis of this group of disorders. Additional investigations, including functional studies of the gene product, are now being undertaken to characterize the role of this protein in the health of the corticospinal tract, which degenerates in HSP.

157

Combination of positional cloning and new generation sequencing identifies 3 novel genes of hereditary spastic paraplegia involved in closely related functions. *G. Stevanin[1,2], C. Tesson[1], M. Nawara[1], M.A.M. Salih[3], M. Zaki[7], E. Mundwiller[1], M. Al Balwi[6], A. Boukhris[9], A. Bouhouche[10], E. Martin[1], A. Al Drees[3], S.A. Elmalik[3], M.M. Kabiraj[4], M.Z. Seidahmed[5], A. Alswaid[6], N. Bouslam[10], L. Orlando[1], F. Mochel[1,2], A. Rastetter[1], A. Durr[1,2], I. Al Abdulkareem[12], M.T. Al Rifai[6], F.M. Santorelli[11], A. Benomar[10], S.A. Al Rasheed[6], C. Mhiri[9], J. Gleeson[8], F. Darios[1], A. Brice[1,2].* 1) Institut du Cerveau et de la Moelle épinière (INSERM / UPMC UMR_S975, CNRS 7225, NEB, EPHE), Pitié-Salpêtrière Hospital, Paris, France; 2) APHP, Dpt of Genetics and Cytogenetics, Pitié-Salpêtrière Hospital, Paris, France; 3) King Saud University, Riyadh, Saudi Arabia; 4) Armed Forces Hospital, Riyadh, Saudi Arabia; 5) Security Forces Hospital, Riyadh, Saudi Arabia; 6) King Abdulaziz Medical City, Riyadh, Saudi Arabia; 7) National Research Institute in Cairo Egypt; 8) University of California, San Diego; 9) Hôpital Habib Bourguiba, Sfax, Tunisia; 10) Hôpital des Spécialités, Rabat, Morocco; 11) IRCCS Fondazione Stella Maris, Calambrone, Pisa, Italy; 12) King Abdullah International Medical Research Center, Riyadh, Saudi Arabia.

Hereditary spastic paraplegias, which encompass a wide range of phenotypes, are clinically and genetically very heterogeneous neurological disorders. Autosomal dominant and recessive as well as X-linked forms have been described, and known genes account for the majority of dominant cases, but for less than 40% of the recessive forms. The combination of classical positional cloning and next generation sequencing in 3 consanguineous families with an autosomal recessive form of these conditions proved very powerful to identify causative mutations. The first two families, from North-Africa, were linked to SPG28 and SPG46 and the disease was shown to segregate with a splice site mutation and a missense mutation affecting a conserved amino-acid, respectively. The third family was linked to a new locus on chromosome 4 (SPG47). The missense mutation found in this family was also identified in 2 other families of the same origin, Saudi Arabia, while a frameshift mutation was identified in the same gene in a fourth kindred from Egypt. All mutations were absent in large series of matched controls. The 3 new genes were involved in close metabolic pathways related to lipids, paving the way for a better understanding of the mechanisms involved in these diseases. Our study underlies the power of next generation sequencing combined with linkage data in rare and genetically heterogeneous disorders.

158

Tremor-Ataxia with Central Hypomyelination (TACH), Leukodystrophy with Oligodontia (LO) and 4H syndrome (Hypomyelinating leukodystrophy with Hypodontia and Hypogonadotropic Hypogonadism) are allelic disorders caused by mutations in the same gene. *G. Bernard[1,3], E. Chouery[2], M.L. Putorti[3], M. Tétreault[3], A. Takanohashi[4], G. Carosso[4], I. Clément[3], K. Choquet[3], S. Fribourg[5], M. Teichmann[5], A. Megarbane[2], R. Schiffmann[6], A. Vanderver[4], B. Brais[3].* 1) Neurology and Neurosurgery, Montreal Children's Hospital, Montreal, Quebec, Canada; 2) Unité de génétique médicale et laboratoire associé, Université Saint-Joseph, Beirut, Lebanon; 3) CRCHUM-Hôpital Notre-Dame, Laboratoire de Neurogénétique, Université de Montréal, Quebec, Canada; 4) Children's National Medical Center, Washington DC, USA; 5) Institut Européen de Chimie et Biologie, Université de Bordeaux, Institut National de la Santé et de la Recherche Médicale, France; 6) Institute of Metabolic Disease, Baylor Research Institute, Dallas, Texas, USA.

PURPOSE: Leukodystrophies are a heterogeneous group of neurodegenerative disorders characterized by abnormal white matter on brain imaging. Three forms of leukodystrophies have overlapping clinical features: Tremor-Ataxia with Central Hypomyelination (TACH), Leukodystrophy with Oligodontia (LO) and 4H syndrome (Hypomyelination with Hypodontia and Hypogonadotropic Hypogonadism). We present the clinical features and the underlying genetic defect of these three disorders. METHODS: We have reviewed the clinical features of six cases of TACH, seven cases of 4H and eight cases of LO. By combining the two published candidate intervals for TACH and LO and the mapping of a predicted historical recombinant in a new French-Canadian TACH family, we narrowed the candidate interval to a 2.99 Mb region on chromosome 10q22.3. The coding regions, 5' and 3' UTR of the genes located in the combined interval were sequenced. RESULTS: There is some clinical and radiological overlap between these three disorders. All patients with TACH, five out of seven patients with 4H and all patients with LO were homozygote or compound heterozygote for mutations in the same gene. Despite the small number of patients, mutations seem to be different for each disorder, suggesting a genotype-phenotype correlation. CONCLUSIONS: TACH, 4H syndrome and LO share some clinical and radiological features. We identified the causative gene for these disorders, allowing physicians to provide a precise diagnosis and genetic counseling to the patients and their families. This discovery will also improve our understanding of neurodegeneration in leukodystrophies and hopefully, provide potential therapeutic targets.

159

IOCDF Genome-wide association study of obsessive-compulsive disorder. *S.E. Stewart[1,2], D. Yu[1], J. Scharf[1], E.R. Gamazon[3], P. Evans[3], J.A. Knowles[4], C. Mathews[5], P. Arnold[6], G. Hanna[7], G. Nestadt[8], M. Wagner[9], D. Denys[10], D. Cath[11], C. Lochner[12], D.L. Murphy[13], A.G. Hounie[14], M.C. Cavallini[15], H. Nicolini[16], R. Delorme[17], D. Stein[12], J. Samuels[7], E. Miguel[14], M. Pato[4], J. Fagerness[1], C. Mayerfeld[1], S. Haddad[1], B. Neale[1], S. Purcell[1], N. Cox[3], D.L. Pauls[1], International OCD Foundation Genetics Collaborative.* 1) Psychiatry, Harvard Med Sch/MGH, Boston, MA; 2) CFRI, University of British Columbia, Canada; 3) University of Chicago; 4) USC, Los Angeles, CA; 5) UCSF, San Francisco, CA; 6) University of Toronto, Canada; 7) University of Michigan Medical School; 8) Johns Hopkins University, Baltimore, MD; 9) University of Bonn, Germany; 10) University of Amsterdam, Netherlands; 11) Utrecht University and University Medical Centre, Netherlands; 12) University of Stellenbosch, Cape Town, South Africa; 13) NIMH Intramural Research Program, NIH Clinical Center, Bethesda, MD; 14) University of Sao Paulo, Brazil; 15) University of Milan Medical School, Italy; 16) Carracci Medical Group, Mexico City; 17) Robert Debre Hospital and French National Science Foundation Paris, France.

The International OCD Foundation Genetics Collaborative (IOCDF-GC) is a multi-national collaboration established to discover the genetic variation that predisposes to obsessive-compulsive disorder (OCD). A set of individuals affected with DSM-IV OCD, a subset of their parents and of individuals without the disorder, were genotyped with several different Illumina SNP microarrays (550K to 1M). After extensive data QC, a total of 1,446 cases (1,249 European ancestry (EU), 93 South African (SA) and 104 Ashkenazi Jewish (AJ)), 6,314 controls (5,791 EU, 365 AJ and 158 SA) and 400 complete trios (299 EU) remained, each with a common set of 469,586 SNPs. Separate association analyses were conducted for each of the case-control ancestry-based subsamples and for the trio samples. The three separate case-control subgroups were combined, and then further merged with the trio analyses, using a weighted-z approach and the Genome Power Calculator based non-centrality parameters. In the trios, but not the combined case-control and trio analyses, one SNP, rs6131295 (11,944,267 bp (hg18) on 20p12.1-2), exceeded the generally accepted threshold for genome-wide significance of P <5 x 10-8 with a P=3.84 x 10-8. This SNP is located ~90 kb 3' to a gene encoding the transcription factor, BTB/POZ domain containing protein 3 isoform b (BTBD3). rs6131295 is an eQTL for BTBD3, BCDIN3D, DHRS11 and ISM1. The expression of these latter two genes in the developing human brain (www.BrainSpan.org) are highly correlated with the expression of FAIM2 and ADCY8, which are the top and fourth ranked hits in the combined case-control-trio analysis of the entire sample. So, although no genome-wide significant associations were found in the entire sample, the convergence of results from both the trio and entire sample raises the possibility that our findings at BTBD3, FAIM2 and ADCY8, implicate these genes in the pathogenesis of OCD. Two regions that had previously been linked to OCD subtypes, including that on chromosome 14 for compulsive hoarding and that on 19q for childhood-onset OCD, also contain clusters of SNPs among our top hits. Moreover, a highly significant enrichment of mQTLs and frontal lobe eQTLs was observed within the SNPs that were most strongly associated with OCD, suggesting that as a set, these identified SNPs have a broad role in gene expression in OCD-related brain regions.

160

Exome sequencing of an isolated Chilean population affected by Specific Language Impairment (SLI). *R. Nudel[1], P. Villanueva[2,3,4], A. Hoischen[5], C. Gilissen[5], L. Carvajal-Carmona[1,6], M. Echeverry[6], L. Jara[2], Z. De Barbieri[4], H.M. Palomino[3], M.A. Fernández[4], H. Palomino[2], J. Veltman[5], A.P. Monaco[1], S.E. Fisher[1,7], D.F. Newbury[1].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, United Kingdom; 2) Human Genetics Program, Institute of Biomedical Sciences (ICBM), Faculty of Medicine, University of Chile, Santiago, Chile; 3) Department of Child and Dental Maxillary Orthopedics, Faculty of Dentistry, University of Chile, Santiago, Chile; 4) School of Speech and Hearing Therapy, Faculty of Medicine, University of Chile, Santiago, Chile; 5) Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences and Institute for Genetic and Metabolic Disorders, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 6) Department of Biology, University of Tolima, Tolima, Colombia; 7) Department of Language and Genetics, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Specific Language Impairment (SLI) is defined as an unexpected and considerable impairment in the acquisition and/or use of language in the absence of other diagnostic features, such as: hearing loss, learning disabilities or autism. SLI affects approximately 5% of English speaking preschool children and has been shown to be heritable. We studied an isolated Chilean founder population (from the Robinson Crusoe Island) in which the prevalence of SLI is very high: 35% of the children from the colonizing families had SLI; in contrast, the prevalence of SLI in immigrant-descendant children was 3.8%, consistent with that reported for mainland Chile (~4%). Genealogical profiling demonstrated that over 80% of affected individuals were related to a single pair of founder brothers. 77% of the current population have a colonizing surname and 14% of registered marriages are consanguineous. This population therefore provides a unique resource for identifying genetic variants which contribute to SLI susceptibility. Exome sequencing of five affected islanders revealed nine novel non-synonymous coding or splice mutations that were present in at least three individuals. The mutations were verified by Sanger sequencing in the entire island cohort (123 individuals; 41 affected; 82 unaffected). One non-synonymous coding change on chromosome 4 was significantly more frequent in affected than in unaffected individuals (genotype frequencies of 44% and 11% respectively, p=0.0002). No individuals were homozygous for this mutation. This variant was not identified in 350 European samples (223 cases, 127 controls) but was present in 8.4% of the 320 Colombian unselected controls we sequenced (we again found no individuals who were homozygous for this mutation). We suggest that this rare coding variant may contribute to the elevated frequency of SLI in this population.

161

Duplication of *GTF2I* results in separation anxiety in mice and humans. *J. Dida[1], C.B. Mervis[2], E. Lam[3], N.A. Crawford-Zelli[2], E.J. Young[3], D.R. Henderson[2], C.A. Morris[4], T. Onay[5], J. Woodruff-Borden[2], J. Yeomans[1], L.R. Osborne[2,6,7].* 1) Psychology, University of Toronto, Toronto, ON, Canada; 2) Department of Psychological and Brain Sciences, University of Louisville, Louisville, KY; 3) Institute of Medical Science, University of Toronto, Toronto, ON, Canada; 4) Department of Pediatrics, University of Nevada School of Medicine, Las Vegas, NV; 5) Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON, Canada; 6) Department of Medicine, University of Toronto, Toronto, ON, Canada; 7) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada.

A 1.5 Mb deletion of chromosome 7q11.23 causes Williams-Beuren syndrome (WBS), a neurodevelopmental disorder characterized by intellectual disability, deficits in visuospatial construction, relative strength in concrete language, lack of stranger anxiety, social disinhibition, and non-social anxiety. Duplication of 7q11.23 results in a contrasting syndrome, Dup7q11.23, associated with speech delay and/or disorder and social anxiety. *GTF2I* is one of 26 commonly deleted or duplicated genes. We generated mice with either reduced or increased *Gtf2i* genomic copy number and corresponding changes in mRNA and protein expression, to examine the phenotypic effect of *Gtf2i* dose. P7 mouse pups with 3 (n=78) or 4 (n=37) genomic copies of *Gtf2i* showed an increase in separation-induced ultrasonic vocalizations (USVs) [157 and 192 USVs respectively on average during the first 2 mins] compared to wild type littermates (n=119) [102 USVs] ($p<.005$; $p<.001$), whereas P7 pups with only a single copy of *Gtf2i* (n=23) showed reduced vocalizations [80 USVs] ($p<0.05$). This pattern suggests that *Gtf2i* has a dose-dependent effect on maternal separation anxiety in mice. To determine if a similar effect was present in humans, we measured separation anxiety in children with Dup7q11.23 (3 copies of *GTF2I*) and children with WBS (1 copy of *GTF2I*). Nineteen children [4-13 yrs] with Dup7q11.23 and 214 age-matched children with WBS were assessed using the Anxiety Disorders Interview Schedule for DSM-IV-Parent Interview (ADIS-P). In addition, parental responses for 14 children with Dup7q11.23 aged 2-5 yrs and 189 age-matched children with WBS were compared on the separation anxiety question of the Child Behavior Checklist (CBCL) for Ages 11/2 -5. 26% of children with Dup7q11.23 were diagnosed with separation anxiety disorder (ADIS-P) and 28.6% had unusual difficulty separating from their parents (CBCL) compared with only 4.2% and 1.1% of those with WBS, respectively ($p = .000$; $p = .000$). These results suggest that GTF2I plays a significant role in the contrasting separation anxiety phenotypes seen in children with Dup7q11.23 and WBS. This study links the copy number of a single gene from 7q11.23 to separation anxiety in both mice and humans and demonstrates the utility of mouse models in dissecting disorders that include multiple genes in humans. This study provides the first evidence of a single gene associated with separation anxiety and offers a molecular target for future therapies.

162

Meta-analysis of genome-wide association data of 20 000 migraine cases identifies two novel gene loci and supports several biologically relevant mechanisms. *V. Anttila[1], A. Aromaa[2], D. Boomsma[3], D. Chasman[4], L. Cherkas[5], M. Dichgans[6], C. van Duijn[7], T. Freilinger[8], M.-R. Järvelin[9], M. Kallela[10], J. Kaprio[11], C. Kubisch[12], T. Kurth[4], L. Launer[13], L. Ligthart[14], A. van den Maagdenberg[14], N. Martin[15], D. Nyholt[15], O. Raitakari[16], M. Schuerks[4], T. Spector[5], H. Stefansson[17], K. Stefansson[17], D. Strachan[18], B. de Vries[14], M. Wessman[19], B. Winswold[20], J.-A. Zwart[21], A. Palotie[1], International Migraine Genetics Meta-analysis Consortium.* 1) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 2) National Institute for Health and Welfare, Helsinki, Finland; 3) Dept. of Biological Psychology, VU University Amsterdam, Amsterdam, the Netherlands; 4) Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 5) Dept. of Twin Research & Genetic Epidemiology, King's College London, London, UK; 6) Institute for Stroke and Dementia Reseach, University of Munich, Munich, Germany; 7) Dept. of Epidemiology, Erasmus University Medical Centre, Rotterdam, the Netherlands; 8) Dept. of Neurology, Klinikum Grosshadern, Ludwig-Maximilians-Universität, Munich, Germany; 9) Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom; 10) Helsinki University Central Hospital, Helsinki, Finland; 11) Dept. of Public Health, University of Helsinki, Helsinki, Finland; 12) Institute of Human Genetics, University of Ulm, Ulm, Germany; 13) National Institute of Aging, Laboratory for Epidemiology, Demography, and Biometry, Bethesda, Maryland, USA; 14) Leiden University Medical Centre, Leiden, the Netherlands; 15) Queensland Institute for Medical Research, Brisbane, Australia; 16) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; 17) deCODE Genetics, Reykjavik, Iceland; 18) Division of Population Health Sciences and Education, St George's, University of London, London, United Kingdom; 19) Folkhälsan Research Centre, Helsinki, Finland; 20) Oslo University Hospital, Oslo, Norway; 21) Dept. of Neuroscience, Norwegian University of Science and Technology, Trondheim, Norway.

Migraine is a complex neurological disorder with chronic, episodic presentation. It affects up to 20% of the Caucasian population, women 3-4 times more often than men, totalling roughly 41 million Europeans and 27.9 million Americans. It is estimated to be among the most costly neurological disorder. Two common forms of migraine are recognized, migraine with and without aura. The heritability of migraine has been estimated as 40-60%, but the genetic determinants have only recently begun to be elucidated. In the clinic-based setting, the IHGC recently reported the first genome-wide significant association of a common SNP, rs1835740 between genes PGCP and MTDH at 8q22.1, with migraine with aura (Nat. Genet., 2010, 42:869). In the population-based setting, the Women's Genome Health Study recently reported the first genome-wide significant associations of common SNPs with migraine irrespective of aura status, with replication from contributors to the IHGC and implicating the genes PRDM16 (1p36.32, rs2651899), TRPM8 (2q37.1, rs10166942), and LRP1 (12q13.3, rs11172113) (Nat. Genet., 2011, in press). Here, we present the results of a large scale meta-analysis of genome-wide association data for migraine, including a total of 20,000 cases compared with 90,000 controls from more than twenty different clinic- and population-based study cohorts from three continents, collectively called the International Migraine Genetics Meta-analysis Consortium. We also report secondary association analyses for migraine with and without aura, as well as some pathway and interaction analyses for the identified genes and loci. On a preliminary basis, we identify two new associations at 5p15 and 6p24 in addition to the previously reported findings at 1p36 and 2q37. The genes implicated at the new loci offer interesting insights into the genetics and pathophysiology of migraine, reinforcing the relevance of suspected pathways, for example relating to glutamate homeostasis. The study design further allows us to compare population- and clinic-based ascertainment, and thereby address the issue of genetic heterogeneity of a complex neurological trait. Finally, we evaluate the benefits of including information on a number of co-morbid disorders and anthropomorphic features available in large population based cohorts for understanding common genetic associations with migraine.

163

Functional polymorphisms in the TERT promoter are associated with risk of epithelial ovarian cancer. *G. Chenevix-Trench[1], J. Beesley[1], H. Pickett[2], S. Johnatty[1], X. Chen[1], D. Rider[3], M. Stutz[2], D. Lambrechts[4], J. Chang-Claude[5], T. Dork[6], M.T. Goodman[7], B. Kiemney[8], E. Bandera[9], L.S. Cook[10], N. Le[11], I. Campbell[12], S. Gayther[13], S. Ramus[13], S. Macgregor[1], E. Goode[3], R. Reddel[2] on behalf of the Ovarian Cancer Association Consortium (OCAC).* 1) Queensland Inst Medical Res, Brisbane, Australia; 2) Children's Medical Research Institute, Sydney, Australia; 3) Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, MN, USA; 4) Vesalius Research Center, VIB and K.U. Leuven, Belgium; 5) Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany; 6) Clinics of Obstetrics and Gynaecology, Hannover Medical School, Hannover, Germany; 7) Cancer Research Center, University of Hawaii, Honolulu, HI, USA; 8) Nijmegen Centre for Evidence Based Practice, NL-6500 HB Nijmegen The Netherlands; 9) The Cancer Institute of New Jersey, Robert Wood Johnson Medical School, New Brunswick, NJ 08903, USA and School of Public Health, University of Medicine and Dentistry of New Jersey, Piscataway, NJ, USA; 10) Division of Epidemiology and Biostatistics, University of New Mexico, Albuquerque, New Mexico and Alberta Health Services-Cancer Care, Calgary, AB, Canada; 11) Dept of Statistics, BC Cancer Agency, Vancouver, BC V5Z 1L3, Canada; 12) Centre for Cancer Genomics and Predictive Medicine, Peter MacCallum Cancer Centre, Melbourne, Australia and Department of Pathology, University of Melbourne, Parkville, Victoria, Australia; 13) Dept of Preventive Medicine, University of Southern California, Los Angeles, CA, USA.

Most cancers depend on expression of telomerase, whose catalytic subunit is encoded by the TERT gene, for their continued proliferation. Genome wide association studies (GWAS) have found that the TERT-CLPTM1L locus is associated with susceptibility to several cancers. Computing a gene-based test (VEGAS) of association at TERT and CLPTM1L yielded evidence in dbGAP data for association with risk of lung, prostate and pancreatic cancer, but not breast cancer; combining all in a cross-cancer meta-analysis, revealed a genome-wide significant gene-based p=4.1e-7 for CLPTM1L (p= 0.008 after correction for 19,000 genes). Using a candidate gene approach, we previously reported an association between an intronic SNP in TERT, rs7726159, and epithelial ovarian cancer (EOC) risk in 3,059 serous invasive cases and 8,905 controls from 16 Ovarian Cancer Association Consortium (OCAC) case-control studies (adjusted ORper-allele 0.88 (0.81-0.96), p= 0.003 with the common allele associated with decreased risk). To further clarify this association we employed a fine-mapping strategy in eight OCAC studies, selecting tagging SNPs from ±250 kb across the TERT-CLPTM1L locus from the 1000 Genomes Project (12/2009) plus four additional SNPs implicated by our previous study of EOC or by cancer GWAS. We analysed 28 SNPs in 2,401 invasive EOC cases, including 871 serous cases, and 3,455 controls, all of Caucasian ancestry. We used single marker and step-wise logistic regression models, adjusted for study and age at a threshold of $p \leq 0.05$ for addition (forward stepwise) or removal (backward stepwise) of SNPs. A SNP in the TERT promoter, rs2736109, showed the strongest association with serous EOC (adj. ORper-allele 0.85 (0.76-0.95), p = 0.004) with the minor A allele associated with decreased risk.To examine the potential consequences of rs2736109 and another promoter SNP, rs2736108, we generated luciferase reporter constructs comprising 3.7 kb of the TERT promoter with various combinations of alleles and transfected them into breast and ovarian cell lines. We found that luciferase expression was diminished by the presence of both the minor alleles (A-A) at rs2736108 and rs2736109, but not when either allele is present alone. Our analysis of 345 Australian controls suggests that the A-A haplotype at rs2736108 and rs2736109 occurs with a frequency of 32%;, suggesting that this relatively common promoter haplotype may lower the risk of serous EOC through decreasing TERT expression.

164

Massively Parallel Sequencing Identifies Inherited Mutations in 12 genes in Women with Ovarian, Peritoneal, and Fallopian Tube Carcinomas. *T. Walsh[1], S. Casadei[1], M. Lee[1], C. Pennil[2], A. Nord[1], A. Thornton[1], W. Roeb[1], M-C. King[1], E. Swisher[1,2].* 1) Medical Genetics, University of Washington, Seattle, WA; 2) Gynecologic Oncology, University of Washington, Seattle, WA.

Ovarian carcinoma is the most deadly of gynecological malignancies; the majority of women are diagnosed with advanced stage disease when the chance of cure is small. Inherited mutations in *BRCA1* and *BRCA2* create a high lifetime risk of ovarian carcinoma of between 20 and 50%. *BRCA1* and *BRCA2* mutations are thought to account for the majority of inherited ovarian carcinomas, with estimates of 13% to 15% of ovarian carcinoma patients in North America carrying germline mutations in one of these two genes. However, the contribution of other genes has not been evaluated and clinical genetic testing is currently done gene-by-gene, with each test costing thousands of dollars. Using targeted capture and massively parallel genomic sequencing, we screened for all classes of germline mutations in 21 tumor suppressor genes. Subjects were women with primary ovarian, peritoneal, or fallopian tube carcinoma, consecutively enrolled at diagnosis and not selected for family history or age at onset. Of 360 subjects, 18% carried germline loss-of-function mutations in *BRCA1* or *BRCA2* and 6% carried germline loss-of-function mutations in one or more of *BARD1*, *BRIP1*, *CHEK2*, *MRE11A*, *MSH6*, *NBN*, *PALB2*, *RAD50*, *RAD51C*, and *TP53*. Six of these genes were not previously implicated in inherited ovarian carcinoma. Inherited mutations in *BRCA1* and *BRCA2* were more associated with serous and undifferentiated carcinomas; inherited mutations in other genes appeared in all histological types. Primary carcinomas were generally characterized by genomic loss of normal alleles of the mutant genes. Of women with inherited mutations, >30% had no family history of breast or ovarian carcinoma and >35% were age 60 years or older at diagnosis. Our results indicate that more than one in five ovarian carcinomas are associated with germline mutations in tumor suppressor genes, and that these mutations are distributed in a larger number of genes than previously appreciated. Comprehensive genetic testing for inherited ovarian carcinoma is warranted for all women with ovarian carcinoma regardless of age or family history. In addition, as the cost of multigene testing continues to fall, it may become cost effective to offer population testing for cancer susceptibility genes in order to allow targeted prevention.

165

Ovarian cancer susceptibility loci and risk of ovarian cancer in BRCA1 and BRCA2 carriers. *S.J. Ramus[1], A.C Antoniou[2], K. Kuchenbaecker[2], P. Soucy[3], L. McGuffog[2], S. Healey[4], O.M. Sinilnikova[5,6], P. Radice[7], D.E. Goldgar[8], S. Peock[2], R.K. Schmutzler[9,10], D. Stoppa-Lyonnet[11,12,13], M.A. Rookus[14], A. Jakubowska[15], kConFab. Investigators[16], J. Simard[17], D.F. Easton[2], F. Couch[18], G. Chenevix-Trench[4] on behalf of the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA).* 1) Dept of Preventive Medicine, University of Southern California, Los Angeles, CA.,USA; 2) Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK,; 3) Cancer Genomics Laboratory, Centre Hospitalier Universitaire de Quebec and Laval University, 2705 Laurier Boulevard, T3-57, Quebec City, QC, Canada; 4) Queensland Institute of Medical Research, Brisbane, Australia; 5) Centre Hospitalier Universitaire de Lyon/Centre Leon Berard, Unite Mixte de Genetique Constitutionnelle des Cancers Frequents, Lyon, France; 6) Equipe labellisee LIGUE 2008, UMR5201 CNRS, Centre Leon Berard, Universite de Lyon, Lyon, France; 7) Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Preventive and Predicted Medicine Fondazione IRCCS Istituto Nazionale Tumori (INT), Milan, Italy; 8) Department of Dermatology, University of Utah School of Medicine, Salt Lake City, USA; 9) Centre of Familial Breast and Ovarian Cancer, Department of Gynaecology and Obstetrics University hospital of Cologne, Cologne, Germany; 10) Centre for Integrated Oncology (CIO), University hospital of Cologne, Cologne, Germany; 11) Service de Genetique Oncologique, University Paris Descartes, Paris, France; 12) Unite INSERM U830, Institut Curie, Paris, France; 13) Faculty of Medecine, University Paris Descartes, Paris, France; 14) Department of Epidemiology, Netherlands Cancer Institute, Amsterdam; 15) International Hereditary Cancer Centre, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland; 16) The Peter MacCallum Cancer Centre, Melbourne, VIC, Australia; 17) Canada Research Chair in Oncogenetics, Cancer Genomics Laboratory, Centre Hospitalier Universitaire de Quebec and Laval University, 2705 Laurier Boulevard, T3-57, Quebec City, QC, Canada; 18) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA.

Germline mutations in BRCA1 and BRCA2 are associated with increased risks of breast and ovarian cancer. A genome wide association study (GWAS) recently identified six alleles associated with risk of ovarian cancer for women in the general population. One of these ovarian cancer risk loci, at 9p22.2, has already been shown to modify the risk for ovarian cancer in BRCA1/2 carriers. We have now used the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA) to evaluate additional loci as potential modifiers of ovarian cancer risk for BRCA1 and BRCA2 mutation carriers. We genotyped four SNPs - rs10088218 (at 8q24), rs2665390 (at 3q25), rs717852 (at 2q31) and rs9303542 (at 17q21) - in 12,599 BRCA1 and 7,132 BRCA2 carriers, which included up to 2,678 ovarian cancer cases. Associations with ovarian and breast cancer risk were assessed using a retrospective likelihood approach. The minor allele of rs2665390 (3q25) was associated with an increased risk of ovarian cancer for both BRCA1 carriers (per-allele hazard ratio (HR) =1.25 [95%CI: 1.10-1.42] P-trend =6.1×10-4), and BRCA2 carriers (per allele HR=1.48 [95%CI: 1.21-1.83] P-trend = 1.8×10-4). The minor allele of rs10088218 (8q24) was associated with a decreased risk of ovarian cancer for both BRCA1 carriers (per-allele HR= 0.89 [95%CI: 0.81-0.99] P-trend =0.029), and BRCA2 carriers (per allele HR=0.81 [95%CI: 0.67 - 0.98] P-trend =0.033). The two remaining loci, rs717852 (2q31) and rs9303542 (17q21), were associated with an increased ovarian cancer risk in BRCA2 carriers only (rs717852 per allele HR=1.25 [95%CI: 1.10-1.42], P-trend =6.6×10-4; rs9303542 per allele HR=1.16 [95%CI: 1.02-1.33] P-trend =0.026). There was no evidence that any of these four SNPs modified breast cancer risk for mutation carriers. The identification of multiple loci modifying ovarian cancer risk may be potentially useful for counselling women with BRCA1 and BRCA2 mutations about their risk of ovarian cancer.

166

Evaluation of transcripts generated by germline mutations identified by massively parallel genomic sequencing in ovarian cancer patients. *S. Casadei[1], T. Walsh[1], M. Lee[1], A. Nord[1], S. Stray[1], A. Thornton[1], C. Pennil[2], A. Wickramanayake[2], B. Norquist[2], M.C. King[1], E. Swisher[2].* 1) Division of Medical Genetics, Department of Medicine, University of Washington School of Medicine, Seattle, WA; 2) Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, University of Washington School of Medicine, Seattle, WA.

Inherited loss-of-function mutations in BRCA1, BRCA2 and other known tumor suppressor genes lead to very high risks of breast and/or ovarian carcinomas. Recent approaches to mutation detection based on next-generation sequencing allow accurate and efficient identification of almost all classes of mutations in these genes. The consequences of mutations potentially altering splice sites (SS) or exonic splice enhancers (ESE), as well as of genomic rearrangements involving coding sequences, still require evaluation. Targeted genomic capture and massively parallel genomic sequencing were used to screen for all germline mutations in a panel of 21 tumor suppressor genes in genomic DNA from 360 women with primary ovarian, peritoneal, or fallopian tube carcinoma. All rare variants within 12 nucleotides of the consensus exon-intron boundaries were flagged for transcript analysis. Both predictive bioinformatics models for splice site and splice enhancer mutations and RT - PCR experiments were carried out to test for aberrant transcription, using RNA isolated from patients' lymphoblasts. In addition, genomic rearrangements causing exonic deletions and duplications were evaluated for aberrant transcription in lymphoblast RNA. The 360 subjects harbored 85 loss-of-function germline mutations in 12 genes. Of these mutations, 6 were splice mutations and 6 were large genomic rearrangements, totaling 14% of all mutations. Three of six splice mutations were in BRCA2 and one each were in BRCA1, MRE11A and RAD51C. The splice mutations in BRCA2, BRCA1, and MRE11A led to frameshifts; the splice mutation in RAD51C led to exon skipping and deletion of 44 amino acids in a critical protein domain. Of the six genomic rearrangements, four deletions and one duplication were detected in BRCA1 and one deletion was detected in BRIP1. Four deletions and the duplication led to frameshifts; one deletion led to an unstable message. Rapid advances in throughput of next-generation DNA sequencing methods leads to identification of more disease-related genes and of a much wider spectrum of loss-of-function mutations than previously encountered. Bioinformatics tools do not adequately predict the consequences of mutations potentially altering splicing and transcript stability. In our view, experimental evaluation of the significance for transcription of splice mutations and structural genomic rearrangements is essential for the proper assessment of clinically relevant mutations.

167

Discovery of new genes for inherited breast cancer by exome sequencing of unresolved high-risk families. *C. Spurrell[1], M. Lee[2], A. Nord[1], S. Casadei[2], A. Thornton[2], J. Mandell[2], T. Walsh[2], M-C. King[1,2].* 1) Genome Sciences, University of Washington, Seattle, WA; 2) Medical Genetics, University of Washington, Seattle, WA.

Breast cancer is the most common cancer among American women and family history is a significant risk factor for its occurrence. Studies in high-risk families have uncovered multiple genes in which inherited mutations lead to significantly increased risk of breast cancer. There is genetic heterogeneity of both loci and alleles, and mutations are usually private to each family. Our lab has developed a method for using next-generation sequencing to screen all known breast cancer genes, including *BRCA1* and *BRCA2*, moderate risk genes and genes implicated in cancer syndromes. We screened 192 women who were diagnosed with breast cancer before the age of 60 and were from high-risk families. Because individual cases within high-risk families may be sporadic, we sequenced multiple individuals in each family. From this set we chose 27 families that did not harbor any deleterious mutations in known genes. We hypothesize that the remaining familial risk of breast cancer is due to individually rare alleles of moderate to high penetrance in as-yet-undiscovered genes. In order to discover new breast cancer susceptibility genes we have performed exome sequencing on these 27 families, selecting the two most distantly related affected individuals to sequence. The exome libraries were sequenced to 125x median coverage with 95% of the exome covered by >10 reads. All variants were filtered for rare events that were co-inherited by both affected individuals. We verified these events through standard sequencing and characterized each candidate variant in all family members to look for co-segregation with breast cancer. This resulted in a list of candidate genes for each family. When possible we tested for loss of heterozygosity in the tumor DNA. We will make a custom capture pool with these candidate genes and screen >1000 high-risk (unsolved) breast cancer families for additional deleterious mutations. The power of this approach is that we can combine co-segregation, LOH, and presence of multiple mutations in the same candidate gene to find new genes that explain the inherited predisposition to cancer in a proportion of high-risk families.

168

Common breast cancer susceptibility alleles are associated with tumor subtypes in *BRCA1* and *BRCA2* mutation carriers: results from the Consortium of Investigators of Modifiers of *BRCA1/2*. *A.C. Antoniou[1], A.M. Mulligan[2], F.J. Couch[3], D. Barrowdale[1], S.M. Domchek[4], D. Eccles[5], H. Nevanlinna[6], S.J. Ramus[7], M. Robson[8], M. Sherman[9], A.B. Spurdle[10], B. Wappenschmidt[11], L. McGuffog[1], J. Simard[12], G. Chenevix-Trench[10], D.F. Easton[1], I.L. Andrulis[13] on behalf of the Consortium of Investigators of Modifiers of BRCA1/2.* 1) Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom; 2) Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada; Department of Laboratory Medicine, and the Keenan Research Centre of the Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, ON, Canada; 3) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA; 4) Abramson Cancer Center, University of Pennsylvania; 5) Faculty of Medicine, University of Southampton, Southampton University Hospitals NHS Trust, Southampton, UK; 6) Department of Obstetrics and Gynecology, Helsinki University Central Hospital, Helsinki, Finland; 7) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, California, USA; 8) Departments of Medicine and Surgery, Memorial Sloan-Kettering, Cancer Center, New York, NY, USA; 9) National Cancer Institute, Division of Cancer Epidemiology and Genetics, Hormonal and Reproductive Epidemiology Branch, Rockville, USA; 10) Queensland Institute of Medical Research, Brisbane, QLD 4006, Australia; 11) Centre of Familial Breast and Ovarian Cancer, Department of Gynaecology and Obstetrics and Centre for Integrated Oncology, University hospital of Cologne, Germany; 12) Canada Research Chair in Oncogenetics, Cancer Genomics Laboratory, Centre Hospitalier Universitaire de Québec and Laval University; 13) Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto; Cancer Care Ontario, Departments of Molecular Genetics and Laboratory Medicine and Pathobiology, University of Toronto, Ontario, Canada.

Previous studies have demonstrated that common breast cancer susceptibility alleles are differentially associated with breast cancer risk for *BRCA1* and/or *BRCA2* mutation carriers. We hypothesized that differences in tumor characteristics between *BRCA1* and *BRCA2* carriers account for these differential allelic associations. We used genotype data on up to 11,421 *BRCA1* and 7,080 *BRCA2* carriers, of whom 4,310 had been affected with breast cancer and had information on either Estrogen (ER) or Progesterone Receptor (PR) status of the tumor, to assess the associations of 12 known breast cancer susceptibility loci with tumor characteristics. Associations were evaluated using a retrospective cohort approach which models the simultaneous effect of each SNP on more than one tumor subtype. We found stronger associations with ER-positive breast cancer than ER-negative for 11 loci in both *BRCA1* and *BRCA2* carriers. Among *BRCA1* carriers, SNP rs2981582 (*FGFR2*) exhibited the biggest difference based on ER status (per-allele HR for ER-positive=1.35, 95%CI:1.17-1.56 vs HR=0.91, 95%CI:0.85-0.98 for ER-negative, P-heterogeneity=6.5×10^{-6}). In contrast, SNP rs2046210 at 6q25.1 near *ESR1* was primarily associated with ER-negative breast cancer risk for both *BRCA1* and *BRCA2* carriers. In BRCA2 carriers, SNPs in *FGFR2*, *TOX3*, *LSP1*, *SLC4A7/NEK10*, 5p12, 2q35, and 1p11.2 were significantly associated with ER-positive but not ER-negative disease. Similar results were observed when differentiating breast cancer cases by PR status. The apparent differences in SNP associations between *BRCA1* and *BRCA2* carriers, and non-carriers, may be explicable by differences in the prevalence of tumor subtypes. Based on these results, we estimate that a *BRCA1* mutation carrier at the 95th percentile of the combined 12 SNP profile distribution would be at 60% risk of developing ER-negative breast cancer by age 80 compared with 43% for a *BRCA1* mutation carrier at the 5th percentile. The risks of developing ER-positive breast cancer would be 18% and 46% by age 80 at the 5th and 95th percentiles of the ER-positive breast cancer risk distribution. Incorporating these associations into breast cancer subtype-specific risk models should improve clinical management for mutation carriers.

169

Common variation at the *C19orf62* and *ZNF365* loci is associated with breast and ovarian cancer risk in *BRCA1* and *BRCA2* mutation carriers. *F.J. Couch[1], M.M. Gaudet[2], A.C. Antoniou[3], S.J. Ramus[4], K. Kuchenbaecker[5], P. Soucy[6], J. Beesley[7], X. Wang[1], T. Kirchhoff[8], L. McGuffog[3], D. Barrowdale[3], O.M. Sinilnikova[9], D. Goldgar[10], S.L. Neuhausen[11], A. Borg[12], A.M. Gerdes[13], T.V.O. Hansen[14], A. Osorio[15], I.L. Andrulis[16], M. Greene[17], S.M. Domchek[18], P. Radice[19], D.F. Easton[3], G. Chenevix-Trench[7], K. Offit[8], J. Simard[6], CIMBA.* 1) Mayo Clinic, Rochester, MN; 2) American Cancer Society, Atlanta, GA; 3) University of Cambridge, Cambridge, UK; 4) University of Southern California, Los Angeles, CA; 5) University College London, UK; 6) Laval University, Quebec, Canada; 7) Queensland Institute of Medical Research, Brisbane, Australia; 8) Memorial Sloan-Kettering Cancer Center, New York, N.Y; 9) Centre Hospitalier Universitaire de Lyon/Centre Léon Bérard, Lyon, France; 10) University of Utah, Salt Lake City, UT; 11) City of Hope, Duarte, CA; 12) Lund University, Lund, Sweden; 13) Rigshospital and Odense University Hospital, Denmark; 14) Copenhagen University Hospitalet, Copenhagen, Denmark; 15) National Cancer Research Centre, Madrid, Spain; 16) Mount Sinai Hospital, Toronto, Ontario, Canada; 17) National Cancer Institute, Bethesda, MD; 18) University of Pennsylvania School of Medicine, Philadelphia, PA; 19) Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy.

Genome wide association studies (GWAS) of *BRCA1* and *BRCA2* mutation carriers have identified common variants at 19p13.1 (MERIT40 locus) and 10q21.2 (ZNF365 locus) associated with risk of breast cancer for *BRCA1* and *BRCA2* mutation carriers, respectively. Variants in the 19p13.1 and 10q21.2 loci have also been associated with ovarian cancer and breast cancer, respectively, in the general population. Here we report on further analysis of these loci using 12,599 *BRCA1* and 7,132 *BRCA2* mutation carriers from the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA). Associations with breast and ovarian cancer risk were assessed using a retrospective likelihood approach. Consistent with the previous GWAS findings, in a combined analysis using all genotype data, the minor alleles of SNPs at 19p13.1 were associated with an increased risk of breast cancer in *BRCA1* but not in *BRCA2* mutation carriers, (rs8170 per allele Hazard ratio (HR)=1.20 [95%Confidence Interval (CI) 1.13-1.28] ptrend=$8.7x10^{-9}$) (rs67397200 per allele HR=1.17 [95%CI 1.11-1.23] ptrend=$2.4x10^{-8}$). Similarly, the minor allele of rs16917302 at 10q21.2 was associated with a decreased breast cancer risk in *BRCA2* mutation carriers only (per allele HR=0.83 [95%CI 0.75-0.93] ptrend=7.0×10^{-4}). Furthermore, associations between rs67397200 (19p13.1) and ER-negative disease as well as between rs16917302 (10q21.2) and ER-positive disease, were observed for both *BRCA1* and *BRCA2* mutation carriers. Most importantly, for the first time the minor alleles of variants from the 19p13.1 locus (rs67397200) were associated with an increased risk of ovarian cancer in both B*RCA1* mutation carriers (per allele HR=1.16 [95% CI 1.05-1.29] ptrend=3.8×10^{-4}) and *BRCA2* mutation carriers (per allele HR=1.30 [95%CI 1.10-1.52] ptrend=1.8×10^{-3}). This is the first variant found to influence both breast and ovarian cancer in either *BRCA1* or *BRCA2* mutation carriers. New understanding of the contribution of these loci to cancer may lead to improved understanding of the biology of tumor development in *BRCA1* and *BRCA2* mutation carriers.

170

Admixture mapping identifies regions on 6q25 and 11p15 associated with breast cancer in U.S. Latinas. *L. Fejerman[1], G. Chen[2], S. Huntsman[1], D. Hu[1], A. Williams[3], B. Pasaniuc[4], E.M. John[5], M. Via[1], C. Gignoux[1], S. Ingles[2], B.E. Henderson[2], L. Le Merchand[6], L.N. Kolonel[6], G. Torres-Mejia[7], E. Perez-Stable[1], E. Gonzalez-Burchard[1], C. Eng[1], C. Haiman[2], E. Ziv[1].* 1) Dept Med, Univ California, San Francisco, San Francisco, CA; 2) University of Southern California, Keck School of Medicine, Los Angeles, CA; 3) Harvard Medical School, Dept of Genetics, Cambridge, MA; 4) Harvard School of Public Health, Dept of Epidemiology, Cambridge, MA; 5) Cancer Prevention Institute of California, Fremont, CA; 6) University of Hawaii Cancer Center, Hawaii; 7) Instituto Nacional de Salud Publica, Cuernavaca, Morelos, Mexico.

Breast cancer incidence varies across different racial/ethnic populations in the U.S. We have previously demonstrated that higher European genetic ancestry among U.S. Latinas and Mexicans is associated with increased risk of breast cancer. We used an admixture mapping approach to search for loci associated with ancestry differences between cases and controls. We used genome wide association data from 1,461 U.S. Latinas with breast cancer and 1,280 U.S. Latina controls and selected a subset of the genotyped SNPs to estimate locus specific genetic ancestry. We applied logistic regression models to test the association between locus specific ancestry and breast cancer risk. We identified two loci that were significantly associated with breast cancer risk after adjusting for multiple hypothesis testing: 6q25 (p=8.2*10-7) and 11p15 (p=2.9*10-6). In both regions the risk of breast cancer increases with higher European ancestry and a protective effect is observed with higher Native American ancestry. The peak of the 6q25 signal includes the estrogen receptor 1 gene (ESR1) and the 5' region of the ESR1 gene, which has been associated with breast cancer in previous genome wide association studies (GWAS) in Asians and Europeans. None of the previously reported variants within this region accounts for the admixture mapping signal. The 11p15 signal is within a region not previously implicated in breast cancer. Combined analysis of the SNPs genotyped and imputed in these regions found no signal variant that explained the admixture signal at either 6q25 or 11p15. However, we were able to explain the locus specific ancestry signals at both 6q25 and 11p15 by means of a multi-SNP model, suggesting that multiple variants may explain the risk affect. These data strongly suggest that the differences in breast cancer incidence between U.S. Latinas and non-Hispanic Whites are due, at least in part, to differences in genetic predisposition.

171

Reference Transcriptomes for 14 Non-Human Primates reveals cross-species and species-specific isoforms and novel, active RNAs. *C.E. Mason[1], A. Akalin[1], G. Schroth[3], M. Katze[2], Non-Human Primate Reference Transcriptome Consortium.* 1) Physiology and Biophysics, Weill Cornell Medical College, New York, NY; 2) University of Washington's Washington National Primate Research Center, Seattle, WA, USA; 3) Illumina, Inc., Hayward, CA, USA.

RNA-Sequencing (RNA-Seq) allows for fast, single-base resolution scans of entire transcriptomes for many samples, and this has fostered the development of pioneering discoveries in human genetics and medicine, such as a demonstration of ~94% of genes showing alternative splicing (Wang et al, 2008) or to reveal low abundance transcripts. Indeed, a key advantage over microarrays is the greater dynamic range that can be achieved with RNA-seq, thereby allowing the detection and accurate quantitation of low abundance transcripts. Also, unlike array technology data, RNA-seq data collection has no dependency on prior knowledge of the sequence or organization of the transcriptional units resident in the host genome. This allows transcriptional profiling to be performed on nonhuman primate (NHP) species that lack genomic sequence data and/or where such data is not fully annotated. Thus, we chose to perform RNA-Seq on 20 tissues from 14 non-human primates as part of the non-human primate reference transcriptome project (NHPRTP), to better understand the evolution of the transcripts in primates and to improve the annotation of existing or pending genomes. This model has been successfully used to improve human gene annotation in the BodyMap project, and we employed similar methods for transcript characterization and mapping. We used three methods of library construction: a duplex-specific nuclease (DSN) method that reveals low-abundance transcripts, standard mRNA-Seq, and directional mRNA-Seq using Uracil-DNA glycosolace (UDG). We sequenced each species to at least 400 million reads. We then used Trans-Abyss to re-construct ab intiio isoforms from the RNA-Seq data. As a test, we first examined the Cynologous macaque isoforms vis-a-vis the human genome to observe phylogenetic divergence. We observed 2,867 (13%) Refseq genes with 100% re-construction of the gene model, but importantly, we were also able to perform partial reconstruction (where we cover 100% of at least one exon and one intron boundary) for 16,746 (75%) of Refseq genes. Also, we observe an enrichment of conserved genes in neurogenesis and signal transduction (p=0.004, Bonferroni). We have also discovered thousands of species-specific and cross-species transcriptionally active regions (TARs), which highlight the utility of using un-biased methods for transcriptome characterization.

172

Massively parallel saturation mutagenesis and functional analysis of mammalian enhancers. *R. Patwardhan[1], J.B. Hiatt[1], M.J. Kim[3,4], D. May[2], R.P. Smith[3,4], J. Wagner[1], N. Ahituv[3,4], L. Pennacchio[2], J. Shendure[1].* 1) University of Washington, Genome Sciences, Seattle, WA; 2) Lawrence Berkeley National Laboratory, Berkeley, CA; 3) Department of Bioengineering and Therapeutic Sciences, UCSF, San Francisco, CA; 4) Institute for Human Genetics, UCSF, San Francisco, CA.

We present a high-throughput method to dissect the architecture of mammalian enhancers at single base resolution through generation and in vivo functional interrogation of tens of thousands of variants of a given enhancer in a single experiment.

While high-throughput genome-wide assays such as ChIP-Seq have proven effective at locating tissue or cell-type specific enhancers, very few of these elements have been studied in detail at base-pair resolution. Classically, methods such as saturation mutagenesis were used to perform such studies, but these methods are inherently low throughput, both in terms of generating mutant constructs as well as in evaluating their relative activities. Here we present a method to perform these experiments in a massively parallel way, and apply it to dissect a 302bp enhancer with strong activity in mouse livers.

A highly complex library of variants of this enhancer was constructed synthetically by stitching 90bp oligos synthesized with a 3% per-base error rate. These were cloned into a plasmid bearing a minimal promoter. Each enhancer variant was tagged by a unique 20bp barcode cloned downstream of the transcription start site. The barcoded library representing more than 100K unique enhancer variants was introduced into the mouse liver via tail-vein assay. Relative impact of each mutation was calculated based on the abundance of corresponding barcoded transcripts in RNA extracted from the mouse livers.

We successfully obtained a comprehensive profile of impact of all single base mutations and tens of thousands of combinations of mutations within this enhancer in a single experiment. In general, the effect sizes of individual mutations were significant, yet surprisingly modest, providing insight into the robustness of enhancers to single nucleotide variants. Positions with relatively high impact were often found to be in clusters, suggesting the likely locations of TF binding sites. We also observed many examples of synergistic and compensatory effects between both distally and proximally located mutations.

Ongoing work is aimed at further modeling of this enhancer, as well as interrogating additional liver-active enhancers. We anticipate that massively parallel saturation mutagenesis will enable the dissection of the fine-scale architecture of enhancers, and more generally allow us to learn the distribution of effect sizes of mutations in regulatory regions.

173

The role of genetic variation underlying transcription factor mRNA and protein levels. *R.J. Hause[1,2,3], A.L. Stark[4], N.N. Antao[5], L. Gorsic[5], D.F. Gill[3], S.H. Chung[3], C.D. Brown[1,3,4], K.P. White[1,3,4], M.E. Dolan[1,3,5], R.B. Jones[1,2,3].* 1) Committee on Genetics, Genomics, and Systems Biology; 2) Ben May Department for Cancer Research; 3) Institute for Genomics and Systems Biology; 4) Department of Human Genetics; 5) Section of Hematology/Oncology, Department of Medicine, University of Chicago, Chicago, IL.

Genome-wide association analyses have discovered many DNA variants that influence complex phenotypes. Because gene expression is a molecular phenotype that acts as an "intermediary" between genetic and physiological variation, there has been enthusiasm about using expression quantitative trait loci (eQTL) mapping to try to better understand complex disease by identifying regulatory variation that affects mRNA expression. An implicit assumption made in these analyses is that eQTLs explain subsequent differences in protein levels, even though experiments have shown imperfect correlations between mRNA and protein expression. To improve our understanding of the genetic basis of protein expression variation and how it affects cellular phenotypes, we have (1) extended micro-western and reverse phase lysate array technologies to quantify protein levels and examine the correlation between mRNA levels, miRNA levels, methylation levels, protein levels, and physiological confounders; (2) compared the genetic determinants of variation in transcriptome and proteome abundances; and (3) identified genetic variants that are associated with cellular phenotypes through their effect on mRNA and/or protein expression. We have quantified over 400 transcription factor and core signaling proteins across three biological replicates for each of 68 unrelated HapMap YRI LCLs. We performed association analyses on these protein levels and compared significant loci to those associated with mRNA transcript levels measured by expression arrays, exon arrays, and RNA sequencing. While many eQTLs replicated at the protein level, we also identified loci that uniquely associated with protein levels, including some whose mRNA transcript levels escaped detection, such as a genome-wide significant protein QTL (pQTL) for EBF2 that is located within one of its introns. We also identified a number of putative "master regulator" SNPs associated with multiple protein expression traits. Finally, to identify potentially functional polymorphisms, we examined the overlap of pQTLs with chemotherapeutic drug susceptibility and disease-associated loci. To our knowledge, this study represents the most comprehensive analysis of transcription factor protein levels across a population to date. Our results offer a first look at the genetic circuitry that underlies both proteome and transcriptome diversity in cells derived from a human population and how they together contribute to phenotypic diversity.

174

RARE AND COMMON TISSUE-SPECIFIC GENETIC EFFECTS ON TRANSCRIPTOME ABUNDANCE AND STRUCTURE USING NEXT GENERATION SEQUENCING. *S.B. Montgomery[1,2], M. Gutierrez-Arcelus[1], T. Lappalainen[1], H. Ongen[1], A. Buil[1], A. Nica[1], I. Padioleau[1], M. Guipponi[1], P. Makrythanasis[1], A. Yurosvsky[1], T. Giger[1], L. Romano[1], C. Gehrig[1], E. Falconnet[1], S.E. Antonarakis[1], E.T. Dermitzakis[1].* 1) Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 2) Present Address: Department of Pathology and Genetics, Stanford University, Stanford, CA.

With the availability of next generation RNA-sequencing we have been increasingly able to understand the diversity of modes through which genetic variation is influencing the transcriptome (Montgomery et al., Nature, 2010, Pickrell et al., Nature, 2010). However, the degree of tissue specificity of these effects is poorly understood. To investigate this, we have used the Gencord collection of umbilical cord samples from newborns of Western European descent and isolated B-cells (transformed to LCLs), CD4 T-cells and Fibroblasts (Dimas et al., Science, 2009). For 55 individuals, we have now collected RNA-Sequencing data for each of the 3 cell types (15-20 million reads/sample) and methylation data for Fibroblasts and LCLs using the Infinium 450K HD Methylation array. We have also genotyped each individual using the Omni 2.5 array and imputed 1000 genomes phase 1 variants. Using this data, we have explored novel gene structure and dissected genetic differences in transcript abundance and UTR termination. As previously, we have discovered increased numbers of eQTL with the use of RNAseq data and high degree of tissue specificity. We have also detected abundant signatures of splicing QTLs and exon length QTLs (~80000 per tissue; FDR 0.1) revealing multiple mechanisms of genetic control of transcript structure. We have observed and related heritable differences in methylation status to changes in gene expression. We have been able to dissect the impact and tissue-specificity of variants affecting all aspects of transcriptome characteristics. RNA-Seq based allele specific expression analyses allowed us to: (i) estimate that 25% of allelic effects are shared between any two tissues; (ii) identify rare regulatory haplotypes and to identify the degree to which rare effects are relevant across cell types. We have used all the above information to explore tissue-sharing of eQTLs which also are correlated to GWA signals and have such interrogated specific transcripts as candidate components underlying specific disease etiologies.

175

The genetic basis of inter-individual variation in mRNA decay rates. *A.A. Pai[1], J.F. Degner[1], S. De Leon[1,2], N. Lewellen[1,2], J.K. Pickrell[1], J.K. Pritchard[1,2], Y. Gilad[1].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Howard Hughes Medical Institute,University of Chicago, Chicago, IL.

Recent eQTL studies have provided a detailed description of inter-individual regulatory variation as well as considerable insight into the genetic basis for such variation. However, a limitation of all approaches based on measurements of steady-state gene expression levels alone is the inability to distinguish between variation in transcription and decay rates. To address this gap, we performed a genome-wide study of variation in transcript-specific mRNA decay rates across individuals. To do so, we used a time course study design to estimate mRNA decay rates for over 16,000 genes in 70 Yoruban HapMap lymphoblastoid cell lines (LCLs), for which steady-state gene expression level data have been previously characterized. As expected, we found that highly expressed genes are generally associated with lower mRNA decay rates. We identified sets of genes with similar functional roles that exhibit correlated rates of decay across individuals and are putatively co-regulated by mRNA degradation. Our approach also allowed us to estimate the relative contributions of transcription and degradation processes to overall regulatory variation between individuals. Specifically, we performed genome-wide mapping of decay rates and identified several dozen mRNA decay QTLs. We identified genetic factors underlying variation in both gene expression levels and decay rates, and found that nearly all of the mRNA decay QTLs are mapped in cis. Finally, by analyzing our data within the context of known eQTLs, we were able to distinguish between eQTLs due to variation in transcription or decay rates.

176

Human *cis*-regulatory SNPs (*cis*-rSNPs) altering transcription factor binding and gene expression. *V. Adoue[1], T. Kwan[1], B. Ge[1], M. Ouimet[2], V. Gagne[2], K. Gunderson[3], D. Sinnett[2], T. Pastinen[1].* 1) Human Genetics, McGill University, Montreal, Quebec, Canada; 2) Hôpital Ste-Justine, Université de Montréal, Montréal, Canada; 3) Illumina Inc, San Diego, CA.

Up to 30% of RefSeq transcripts are regulated by *cis*-rSNPs leading to phenotypic differences. These variants can be identified by mapping differences in allelic expression (AE) on Illumina1M/2.5M BeadChips. Mapped *cis*-rSNPs explain >50%; of population variance in AE. Our systematic approach to isolate causal *cis*-rSNPs includes integrative analysis of AE-mapping data in CEU and YRI LCLs with the intersection of common SNPs (1000 Genomes) and functional non-coding elements (wgENCODE) catalogues, along with ChIP-seq assays and assessing promoters and enhancers at high coverage. To allow measurement of variation at the population level, we have also pursued allele-specific assays on chromatin immunoprecipitation (AS-ChIP) using high density Illumina Omni 1M/2.5M BeadChips interrogating almost 3 million SNPs. Fine-mapped *cis*-rSNPs are enriched in regulatory elements and show predominant localization to the 5' transcription start site. Reporter gene assays validated 62% of promoter *cis*-rSNPs. Integrative genomic approaches can successfully isolate causal *cis*-rSNPs beyond promoters: *rs909685* is strongly associated with differential AE of SYNGR1, shows association by RNA-seq in LCLs, is located in active chromatin (DNaseI HS), and shows an allelic H3K4me1 signal in our ChIP-seq analysis. The allelic enhancer activity at this site was verified by reporter gene assays, underscoring the capacity of population genomic approaches in revealing function and variation of non-coding sequence elements. Multiple other sites show converging functional data suggesting specific mechanisms for common *cis*-rSNP action. Preliminary AS-ChIP results show patterns of allele-specific immunoprecipitation and converging allelic biases (up to 94%) by this genotyping chip-based approach as compared to next generation sequencing (ChIP-seq). However, despite numerous functional genomic datasets, only 50% of mapped *cis*-rSNPs yield straightforward hypotheses, suggesting the need for further tools for assessing non-coding DNA. Approximately 8% of disease SNPs in current GWAS catalogues can be fine mapped using our approach in LCLs. Additional information can be derived from our comparisons of *cis*-rSNPs observed in LCLs with those from other cell lineages and will provide a guide on how to more comprehensively enumerate causal *cis*-rSNPs in the human genome.

177

Genetic Variation of Gene Co-Expression Networks in Three Tissues.
A. Buil[1], A.K. Hedman[2], K. Small[3], E. Grundberg[4], A.C. Nica[1], T.D. Spector[3], M.I. McCarthy[2], P. Deloukas[4], E.T. Dermitzakis[1], the MuTHER Consortium. 1) Genetics and Development, University of Geneva, Geneva, Switzerland; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 3) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 4) Wellcome Trust Sanger Institute, Genome Campus, Hinxton, UK.

Genome-wide gene expression measures can be represented as a network of co-expressed genes. Our goal is to measure the effects of genetic variation on the topology of these networks. For the analysis we have used the sample from the MuTHER Project that consists of ~400 female twin pairs. For each individual we have HapMapII imputed genotypes and gene expression levels measured in LCLs, skin and fat using Illumina's expression array HumanHT-12 version 3. We built networks and defined modules of co-expressed genes for each tissue using the Weighted Gene Co-expression Network Analysis approach. We found that gene co-expression networks present large differences among tissues in their topology as well as in the modules of co-expressed genes observed in each tissue. Our results suggest that a fraction of the connections in these networks is genetically controlled. After correcting for the effect of tissue specific cis eQTLs, the global connectivity was reduced by 10%, 23%, 26% in the LCLs, fat and skin networks respectively. However, cis eQTLs common to the three tissues showed a weak effect on the global connectivity of the networks. On the other hand, 37 bona fide trans eQTLs in LCL reduced a 7% the total connectivity in the LCL network. We observed that genes that have tissue specific cis eQTLs with many trans effects are more connected in the gene co-expression networks. We estimated that the correlation between connectivity and a measure of trans activity for each gene is 0.19 (p=1e-12), 0.19 (p=1e-16) and 0.27 (p=4e-14) for LCLs, fat and skin respectively. However, this effect was negligible for cis eQTLs common to the three tissues. We found that cis eQTL have an average of 70% more trans effects on genes in his own co-expression module than in genes in other modules. Finally we found that highly connected genes are less prone to harbor cis eQTLs as well as disease SNPs. On the one hand, we saw that gene expression variation of highly connected genes is less likely controlled by a cis eQTL ($p<1e-16$). On the other hand, we observed that disease eQTLs common to the three tissues tend to be in genes with low connectivity (p=0.001).

178

Whole Omics Profiling Reveals Medical and Complex Molecular Phenotypes.
M. Snyder[1], R. Chen[1], H. Lam[1], J. Li-Pook-Than[1], G. Mias[1], L. Jiang[1], K. Karczewski[1], F. Dewey[2], M. Hariharan[1], S. Hillenmeyer[1], R. Haraksingh[1], M. O'Huallachain[1], L. Habbeger[3], R. Chen[1], M.J. Clark[1], J. Dudley[5], S. Balasubramanian[6], T. Kein[7], R. Altman[1], A. Butte[4], E. Ashley[8], M. Gerstein[3], H. Tang[1]. 1) Genetics, Stanford University, Stanford, CA; 2) School of Medicine, Stanford University, Stanford, CA; 3) Computational Biology & Bioinformatics, Yale University, New Haven, CT; 4) Pediatric Department, Stanford University, Stanford, CA; 5) Biomedical Informatics, Stanford University, Stanford, CA; 6) Molecular Biophysics and Biochemistry Department, Yale University, New Haven, CT; 7) PharmGKP, Stanford University, Stanford, CA; 8) Cardiovascular, Stanford University, Stanford, CA.

Personalized medicine requires the integrated analysis of genomic and other omics information. We have determined the whole genome sequence of an individual at high accuracy and performed an integrated analysis of omics profiles for a 16-month period that includes healthy and two virally infected states. Omics profiling of transcriptomes, proteomes, cytokines, metabolomes and autoantibodyomes have revealed dynamic and broad changes in molecular components that occurred during infection. In particular, analyses of heteroallelic expression and RNA editing suggested significant allele- and edit-specific expression changes in RNA and protein, in disease versus healthy states. More importantly, genomic information was also used to predict medical risks, including a previously unnoticed onset of diabetes, which was validated using standard clinical tests and molecular profiles. Our study is the first to relate personal genomic information to global functional omics activity for physiological and medical interpretation of healthy and disease states.

179

Hippocampal dysgenesis occurs frequently in Bardet-Bield patients independently from the BBS genotype. *H. Dollfus[1,2,5], V. Bennouna-Greene[3], S. Kremer[4], A. Danion[3], V. Marion[2], J-L. Dietemann[4], C. Stoetzel[2].* 1) Service de Génétique Médicale, CHRU Strasbourg F67098, France; 2) Laboratoire de Génétique Médicale, Equipe Avenir-INSERM EA3949, Université de Strasbourg, F67085, France; 3) Pôle de Psychiatrie et Santé Mentale, CHRU Strasbourg, F6709, France; 4) UMR-7191 (LINC), Service de Radiologie 2, Hôpital de Hautepierre, Université de Strasbourg, CHRU Strasbourg, F67098, France; 5) Centre de Référence pour les Affections Rares en Génétique Ophalmologique (CARGO), CHRU Strasbourg, F67091, France.

Introduction: The Bardet-Biedl syndrome (BBS) is an emblematic autosomal recessive ciliopathy defined by the association of retinitis pigmentosa, polydactyly, obesity, kidney disease and cognitive impairment. Indeed, developmental delay, mental retardation, speech deficiency and behavioral problems have been reported by many previous studies. However, variability between patients has been observed. The study was designed to evaluate at the best (taking into account the visual impairment) cognitive and psychiatric phenotypes as well as the brain morphology. Material and methods: The cognitive functioning, behavioral phenotype, prevalence of psychiatric diseases and memory performances in a cohort of 34 patients with BBS were evaluated in detail and a systemic brain magnetic resonance imaging (MRI) was performed. All patients were genotyped for known BBS genes. In parallel, the expression of 12 BBS mRNAs was studied in the Human brain and the Human hippocampus by qPCR. Results The patients cognitive performances were of marked variable efficiency for the whole cohort as well as the neuropsychological disorders observed in a subset of patients. The hippocampus dysgenesis was observed in 42 % of the patients. Mutations (respectively in BBS1, BBS2, BBS6, BBS9, BBS10, BBS12) were identified for at least two alleles for all except 2 patients. No phenotype genotype correlation could be evidenced. High expression of BBS genes was showed in the hippocampus compared to the brain. Conclusions: Hippocampal dysgenesis occurs frequently in BBS confirming the data of another group (Baker et al, 2011). The results point to the requirement of an intact primary cilium for hippocampal neurogenesis recently underlined by data obtained by other groups on the mouse. The exact cognitive impact in relation to the hippocampal abnormality remains to be addressed taking into account the visual handicap and the important variability of the neurocognitive phenotype. A major conclusion is also that so called “mental retardation” often described as an obligatory feature is not a systematic feature in BBS.

180

Exome capture reveals mutations in CEP164 causing a retinal-renal ciliopathy with a possible role in DNA damage response signaling. *M. Chaki[1], R. Airik[1], A.K. Ghosh[1], E.A. Otto[1], W. Zhou[1], T.W. Hurd[1], C. Antignac[2], S. Saunier[2], R.K. Koenekoop[3], R. Chen[4], F. Hildebrandt[1,5,6].* 1) Pediatrics - Nephrology, University of Michigan, Ann Arbor, MI; 2) Dept of Genetics, Hôpital Necker-Enfants Malades, Paris, France; 3) McGill Ocular Genetics Laboratory, McGill University Health Centre, Montreal, Quebec, Canada; 4) Dept of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 5) Dept of Human Genetics, University of Michigan, Ann Arbor, MI; 6) Howard Hughes Medical Institute.

Nephronophthisis-related ciliopathies (NPHP-RC) are recessive disorders featuring dysplasia or degeneration of retina, kidney and cerebellum. Cilia/centrosome localization of NPHP gene products characterizes them as `ciliopathies', but the disease mechanism remains poorly understood. To gain further insight into the pathogenesis of NPHP-RC, we sought to identify additional causative genes. Using homozygosity mapping followed by whole exome capture and massively parallel sequencing, we here identify recessive loss of function mutations of centrosomal protein *CEP164* in 4 different families as a novel cause of NPHP-RC. We demonstrate that Cep164 persists at the centrioles throughout mitosis and it localizes to the mitotic spindle poles and midbody during cell cycle. Our immunofluorescence studies reveal that Cep164 is located in mouse photoreceptor connecting cilia. Strikingly, we found that Cep164 additionally localizes to TIP60 positive nuclear foci. TIP60 is known to activate the DNA damage response protein ATM at sites of DNA damage. The dual expression of Cep164 at centrosomes and nuclear foci (Sivasubramaniam *et al.*, Genes & Dev 2008) points towards a disease mechanism of ciliopathy proteins that may relate to DNA damage response signaling. Our findings validate the importance of whole exome capture analysis for broadly heterogeneous single-gene disorders and provide new insights into mechanisms of degenerative diseases of retina and kidney.

181

Exome sequencing and cis-regulatory mapping identify mutations in *MAK*, a gene encoding a regulator of ciliary length, as a cause of retinitis pigmentosa. *A. den Hollander[1,2], A.M. Siemiatkowska[1,2], D. Yücel[3], C.A. Myers[4], R.W.J. Collin[1,2], M.N. Zonneveld[2], A. Beryozkin[5], E. Banin[5], C.B. Hoyng[1], L.I. van den Born[6], R. Bose[7], W. Shen[7], D. Sharon[5], F.P.M. Cremers[2], B.J. Klevering[1], R. Köksal Özgül[3], J.C. Corbo[4], the European Retinal Disease Consortium.* 1) Department of Ophthalmology, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands; 2) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 3) Institute of Child Health and Metabolism Unit, Department of Pediatrics, Hacettepe University, Ankara, Turkey; 4) Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO, USA; 5) Department of Ophthalmology, Hadassah-Hebrew University Medical Center, Jerusalem, Israel; 6) The Rotterdam Eye Hospital, Rotterdam, The Netherlands; 7) Division of Oncology, Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA.

A fundamental challenge in analyzing exome sequence data is distinguishing pathogenic mutations from background polymorphisms. To address this problem in the context of a genetically heterogeneous disease, retinitis pigmentosa (RP), we devised a candidate gene prioritization strategy called `cis-regulatory mapping' which utilizes ChIP-seq data for the photoreceptor transcription factor CRX to rank candidate genes. Exome sequencing combined with this approach identified a homozygous nonsense mutation in male germ cell-associated kinase (*MAK*) in the single affected member of a consanguineous Turkish family with RP. *MAK* encodes a cilium-associated MAP kinase whose function is conserved from the ciliated alga, Chlamydomonas reinhardtii, to humans. Mutations in *MAK* homologs in mice and other model organisms result in abnormally long cilia and, in mice, rapid photoreceptor degeneration. Subsequent sequence analyses of additional RP patients identified four probands with missense mutations in *MAK*. Two of these mutations alter amino acids which are conserved in all known kinases, suggesting that kinase activity is critical for MAK function in humans. This study highlights a previously underappreciated role for CRX as a direct transcriptional regulator of ciliary genes in photoreceptors. In addition, it demonstrates the effectiveness of CRX-based cis-regulatory mapping in prioritizing candidate genes from exome data, and suggests that this strategy should be generally applicable to a range of retinal diseases.

182

Exome sequencing and analysis of induced pluripotent stem cells identify the cilia-related gene MAK as a cause of retinitis pigmentosa. *B.A. Tucker[1], T.E. Scheetz[1], R.F. Mullins[1], A.P. DeLuca[1], J.M. Hoffmann[1], R.M. Johnston[1], S.G. Jacobsen[2], V.C. Sheffield[1,3,4], E.M. Stone[1,4].* 1) Department of Ophthalmology and Visual Sciences, University of Iowa Carver College of Medicine, Iowa City, Iowa; 2) Scheie Eye Institute, Department of Ophthalmology, University of Pennsylvania, Philadelphia; 3) Department of Pediatrics the University of Iowa Carver College of Medicine, Iowa City, Iowa; 4) Howard Hughes Medical Institute, the University of Iowa Carver College of Medicine, Iowa City, Iowa.

The use of exome and whole genome sequencing to identify disease genes presents an opportunity as well as a major challenge. The challenge is to identify disease-causing mutations among numerous non-pathological sequence variations. We used exome sequencing to identify a homozygous Alu insertion in exon 9 of MAK as the cause of disease in an isolated individual with retinitis pigmentosa (RP), an extremely genetically heterogeneous disease characterized by apoptotic death of photoreceptor cells leading to blindness. Subsequent screening of 1798 unrelated RP patients identified 20 additional probands homozygous for this insertion (1.2%). The insertion was not observed among 2952 controls. MAK encodes a kinase involved in the regulation of photoreceptor connecting cilium length. Immunohistochemistry of human donor tissue revealed that MAK is expressed in the inner segments, cell bodies and axons of rod and cone photoreceptors. Several isoforms of MAK that result from alternative splicing were identified. Functional proof that the Alu insertion interrupts normal MAK expression was obtained using induced pluripotent stem cells (iPSCs). iPSCs were derived from the skin of the proband and a control patient with non-MAK-associated RP (RP control). In the RP control patient, we found that a transcript lacking exon 9 was predominant in undifferentiated cells, while a transcript bearing exon 9 and a previously unrecognized exon 12 predominated in cells that were differentiated into retinal precursors. Notably, in the proband with the Alu insertion, the developmental switch to the MAK transcript bearing exons 9 and 12 did not occur. Besides demonstrating the utility of iPSCs to efficiently evaluate the pathogenicity of specific mutations in relatively inaccessible tissues like retina, this study reveals algorithmic and molecular obstacles to the discovery of pathogenic insertions and suggests specific changes in strategy that can be implemented to more fully harness the mutation identification power of new sequencing technologies.

183

The mutational load of the intraflagellar transport complex in ciliopathies. *E.E. Davis[1,2], D.S. Parker[1,3], J.R. Willer[1,3], C. Golzio[1,3], I.-C. Tsai[1,3], J. Hartley[4], K. Szymanska[5], A.C. Young[6], P. Cruz[6], P. Cherukuri[6], B. Maskeri[6], N.F. Hansen[6], J.C. Mullikin[6], R.W. Blakesley[6], G.G. Bouffard[6], D.M. Muzny[7], D.A. Wheeler[7], R.A. Lewis[8,9,10,11], C. Bergmann[12,13], E.A. Otto[14], S. Saunier[15], P.J. Scambler[16], P.L. Beales[16], E.R. Maher[4], T. Attié-Bitach[17], C.A. Johnson[5], F. Hildebrandt[14,18], R.A. Gibbs[7], E.D. Green[6], N. Katsanis[1,2,3].* 1) Center for Human Disease Modeling, Duke University Medical Center, Durham, NC; 2) Department of Pediatrics, Duke University Medical Center, Durham, North Carolina 27710, USA; 3) Department of Cell Biology, Duke University Medical Center, Durham, North Carolina 27710, USA; 4) Department of Medical and Molecular Genetics, Institute of Biomedical Research, University of Birmingham, Birmingham, United Kingdom; 5) Section of Ophthalmology and Neurosciences, Leeds Institute of Molecular Medicine, St. James's University Hospital, Leeds, United Kingdom; 6) NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA; 7) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA; 8) Department of Ophthalmology, Baylor College of Medicine, Houston, Texas 77030, USA; 9) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA; 10) Department of Pediatrics, Baylor College of Medicine, Houston, Texas 77030, USA; 11) Department of Medicine, Baylor College of Medicine, Houston, Texas 77030, USA; 12) Center for Human Genetics, Bioscientia, 55218 Ingelheim, Germany; 13) Department of Human Genetics, RWTH University of Aachen, 52074 Aachen, Germany; 14) Department of Pediatrics, University of Michigan, Ann Arbor, Michigan 48105, USA; 15) Inserm, U983, Université Paris Descartes, Hôpital Necker, Paris, France; 16) Molecular Medicine Unit, Institute of Child Health, University College London, London WC1N 1EH, United Kingdom; 17) Département de Génétique et INSERM U-781, Hôpital Necker-Enfants Malades, Université Paris Descartes, Paris Cedex 15, France; 18) Howard Hughes Medical Institute and Department of Human Genetics, University of Michigan, Ann Arbor, Michigan 48105, USA.

Structural and functional defects in cilia result in a broad range of phenotypically overlapping human genetic disorders, the ciliopathies, that manifest with highly variable penetrance and expressivity. These disorders are also underscored by substantial genetic overlap in which lesions at one locus can contribute either causal or modifying alleles to multiple clinical entities. These observations have raised the possibility that a significant contributor to the specification of clinical phenotypes might not be allelism at a single locus, but the cumulative mutational load across the module. To explore this hypothesis, we conducted medical resequencing of all genes that encode a biochemically characterized functional module in the cilium, the intraflagellar transport complex (IFT), in a clinically diverse cohort of 384 ciliopathy patients encompassing the mild, moderate, and severe aspects of the phenotypic spectrum. Here, we report the comprehensive functional analysis of 165 different novel nonsynonymous changes using both in vivo complementation in a zebrafish model overlaid with in vitro complementation assays measuring structural defects in ciliated mammalian cells. These studies indicate that pathogenic lesions in multiple IFT genes are sufficient to cause each of Meckel-Gruber Syndrome, Jeune Asphyxiating Thoracic Dystrophy, and isolated nephronophthisis under a recessive model of inheritance. In addition, ~15% of patients harbor pathogenic heterozygous variants in multiple IFT genes in trans implicating the combined effect of mutational load across the functional system in the manifestation of disease. Moreover, we found a non-uniform distribution of mutations, with a markedly increased load in IFT complex A (retrograde) versus complex B (anterograde) proteins. Finally, in support of a systems-based approach to understand the clinical variability in ciliopathies, we observed a progressive increase in the mutational load of this module within each clinical subgroup that is consistent with disease severity. This study potentially represents the first saturated, functionally annotated interrogation of a functional cellular module and suggests that the interpretation of mutational distribution in biological systems can inform the variable penetrance and expressivity that typify most human genetic disorders.

184

Genetic interactions revealed by mouse models of CEP290 ciliopathies. *A. Swaroop[1], R. Rachel[1], A. Hackett[1], H. May-Semera[2], L. Dong[1], T. Friedman[2], M. Kelley[2].* 1) National Eye Inst, NIH, Bethesda, MD; 2) National Institute of Deafness and Communicative Disorders.

Mutations in centrosomal/ciliary protein CEP290 cause a broad spectrum of phenotypes resulting in ciliopathies, including Leber congenital amaurosis. Understanding the function of specific domains of CEP290 should shed light on disease pathogenesis and allow treatment paradigms, as will additional knowledge of interactions between Cep290 and other ciliopathy genes. We characterized brain, cerebellar, and retinal degeneration phenotypes of mice with two Cep290 alleles: knockout (null), and rd16 (ΔDSD). Knockout mice were generated by replacing exons 1-4 of the Cep290 gene with a lacZ neo cassette, resulting in a null allele. Cep290rd16 mice have a spontaneous in-frame deletion of exons 35-39 resulting in removal of the DSD domain. This domain interacts with several other ciliopathy-related proteins including MKKS. Brain phenotypes and development and degeneration of the retina were evaluated by histology, immunohistochemistry, and electron microscopy in both lines, including compound heterozygotes. Cep290-ko mice develop hydrocephalus around weaning age, resulting in postnatal death in 80% of homozygotes. Age of death shows is correlated with genetic strain, with mice having a mixed C57BL/6 and 129/SvJ background showing more robust viability. In addition, we crossed mice carrying each Cep290 allele with mice having a null allele of Mkks (Bbs6). We find that in mice doubly mutant for the hypomorphic Cep290-rd16 and Mkks-ko allele, the retinal phenotype is ameliorated and both single and double homozygotes have a normal lifespan. However, mice having null mutations in both genes (Cep290-ko and Mkks-ko) results in prenatal lethality of double homozygotes, and reduced viability of tri-allelic animals. These findings point to complex interactions between Cep290 and other ciliopathy genes, and suggest that viable treatment options for various conditions should not be limited to the defective gene. Such treatment concepts widen the variety of therapeutic considerations for many ciliopathy patients.

185

The centrosomal protein ninl functions in ciliogenesis and acts upstream of mkks/bbs6 in establishing planar cell polarity in zebrafish development. *H. Kremer[1], E. van Wijk[1], N.A. Zaghloul[2,3], R. Bachmann[4], F.F.J. Kersten[5,6], J.M. Gerdes[2], D.A. Mans[5], T.A. Peters[1], H.H. Arts[5], E. Davis[2], C.C. Leitch[2], H. May-Simera[7], P.L. Beales[7], D. Doherty[8], C.B. Moens[4], N. Katsanis[9], R. Roepman[5].* 1) Dept Otorhinolaryngology, Radboud Univ Nijmegen Med Ctr, Nijmegen, Netherlands; 2) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA; 3) Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA; 4) Howard Hughes Medical Institute and Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, USA; 5) Dept. Human Genetics, Radboud Univ Nijmegen Med Ctr, Nijmegen, Netherlands; 6) Dept. Ophthalmology, Radboud Univ Nijmegen Med Ctr, Nijmegen, Netherlands; 7) Molecular Medicine Unit, Institute of Child Health, University College London, UK; 8) Dept. Pediatrics, University of Washington School of Medicine and Seattle Children's Hospital, Seattle, USA; 9) Center for Human Disease Modeling, Dept Cell Biology, Duke University, Durham, NC, USA.

Primary cilia, the sensory organelles present on the surface of most eukaryotic cells, are involved in the interpretation and propagation of an increasing number of receptor-dependent paracrine signalling cascades. Dysfunction of the ciliary basal body by downregulation of centrosomal proteins often leads to defective cilium formation and/or function. We recently identified the centrosome and basal body-associated ninein-like protein (NINL) as a molecular connector of the retinal ciliopathies Usher syndrome (USH) and Leber congenital amaurosis (LCA). NINL is involved in microtubule anchoring, nucleation and outgrowth in cultured cells. In this study, we have scrutinized the molecular disease mechanism that underlies both USH and LCA by studying ninl function during zebrafish embryonic development. Downregulation of ninl expression induced defects in ciliogenesis in cultured IMCD3 cells and in photoreceptor outer segment formation in zebrafish embryos. In addition, we show that ninl acts in the switch between canonical and non-canonical Wnt signaling pathways, more specifically upstream of mkks/bbs6, as ninl depletion perturbed planar cell polarity and simultaneously upregulated) -catenin-dependent Wnt signaling. The severity of the ciliary defects in photoreceptor cells indicates that ninl acts upstream of mkks/bbs6, lebercilin, and ush2a. Several pieces of evidence suggest that this role is linked to (vesicular) protein transport.

186

Identification of the genetic basis of heterotaxy-spectrum congenital heart defects. *J. Cowan[1], M. Tariq[1], A. Cast[1], J. Lander[1], J.A. Towbin[1], B. Mohapatra[2], J.W. Belmont[2], C. Shaw[2], T. Smolarek[1], S. Lalani[2], S.M. Ware[1].* 1) Cincinnati Children's Hospital, Cincinnati, OH; 2) Baylor College of Medicine, Houston, TX.

Heterotaxy is a multiple congenital anomaly syndrome resulting from abnormalities of the proper specification of left-right (LR) asymmetry during embryonic development. It is characterized by complex cardiovascular malformations that may occur as syndromic or isolated defects. Although heterotaxy is the most highly heritable congenital heart disease, the majority of cases remain unexplained indicating a need for novel approaches to identify the molecular underpinnings of this heritable disorder. We undertook comprehensive genetic analyses using whole genome SNP microarray, genome wide array based comparative genomic hybridization (aCGH), and targeted resequencing in patients with heterotaxy and heterotaxy-spectrum congenital heart disease. We analyzed copy number variation (CNV) in 224 unrelated heterotaxy subjects by SNP microarray and aCGH. Disease causing abnormalities ranged from large cryptic unbalanced translocations to small, single exon deletions in *ZIC3*, the gene known to cause the X-linked form of heterotaxy. We identified novel, rare CNVs of potential clinical significance in 39% of patients. Functional analysis of candidate genes within CNVs was performed by developmental gene expression analyses and morpholino knockdown in *Xenopus laevis*. Morphants for *Rock2*, *Tgfbr2*, and *Meox2* showed abnormal molecular markers of LR asymmetry as well as abnormal heart and gut looping, suggesting these as novel genes associated with heterotaxy. In parallel, we sequenced 9 genes in the NODAL signal transduction pathway, a critical developmental pathway controlling LR asymmetry, in over 227 sporadic heterotaxy cases and 298 controls. Logistic regression controlling for ethnicity demonstrated a significant association of non-synonymous rare variants in NODAL pathway genes with affected cases (odds ratio = 4.1; 95% CI 2.2 - 7.6; p<.0001). Taken together, the results indicate a high diagnostic yield for both isolated and syndromic congenital heart defects, identify novel pathogenic submicroscopic gains and losses, delineate loci with strong novel candidates for heterotaxy, and illustrate the importance of combined diagnostic approaches to maximize the understanding of genetic architecture, genetic heterogeneity, and disease association.

187

Quantitative Trait Analysis in the NHLBI Exome Sequencing Project. *D.Y. Lin[1], D. Zeng[1], L. Lange[2], P. Auer[3], C. Carlson[3], R. Jackson[4], K. North[5] on behalf of the NHLBI Exome Sequencing Project.* 1) Dept Biostatistics, CB #7420, Univ North Carolina, Chapel Hill, NC; 2) Dept Genetics, Univ North Carolina, Chapel Hill, NC; 3) Fred Hutchinson Cancer Research Center, Seattle, WA; 4) College of Medicine, Ohio State University, Columbus, OH; 5) Dept Epidemiology, Univ North Carolina, Chapel Hill, NC.

It is not economically feasible to sequence all study subjects in a large cohort. A cost-effective strategy is to sequence only the subjects with the extreme values of a quantitative trait. Such trait-dependent sampling poses considerable analytical challenges, especially in the meta-analysis of quantitative-trait data across sequencing studies with different selection criteria. In the NHLBI Exome Sequencing Project, 267 subjects with BMI values >40 and 178 subjects with BMI values <25 (and without diabetes) were selected for sequencing out of a total of 11,468 subjects from the Women's Health Initiative. Similar designs were used for LDL and blood pressures, although the sampling was based on age- and gender-adjusted residuals rather than raw measurements. The traditional case-control design was used for early-onset MI. The case-control testing is a valid option for comparing the two extremes of a quantitative trait, such as BMI, LDL and blood pressures, but will be inefficient if the underlying association is quantitative. In addition, efficiency will be dramatically improved if the analysis of a particular quantitative trait, say BMI, uses not only its own sample (i.e., the BMI sample) but also all other samples (i.e., the LDL, blood-pressure and MI samples). This is extremely important because there is very little power to detect rare variants in small samples. Because the sampling depends on the trait values, standard analysis methods, such as least-squares estimation, are not appropriate. We have developed valid and efficient statistical methods for quantitative trait analysis under complex trait-dependent sampling. Our methods can be used to perform quantitative trait analysis not only for the trait that is used to select the subjects for sequencing but also for any other quantitative traits that are measured. For any particular quantitative trait, the association results from all studies with measurements of that trait are combined through meta-analysis. The efficiency gain of such meta-analysis over the analysis of each trait based on its own sample alone is enormous and can lead to many more discoveries. The advantages of the new methods over the existing ones are demonstrated through simulation studies and analysis of data from the NHLBI Exome Sequencing Project. The relevant software is freely available.

188

Whole Genome Sequencing of 1000 Individuals in an Isolated Population. *C. Sidore[1,2,3], S. Sanna[3], F. Busonero[1,3], W. Chen[1], H.M. Kang[1], C. Fuchsberger[1], F. Reinier[4], R. Berutti[2,4], D. Hovelson[1], M.F. Urru[4], M. Marcelli[4], R. Cusano[3], M. Oppo[4], A. Maschio[1,3], M. Pitzalis[4], M. Zoledziewska[4], A. Angius[4], R. Nagaraja[5], M. Uda[3], D. Schlessinger[5], C. Jones[4], F. Cucca[2,3], G. Abecasis[1].* 1) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 2) Dipartimento di Scienze Biomediche, Università di Sassari, Sassari, Italy; 3) Istituto di Ricerca Genetica e Biomedica, CNR, Monserrato(CA), Italy; 4) Center for Advanced Genomics Computing Technology, Pula (CA), Italy; 5) NIA, Laboratory of Genetics, Baltimore, MD.

Genome-wide association studies (GWAS) have furthered our understanding of the molecular basis of many complex traits by finding identifying new disease associated loci and pathways. However, GWAS studies based commercial genotyping arrays and imputation fail to capture rarer variants and, in particular, population specific variants. To advance our understanding of the genetics of a variety of traits in the Sardinian population, we are studying a sample of >6,000 individuals recruited from the population of a cluster of 4 small towns. Using whole genome sequencing, we sequenced DNAs from ~1000 individuals enrolled either in this project or in a parallel project to study autoimmune diseases on the island, at an average depth of coverage of >4X. A preliminary analysis of sequence data from 505 individuals has identified and permitted the genotyping of >13M single nucleotide polymorphisms. Genotyping error rate (which is expected to decrease as additional individuals are sequenced) is currently ~0.5% in this preliminary analysis of 505 individuals (in earlier analyses of 66 and 226 individuals, the genotype error rate was estimated at 4.2 and 1.5%, respectively). To increase the power to detect association, we are using haplotypes derived by sequencing of these individuals to impute missing genotypes in other individuals in the cohorts under study, who have already been genotyped with Affymetrix arrays and the Metabochip. Our results show that Sardinian specific loci, missing in European populations, can be identified using imputation, without the need for additional genotyping. Furthermore, we find that genotypes from the initial 505 individuals can already impute missing genotypes for the rest of the cohort more accurately than an equal size reference panel of individuals from elsewhere in Europe. For example, alleles with 1-3% frequency can be imputed with an average r2 > 0.90 (versus 0.75 with a comparable panel of haplotypes from elsewhere in Europe). We will present results of our analysis of blood lipid levels, where our results already show that a more complete examination of genetic variation in GWAS loci previously identified in Sardinia as well as elsewhere greatly increases the proportion of genetic variance explained. We will also present how our current strategy allows the rediscover of a coding variant (Q39X) specific to Sardinia, known to affect LDL by direct sequencing, but undetectable using standard GWAS genotyping and imputation.

189

Likelihood Based Deletion Analysis in a Sample of Sequenced Sardinian Individuals. *S. Rashkin[1], T. Blackwell[1], C. Sidore[1,2,3], S. Sanna[3], F. Busonero[1,3], W. Chen[1], H.M. Kang[1], C. Fuchsberger[1], F. Reinier[4], R. Berutti[2,4], F. Deidda[2], M.F. Urru[4], M. Marcelli[4], R. Cusano[3], M. Oppo[4], A. Maschio[3], M. Pitzalis[4], M. Zoledziewska[4], A. Angius[4], R. Nagaraja[5], M. Uda[3], D. Schlessinger[5], C. Jones[4], F. Cucca[2,3], G. Abecasis[1].* 1) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 2) Dipartimento di Scienze Biomediche, Università di Sassari, Sassari, Italy; 3) Istituto di Ricerca Genetica e Biomedica, CNR, Monserrato, Cagliari, Italy; 4) CRS4, Center for Advanced Genomics Computing Technology, Pula (CA), Italy; 5) NIA, Laboratory of Genetics, Baltimore, MD, USA.

Copy number variants (CNVs) are regions of the genome where the number of copies of the region varies between individuals, such as deletions and duplications. It is now clear that CNVs can be associated with susceptibility or resistance to a variety of diseases. We reasoned that whole genome sequence data would provide a powerful strategy for analysis of the patterns of structural variation across the entire genome. Here, we summarize an analysis of >500 sequenced individuals (each sequenced with 100 to 120-bp paired end reads to an average depth of >4x).

We used a likelihood based method to estimate copy number for each individual at sites overlapping CNVs described by the 1000 Genomes Project. The method uses read depth information for each individual (and, in particular, the fraction of reads overlapping each potential deletion) to estimate copy number, using a set of parameters estimated from the full sample. When applied to a set of 1000 Genome Project individuals with known genotypes at CNV loci, we estimated that our method has ~93% accuracy (compared to calls generated independently by the 1000 Genomes Project).

Here, we will report on the patterns of copy number variation observed in the sequenced Sardinian individuals.

190

Sequencing and Genotyping Thousands of European Genomes and Exomes to Better Understand the Genetic Architecture of Type 2 Diabetes: the GoT2D Study. *H.M. Kang[1], K. Gaulton[2], B.F. Voight[3], C. Fuchsberger[1], R.D. Pearson[2], J. Maguire[3], T. Teslovich[1], Y. Chen[2], J. Flannick[3], L.J. Scott[1], L. Moutsianas[2], R. Poplin[3], P. Chines[1], J. Perry[2], C. Sidore[1], M. Rivas[2], T. Blackwell[1], I. Prokopenko[2], T. Fennell[3], G. Jun[1], T. Frayling[2], N. Burtt[3], G.R. Abecasis[1], P. Donnelly[2], L. Groop[3], M. Boehnke[1], M. McCarthy[2], D. Altshuler[3] on behalf of the Genetics of Type-2 Diabetes (Go-T2D) Consortium.* 1) Department of Biostatistics, University of Michigan, Ann Arbor, Ann Arbor, MI, USA; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 3) Medical Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA.

In the past five years, >40 common SNP variants have been identified as associated with type 2 diabetes (T2D), spurred by the rapid advance of genotyping technologies. Advances in next-generation sequencing technologies now allow a more comprehensive investigation of the genetic architecture of T2D across a more complete spectrum of genetic variation including lower frequency SNP variants as well as indels and structural variants. In the Genetics of Type 2 Diabetes (GoT2D) study, we aim to identify genetic variants associated with T2D via low-coverage whole-genome sequencing, deep exome sequencing, and next-generation 2.5M SNP chip genotyping of 1,325 cases and 1,325 controls selected in the phenotype extremes of T2D from Northern Europe, followed by genotype imputation into an additional ~28,000 individuals.

Preliminary analysis of 442 individuals identified >14 million SNPs from low pass whole genome sequencing and >120,000 coding variants from exome sequencing with 99.4% genotype concordance between these approaches. Combining the sequence data with the 2.5M SNP chip genotypes and imputing genotypes into >9,000 samples, we (a) recapitulated many previously identified T2D association signals, (b) observed substantial enrichment of low-frequency SNPs in sequenced T2D cases at loci involved in monogenic forms of T2D, and (c) observed that by collecting samples in the phenotypic extremes of T2D, effect sizes for known T2D risk variants were substantially increased.

At the time of abstract submission, we have sequenced 1,520 low-coverage genomes and 1,067 deep exomes, and have genotyped 2,350 samples on 2.5M SNP chip. Sequencing and genotyping of all 2,650 samples is expected to be complete in September. With these data, we will have >80% power to identify T2D-associated loci with genotype relative risks of 1.3 and 1.6 at 5% and 1% minor allele frequency (MAF), respectively, at genome-wide significance. We expect our results to extend our understanding of the genetic architecture of T2D, and to provide new tools, resources, and study designs to accelerate the unraveling of the genetic basis of many complex diseases and traits.

191

A longevity reference genome generated from the world's oldest woman. *H. Holstege[1], D. Sie[2], T. Harkins[3], C. Lee[3], T. Ross[3], S. McLaughlin[3], M. Shah[3], B. Ylstra[2], G. Meijer[2], H. Meijers-Heijboer[1], P. Heutink[1], S. Shaw Murray[4], M. Reinders[5], G. Holstege[6], E. Sistermans[1], S. Levy[4].* 1) Department of Clinical Genetics, VU University Medical Center, van der Boechorststraat 7, 1081 BT, Amsterdam, the Netherlands; 2) Department of Pathology, VU University Medical Centre, de Boelelaan 1117, 1081 HV, Amsterdam, the Netherlands; 3) Life Technologies, 500 Cummings Center, Suite 2450, Beverly MA, 01915, USA; 4) Scripps Translational Science Institute, Scripps Genomic Medicine,, San Diego CA, 92037, USA; 5) Delft Bioinformatics Lab, Delft University of Technology, Mekelweg 4, 2628 CD Delft; 6) Center for Uroneurology, University of Groningen, University Medical Center Groningen, 9700 RB Groningen, The Netherlands.

Much effort is currently spent on identifying the genetic variants that are associated with longevity using genetic association approaches. New insight suggests that for genome-wide association studies a reference genome with the same trait is indispensible. Therefore, an anonymous reference, such as Hg19, could introduce a reference bias obscuring the detection of the true variants for longevity. Therefore, we generated a de novo assembly of a unique genome specific for the coding of a long, healthy life, which represents a more accurate biological reference genome for longevity projects. We had the opportunity to study the genome of a woman, W115 who, at the time of her death, was the oldest human being alive in the world. She lived until the age of 115 years without experiencing any problems associated with vascular disease or dementia. Her performance in neuropsychological examinations at the age of 112-113 was above average for healthy adults between 60-75 years. Post-mortal examination showed that she died of a stomach tumor but more importantly, that she had almost no signs of atherosclerotic plaque formation, which is one of the leading causes of death among older people, and the associated vascular dementia is a leading cause of dementia. Using paired end and long mate pair reads adding up to more than 120x genomic coverage, we assembled a de novo reference genome and an extensive catalog of all variants including SNV's, CNV's, and large and small structural variations. Initial analysis showed that this genome has an overrepresentation of rare alleles in DNA variants associated with atherosclerosis, heart disease, Parkinson's Disease and Alzheimer's Disease. Extending beyond personal genomes, W115 will be used as s unique reference genome for population studies focused on longevity.

192

Joint Variant Calling and Analysis across >4,000 Exomes of European and African American Ancestry. *G. Jun[1], H.M. Kang[1], M. Rieder[2], M. DePristo[3], G. Abecasis[1] on behalf of the NHLBI Exome Sequencing Project.* 1) Department of Biostatistics, Univ Michigan, Ann Arbor, Ann Arbor, MI; 2) Department of Genome Sciences, University of Washington, Seattle, WA, USA; 3) Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Boston, MA, USA.

Deep sequencing of human exomes is an effective means to identify causative genetic variants in Mendelian disorders. Enabled by improvements in sequencing technology, large-scale exome sequencing efforts are now using deep exome sequencing to study the role of rare coding variants affecting in complex human diseases and traits. As part of the NHLBI GO Exome Sequencing Project (ESP) we are using next-generation sequencing of the protein coding regions of the human genome in diverse rich-phenotyped populations to study the genes and mechanisms contributing to heart, lung, blood disorders. Jointly analyzing coding variants across thousands of exomes is computationally demanding and requires integration of data generated across different sites, using different capture reagents and DNA sample preparation protocols. Here, we described our scheme for detecting variant sites and calling genotypes for each individual. Our process starts with alignment of sequence reads to the genome, indel realignment and quality score recalibration, generation of compact genotype likelihoods for each individual, and joint analysis of these likelihoods across all individuals. After this initial analysis phase, a series of filtering steps are used to flag sites that appear to perform differentially across sequencing centers. The results of the calling process provide a starting point for detailed analysis of ESP phenotypes but also allow insights into human evolutionary and demographic history as described in a companion abstract by Bigham et al. on behalf of the ESP Population Genetics Project Group. By the time of abstract submission, we performed initial joint variant discovery across 2,520 exomes in European and African American ancestry on extended (±50bp) target regions using our efficient multi-sample variant calling pipeline based on genotype likelihoods. Our results identify >600,000 coding and >350,000 non-coding SNP variants across the 2,520 exomes. Among coding variants, ~60% are predicted to change protein sequences, and the majority of these are singletons. By the time of presentation, we expect to obtain a more comprehensive profile of coding variation in collaboration with analysis groups by delivering the results of our variant calling pipeline on >4,000 exomes sequenced at the time of abstract submission.

193

SNVer: a statistical tool for variant calling in analysis of pooled or individual next-generation sequencing data. *Z. Wei[1], W. Wang[1], P. Hu[2], G.L. Lyon[3], H. Hakonarson[3].* 1) Department of Computer Science, New Jersey Institute of Technology, Newark, NJ; 2) The Centre for Applied Genomics (TCAG) and Program in Genetics and Genome Biology, the Hospital for Sick Children, Toronto, ON, Canada; 3) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.

We develop a statistical tool SNVer for calling common and rare variants in analysis of pooled or individual next-generation sequencing (NGS) data. Different from pre-existing methods, we formulate variant calling as a hypothesis testing problem and develop a unified statistical framework for both pooled and individual NGS data. Most variant calling methods simply discard loci with low depth of coverage to achieve reliable variant calls. Our statistical model does not discriminate against poorly covered loci. Loci with any (low) coverage can be tested and depth of coverage will be quantitatively factored into final significance calculation. SNVer reports one single overall p-value for evaluating the significance of a candidate locus being a variant, based on which multiplicity control can be obtained. This is particularly desirable because tens of thousands of loci are simultaneously examined in typical NGS experiments. Each user can choose the type I error rate threshold he or she considers appropriate, instead of just the dichotomous decisions of whether to “accept or reject the candidates” provided by most existing methods. We apply our method to analyze three real datasets, covering three sequencing platforms: Illumina, SOLiD and 454, three diseases: Attention Deficit/Hyperactivity Disorder (ADHD), Autism and Type 1 Diabetes (T1D), and two sequencing strategies: pooled targeted re-sequencing (Autism, T1D) and individual whole-exome sequencing (ADHD). Our analysis results demonstrate the superior performance of our program in comparison with existing methods including CRISP, GATK, SAMtools and SOAP. SNVer is implemented in Java, which therefore is platform-free and easy to install and run. Our program accepts the standard pileup file produced by SAMtools as input, and outputs variant calls in the standard VCF format. This user-friendly feature makes upstream pre-processing easy and downstream analysis convenient, and therefore it can be integrated seamlessly into any NGS data analysis pipelines. SNVer runs very fast and can complete testing 300K loci within an hour. This excellent scalability makes it feasible for analysis of whole-exome sequencing data, or even whole-genome sequencing data using high performance computing cluster.

194

Improvement of Haplotype Phasing using Next-Generation Sequence Data. *F. Zakharia, J. Kidd, S. Gravel, C.D. Bustamante.* Stanford University, Stanford, CA.

Accurate identification of haplotypes in sequenced human genomes can provide invaluable information about population demography and fine-scale correlations along the genome, thus empowering both population genomic and medical association studies. While haplotype reconstruction from genotype data is generally straightforward and highly accurate in related individuals (such as trio data), it remains a challenging problem in the case of unrelated individuals—generally yielding switch errors between approximately 6% of neighboring heterozygous sites. The advent of massively parallel sequencing technologies and the increasing affordability of these methods present a unique opportunity to improve existing phasing algorithms. We present a novel statistical method that expands on fastPHASE to incorporate phasing information from paired-end read data. The algorithm harnesses the haplotype cluster information estimated by fastPHASE from genotypes across the population to sample the most likely haplotypes compatible with the sequencing data. This constraint on the search space of possible haplotypes comes at little added computational cost. Applying the method to HapMap individuals genotyped on the Affymetrix 2.5M Axiom chip, we found that the inclusion of simulated paired end read data significantly improved phasing. In particular, heterozygous sites spanned by mated reads were phased at higher accuracy versus those without sequencing data (with average switch errors of 2.0-2.6% for inserts ranging from 300bp to 2kb). As expected, longer read sizes and higher throughput translated to greater decreases in switching error, as did higher variance in the size of the insert separating the two reads—suggesting that multi-platform next generation sequencing may be exploited to yield particularly accurate haplotypes. To assess the performance of the method on empirical sequencing data, we also applied the algorithm to parents from trios sequenced by Complete Genomics; preliminary results suggest that despite the relatively small size of the reads and mate gaps (35bp and 300bp, respectively), the algorithm still yields switch error rates that are about 10% lower versus fastPHASE. Overall, the phasing improvements afforded by this new method highlight the power of genotyping by sequencing (GBS) for reconstructing haplotypes in unrelated individuals.

195

The Roadmap Epigenomic Mapping Consortium: a Community Human Epigenomics Resource. *L. Chadwick, The NIH Roadmap Epigenomic Mapping Consortium.* Division of Extramural Research and Training, National Institute of Environmental Health Sciences, Research Triangle Park, NC.

The NIH Roadmap Epigenomics program (http://commonfund.nih.gov/epigenomics) was established around the hypothesis that epigenetics plays a critical role in human health, and that epigenetic misregulation may be an important contributor to disease pathogenesis. The program supports development of epigenetic analysis and imaging technologies, investigations into the epigenetic basis of a variety of diseases, and the NIH Roadmap Epigenomic Mapping Consortium, a group of four Reference Epigenome Mapping Centers (REMCs) and an Epigenomics Data Analysis and Coordination Center (EDACC). The purpose of this consortium is to develop a well-curated community resource containing reference epigenomic maps for a variety of primary human cell and tissue types, including both adult and fetal tissues, as well as a number of embryonic and induced pluripotent stem cells. An additional goal of the program is to make protocols and data quality metrics freely available. As of the time of writing (May 2011), at least one epigenomic data set was available for over one hundred unique human cell or tissue types. Ultimately, the majority of reference epigenomes generated will contain genome-wide information about DNA methylation, a core set of the most informative histone modifications (5-6 different modifications), chromatin accessibility, and gene expression. As of May 2011, fifteen such reference epigenomes were available. A subset of reference epigenomes will also contain an expanded set of at least twenty additional histone modifications. As of May 2011, one such reference epigenome was available, with others nearing completion. Additional data sets are being generated continuously. The data generated by the program can be downloaded and used freely by the scientific community. RoadmapEpigenomics.org (http://roadmapepigenomics.org) serves as the home page for the consortium. All protocols, information about quality metrics, as well as links to the data and a variety of analysis tools can be found at this site. In addition, the program has developed a number of options for visualizing and downloading the data, which will be described in this presentation (links to all of these sites can be found at http://roadmapepigenomics.org).

196

Dissecting the role of NLRP7 in regulation of imprinting in the trophoblast. *S. Wen[1], S. Mahadevan[1], S. Otta[1], A. Balasa[2], H. Peng[1], I. Van den Veyver[1,3].* 1) Department of OB/GYN, Baylor College of Medicine, Houston, TX; 2) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Women with autosomal recessive mutations in NLRP7 in 19q13.42 have recurrent biparental hydatidiform moles with loss of DNA methylation at various paternally maternally methylated imprinting control regions (ICRS). Thus we hypothesize that NLRP7 may play a critical role in establishment of imprinting during oogenesis or its maintenance in early development and used complementary approaches to address this hypothesis. First, we used overexpression and co-immunoprecipitation in human cell lines to screen candidate genes that have a known function related to the imprinting process. We found that CTCF and YY1 can bind to NLRP7. CTCF and YY1 can form a complex, directly bind to DNA, and regulate the methylation status of imprinted genes. Hence, we performed chromatin-immunoprecipitation with an antibody to overexpressed NLRP7 and detected immunoprecipitation of chromosome regions near imprinted genes to which CTCF also binds. Second, as NLRP7 is not present in the rodent genome, we established and used human embryonic stem cell (hESC) differentiation into trophoblast as a model to examine the function of NLRP7 in trophoblast differentiation. We performed transient and stable knockdown of endogenously expressed NLRP7 in these cells. Upon differentiation of the NLRP7-knockdown hESCs, cell proliferation rates are not affected, but the RNA level of a trophoblast marker CGB5 (chorionic gonadotropin beta 5 polypeptide) is significantly increased. In the culture media, the secreted hCG level is also significantly increased. We are currently performing bisulfite sequencing of paternally imprinted genes to determine if knockdown of NLRP7 affects the imprinting status of paternally imprinted loci. Finally, we will use 450K Infinium Methylation BeadChips to analyze at a genome wide level if NLRP7 knockdown affects methylation of other DNA regions during trophoblast differentiation. Together, our data support that NLRP7 interacts with imprinting marks in the developing embryo and affects trophoblast differentiation. Additional studies are needed to determine the exact mechanisms by which it performs this function.

197

Aberrant methylation by mutations of DNA methyltransferase 1 cause peripheral and central axonal degeneration. *C. Klein[1], M. Botuyan[2], Y. Wu[3], C. Ward[4], G. Nicholson[5], S. Hammans[6], K. Hojo[7], H. Yamanishi[7], A. Karpf[8], S. James[8], D. Wallace[9], M. Simon[9], C. Lander[10], J. Cunningham[3], G. Smith[11], W. Litchy[1], B. Boes[12], E. Atkinson[13], S. Middha[13], P. Dyck[1], J. Parisi[14], G. Mer[2], D. Smith[3], J. Dyck[1].* 1) Department of Neurology, Mayo Clinic, Rochester, Rochester, MN.USA; 2) Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester MN, USA; 3) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester MN, USA; 4) Nephrology and Hypertension Research, Mayo Clinic, Rochester MN, USA; 5) University of Sydney, Molecular Medicine Laboratory & ANZAC Research Institute, Australia; 6) Southampton University Hospitals NHS Trust, Department of Neurology, Southampton, United Kingdom; 7) Division of Neuropsychiatry, Harima Sanatorium, Hyogo, Japan; 8) Roswell Park Cancer Institute, Department of Pharmacology and Therapeutics, Buffalo, NY, USA; 9) Center for Molecular & Mitochondrial Medicine and Genetics, University of California, Irvine, USA; 10) Queensland Health, Royal Brisbane Hospital, Herston, Australia; 11) Division of Psychology, Mayo Clinic, Rochester, MN; 12) Roche Applied Science Genomic Sequencing, Indianapolis, IN, USA; 13) Biomedical informatics and Statistics, Mayo Clinic, Rochester, MN, USA; 14) Department of Neuropathology, Mayo Clinic, Rochester, MN, USA.

DNA methyltransferase 1 (DNMT1) is an essential component of genomic methylation. Neural gene expression, DNA mismatch repair and cell cycle regulation are all influenced by DNA methylation. Here we show mutations in DNMT1 cause both central and peripheral axonal degeneration in one form of hereditary sensory and autonomic neuropathy with dementia and hearing loss. Two American, one European and one Japanese kindreds were studied. Next generation sequencing led to the identification of two DNMT1 mutations, c.A1484G (p.Y495C) in the American and Japanese kindreds and c.1470TCC-1472ATA (p.D490E-P491Y) in the European kindred. All mutations are within targeting sequence domain of DNMT1. Functional analysis showed premature degradation of mutant DNMT1 and reduced methyltransferase activity. Mutant DNMT1 lost heterochromatin binding ability during the G2 cell cycle, leading to global hypomethylation measured by mass spectrometry analysis. Methylation of satellite 2 repetitive elements is preferentially decreased. Global hypomethylation and regional hypermethylation is shown in affected persons, a pattern commonly seen in unregulated cancer cells, suggesting post mitotic neural cells may undergo axonal degeneration by loss of cell cycle arrest. The discovered mutations in DNMT1 provide a new framework for the study of neurodegenerative diseases.

198

Factors playing a role in epigenetic mutations in Beckwith Wiedemann Syndrome. *V. Dagar[1,2], J. Mann[1,2], D. Amor[2,3], E. Algar[1,2].* 1) Murdoch Childrens Research Institute, Melbourne, Australia; 2) University of Melbourne, Australia; 3) Genetic Health Services Victoria, MCRI, Melbourne, Australia.

Beckwith Wiedemann Syndrome (BWS), a childhood cancer pre-disposition syndrome is an imprinting disorder characterised by overgrowth. It is one of the most common imprinting disorders with a prevalence of approximately 1 in 13700 live births, which is believed to be an underestimate due to the presence of milder, undiagnosed cases of BWS. The known causes of BWS are varied and encompass epigenetic mutations including loss of methylation (LOM) at KCNQ1OT1, gain of methylation at H19 and UPD of chromosome 11p15.5, CDKN1C mutations and a minority group with cytogenetic abnormalities affecting chromosome 11p15. The maintenance of methylation at imprinted centres post-fertilization is performed by DNMT1 which is the principal methyltransferase in vertebrates. We have hypothesised that as yet unidentified variations or mutations in the DNMT1 gene could result in failure of maintenance of methylation at the KCNQ1OT1 imprinting centre. To date we have identified 3 unique heterozygous substitutions in 3 cases of BWS with complete LOM at KCNQ1OT1. These are NM_001130823.1:[c.406 C>T]+[=], p.R136RC in exon 4, [c.3353A>G]+[=].p.H1118HR in exon 31 and [c.3668G>A]+[=],p.R1223RH in exon 33. All three variants have not been previously reported and were also not found in 100 controls. The variation identified in exon 31 is located just outside the Bromo Adjacent Homology domain II (amino acids 981-1117) where as the variation in exon 33 is located within the S-adenosylmethionine-dependant methyltransferases (SAM or AdoMet-MTase) class I domain (amino acids 1155-1384). As both these domains play a role in establishment of methylation it is hypothesised that they may be involved in loss of methylation in BWS and functional studies are in progress. We have also screened genes in the one-carbon folate pathway including MTHFR, MTR, MTRR, CBS and MAT1A in 50 cases of BWS with isolated LOM. We have identified an increased prevalence of the MTHFR 677C>T polymorphism in these cases. In the cohort the T allele frequency was 36% (36/100) compared to a control frequency of 26% (51/200) ($P<0.05$). Other SNPs in MTHFR (1298A>C, 1793G>A), MTR (2756A>G), MTRR (66A>G), CBS (450G>C, 1350C>T, 1351G>C, 1351G>A) and MAT1A (531C>A, 785G>A, 1046G>A, 1316A>G, 1321C>T) were equally represented in BWS and control cases. The evidence thus far suggests that variations in DNMT1 and polymorphisms in genes in the one-carbon pathway have a role to play in the loss of methylation in BWS cases.

199

5-hydroxymethylcytosine-mediated epigenetic dynamics during neurodevelopment and aging. *K. Szulwach[1], X. Li[1], Y. Li[1], CX. Song[2], H. Wu[3], Q. Dai[2], H. Irier[1], M. Gearing[4], AI. Levey[4], A. Vasanthakumar[5], LA. Godley[5], Q. Chang[6], C. He[2], P. Jin[1].* 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 2) Department of Chemistry and Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL; 3) Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA; 4) Center for Neurodegenerative Disease, Emory University School of Medicine, Atlanta, GA; 5) Section of Hematology/Oncology, Department of Medicine, The University of Chicago, Chicago, IL; 6) Department of Genetics and Neurology, Waisman Center, University of Wisconsin-Madison, Madison, WI.

DNA methylation dynamics influence brain function and are altered in neurological disorders. 5-hydroxymethylcytosine (5-hmC), a DNA base derived from 5-methylcytosine (5mC) accounts for ~40% of modified cytosine in brain, and has been implicated in DNA methylation-related plasticity. Here we utilize a novel strategy for 5-hmC specific chemical labeling coupled with affinity purification and high-throughput sequencing to map 5-hmC genome-wide in human and mouse brain as well as mouse models of the human neurodevelopmental disorder, Rett Syndrome, which is mainly caused by mutations in a gene encoding a 5-mCpG specific binding protein, *MECP2*. We find developmentally programmed acquisition of 5-hmC in neuronal cells. Epigenomic localization of 5-hmC-regulated regions across three ages in mouse hippocampus and cerebellum reveals the relative stability and dynamic regulation of distinct 5-hmC enriched loci during both neurodevelopment and aging. Profiling of 5-hmC in human cerebellum from multiple individuals establishes conserved genomic features of 5-hmC in mammalian brain, including enrichment within gene bodies, SINEs, and strong depletion on chromosome X. Additional epigenomic profiles of 5-hmC in mouse models of Rett Syndrome indicate that 5-hmC levels are sensitive to methyl-CpG-binding protein dosage, while intragenic and repeat-associated 5-hmC are enriched in mice without methyl-CpG-binding protein, Mecp2. These data establish comprehensive genome-wide maps of 5-hmC in mammalian brain and suggest previously unappreciated dynamic regulation of DNA methylation during both normal neurodevelopment and human neurodevelopmental disease.

200

Methylomic profiling across brain and blood: brain area-specific differentially methylated regions, individual differences, and allele-specific DNA methylation. *J. Mill, M. Davies, M. Volta, R. Dobson, E. Meaburn, L. Schalkwyk.* Institute of Psychiatry, King's College London, London, United Kingdom.

Epigenetic processes play an important role in the development and function of the central nervous system, and are being increasingly implicated in human psychiatric disease. Little is known about tissue-specific differences in DNA methylation across brain regions, nor how the brain methylome differs to that of easily accessible peripheral tissues such as blood. To create a detailed `methylomic atlas' of the brain, we performed methylated DNA immunoprecipitation followed by ultra-deep sequencing (MeDIP-seq) on tissue obtained from 8 brain regions and blood from multiple individuals. Between-tissue differences in DNA methylation greatly exceeds between-individual differences within any one tissue. Interestingly, brain area-specific differentially methylated regions (BA-DMRs) are strikingly under-represented in classic promoter CpG islands, being located primarily in intragenic and intergenic locations, and are dramatically enriched for genes known to be involved in brain development and neurobiological function. Samples obtained from cortex, cerebellum and blood can be discriminated with 100% accuracy based on DNA methylation across a core set of BA-DMRs. Of relevance to epigenetic studies of neuropsychiatric disease, much between-individual variation in DNA methylation is reflected across brain areas and blood. Finally, we employed a genome-wide analysis of allele-specific DNA methylation (ASM) across the same set of samples, observing considerable tissue-specificity in both cis- and parental origin-mediated ASM. These data have implications for our understanding about the role of DNA methylation in neurodevelopment and etiological studies of neuropsychiatric disease, and are freely available as a web-resource to the (epi)genetic research community.

201

Stochastic choice of allelic expression in neural progenitor cells. *A.R. Jeffries[1], L. Perfect[1], J. Mill[2], N.J. Bray[1], J. Ledderose[1], J. Price[1].* 1) Department of Neuroscience (CCBB), Institute of Psychiatry, King's College London, London, United Kingdom; 2) SGDP Research Centre, Institute of Psychiatry, King's College London, London, United Kingdom.

While most genes are expressed biallelically, a number of exceptions exist. Chromosome X inactivation and autosomal imprinted genes are two well known examples of monoallelic expression. A third category also exists where the choice of allele is made in a random or stochastic manner within individual cells early in development. This leads to clonal variation or cell self identity within a complex tissue. Initially thought to be limited to small number of gene families such as the olfactory receptors, immune receptors and protocadherins, more recent reports suggest a much greater number of autosomal genes may show stochastic allelic choice.

We used the Illumina Omni1-quad beadchip to carry out a genome wide allelic expression study on clonal human neural progenitor cells (100,000 informative SNPs/10,000 genes). Clones with the same DNA genotype were previously derived from three independent tissue donors: fetal cerebral cortex, striatum and spinal cord. A significant number of autosomal genes showed varying degrees of allelic imbalance. Monoallelic expression, which we defined as allelic ratios exceeding 4:1, was found in 108, 120 and 72 genes from the three donors, of which 45%, 44% and 22% of those identified showed stochastic allelic choice (where one clone shows monoallelic expression and a second clone is monoallelic for the alternate allele or displays biallelic expression). The cortical and striatal donors showed a significant number of proteins localised to the plasma membrane and extracellular region expressed in a stochastic allelic manner. These regions are important in processes such as cell-cell signalling and guidance cues within neurodevelopment. Stochastic allelic choice may therefore provide increased cellular diversity within tissue such as the developing nervous system.

In summary, this is the first study to carry out a global assessment of allelic expression in clonal neural progenitor cells. We identified a number of genes showing monoallelic expression, some of which show stochastic allelic choice. Many of these stochastic allelic choice genes are linked to processes and cell locations important in the developing nervous system.

202

Age effects on DNA methylation patterns in humans. *S. Horvath[1], Y. Zhang[1], K.R. van Eijk[2,3], E. Strengman[2,4], R.S. Kahn[3], M.P.M. Boks[3,4], R.A. Ophoff[1,3,4].* 1) Human Genetics, University of California, Los Angeles, Los Angeles, CA; 2) Medical Genetics, University Medical Center Utrecht, The Netherlands; 3) Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, The Netherlands; 4) Center for Neurobehavioral Genetics, Neuropsychiatric Institute, University of California, Los Angeles, USA.

Cytosine-5 methylation within CpG dinucleotides is an important mechanism of epigenetic control. Several studies have provided evidence that age has a significant effect on DNA methylation. In our primary analysis, we used blood samples from a large Dutch cohort to investigate the relationship of age on DNA methylation patterns . Methylation patterns were measured using the Illumina Infinium HumanMethylation27 Beadchip (Illumina 27k) microarray. Initially, we focused on 273 healthy control subjects since our main interest was to study “normal” aging effects but we also investigated aging effects in diseased individuals (293 schizophrenic samples). Since the relationship between age and DNA methylation may be different between the cases and the controls, we analyzed the two groups separately. We found high overlap between aging in healthy and diseased individuals. To protect against possible batch effects, we used a stratified analysis approach. A False Discovery Rate (FDR) threshold of 0.05 led to the identification of 1,716 age related probes in the control group and 1,210 probes in the diseased group. 865 age-related probes were shared among the 2 lists of genes. Among the 1,716 significant probes in the control group, DNA methylation of 763 probes was positively associated with subjects' age, while that of 953 probes was negatively associated with subjects' age. A gene ontology enrichment analysis found that genes whose methylation pattern decreased with age were significantly enriched for glycoproteins, cell adhesion molecules, and signaling molecules. A gene ontology enrichment analysis of the top 200 most positively age-related probes and the top 200 most negatively age-related probes revealed shared terms such as signal, signal peptide, and glycoprotein. Stratifying the analysis by sex led to the identification of sex-specific methylation changes with age. We also carried out a weighted correlation network analysis (WGCNA) to identify age related methylation modules. In a secondary analysis, we cross-referenced our findings in blood to those in different brain regions using a publicly available brain methylation data set. We found that 158 of the age related probes found in blood also showed an age effect in the PONS frontal cortex and temporal cortex. However, cerebellum showed remarkably distinct aging patterns which suggests that age related methylation changes cannot be consistently identified in all tissues.

203

Analysis of Noncoding RNA expression in Late-Onset Alzheimer's disease (LOAD). *C.E. Humphries[1,2], M.A. Kohli[2], P.L. Whitehead[2], D.C. Mash[2], M.A. Pericak-Vance[2], J.R Gilbert[2].* 1) Department of Human Genetics and Genomics, University of Miami, Miami, FL; 2) Hussman Inst for Human Genomics, University of Miami, MIAMI, FL.

Whole transcriptome pathway analysis (RNA-Seq) using next-generation sequencing (NGS) permits quantization of gene transcription and identification of novel splicing events, coding RNAs, and non-coding RNAs (ncRNA) at a depth of transcription and level of complexity previously unknown. ncRNAs are ubiquitous and involved in many aspects of gene transcription and regulation and thus are likely to play a significant role in AD pathology. Numerous studies have looked at the transcription of coding genes, but few published studies have looked at ncRNA expression in AD. To investigate the molecular networks and ncRNAs transcription in LOAD, RNA-Seq was performed using RNA isolated from brain temporal cortex (BA 38) derived from LOAD, Lewy-Body with Dementia (LBD; non-LOAD dementia), and cognitively normal patients (CN). Samples were matched for age (71-86), sex, race (white) and, for LOAD patients, Braak stage (IV). Sequence libraries were prepared from extracted total RNA using the RiboMinus Eukaryote kit for RNA-seq followed by SOLID Whole Transcriptome preparation. Each sample was run on one quad spot on the SOLiD4 sequencer. TopHat aligned an average of 65 million reads per sample to the reference genome. Using Cufflinks, initial data show an average of 444,298 unique transcripts found in each sample with ~17,200 transcripts aligning to known genes. 42,579 novel isoforms of annotated reference sequence genes were identified including isoforms in 18 genes previously associated with AD (e.g. BIN1, PICALM, PSEN1, and APP). Analyses of noncoding transcripts show potential ncRNAs are found in equal amounts in both CN controls and LOAD cases. Regions involved in ncRNA transcription in LOAD cases and LBD disease controls differ significantly from CN controls. Using Webgestalt, transcribed regions in AD and LBD cases are shown to aggregate preferentially near genes involved in AD pathology; chemokine signaling (p=1.37E-5), ATP Binding Cassette (ABC) transporters (p=5.75E-5), and taste transduction ($p<0.05$), as well as binding sites for SP1, a transcription factor known to be deregulated in AD is found (p=5e-14) near these regions, such as ABC transporters. Noncoding transcription variation is observed specifically near genes shown to be associated with AD (i.e.-PICALM, CLU, MTHFD1L, CD2AP, and APP). These findings suggest widespread uncharacterized ncRNA transcriptional deregulation occurring in AD patients that likely plays an important role in disease etiology.

204

Components of Heritability in an Icelandic Cohort. *N. Zaitlen[1,2,3], A. Helgason[4], N. Patterson[3], B. Pasaniuc[1,2,3], D. Gudbjartsson[1,2,3], P. Kraft[4], A. Kong[4], A.L. Price[1,2,3], K. Stefansson[4].* 1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) Department of Biostatistics, Harvard School of Public Health, Boston, MA; 3) Broad Institute, Cambridge, MA; 4) deCODE genetics, Reykjavík, Iceland.

The combined genotypes and genealogy of 38,167 Icelanders provides the opportunity to examine the contribution of different components of genetic variation to the heritability of complex phenotypes. In this work we focus on both parent-of-origin effects, and the contribution of typed versus untyped variants. Genetic variation can alter phenotype in a parent-of-origin specific manner, for example, via imprinting. Kong et al [1] identified three Type 2 Diabetes (T2D) variants including rs2334499, which is protective for T2D when inherited maternally, confers risk when inherited paternally, and lies in an imprinted region of the genome. The genealogy provides a means of resolving not only highly accurate phasing, but also the parent-of-origin of each typed variant. Using this information we develop methods to examine the total contribution of parent-of origin effects, as well as the differences between paternal and maternal contributions to the heritability of complex phenotypes. We find that variation of height shows little evidence of contribution from parent-of-origin effects, while T2D shows highly significant evidence (p-value < 1.9×10^{-5}), suggesting that many more variants such as rs2334499 remain to be found. In addition to parent-of-origin effects, the unique long range phasing the deCode data provide a means of efficiently estimating the fraction of the genome shared identical by descent (IBD) as well as identical by state (IBS). As has been recently demonstrated [2], IBS sharing can be used to estimate the contribution to phenotype of genotyped SNPs and SNPs in linkage disequilibrium (LD) to genotyped SNPs. IBD sharing estimates the total narrow sense heritability of the phenotype, that is, the additive contribution of all SNPs. Since both are readily available in these data we are able to compare the IBD based estimates to the IBS based estimates and show that for all phenotypes examined the majority of heritability is well captured by the genotyped SNPs. We conclude with a discussion of confounding effects in mixed model estimates of heritability such as those provided by GCTA, showing that population structure can lead to biased estimates of heritability even when corrected for by principal component adjustments. 1 Kong et al. Parental origin of sequence variants associated with complex diseases. Nature 2009. 2 Yang el al. Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet 2011.

205

Gene-phenotype relationships in a deep-resequencing study of 202 genes in >14,000 subjects including 12 complex diseases. *M.G. Ehm[1], M.R. Nelson[1], L. Warren[1], L. Li[1], P.L. St. Jean[1], J. Shen[1], D. Fraser[1], J.L. Aponte[1], J. Rubio[1], S. Zöllner[3], L.R. Cardon[1], J. Novembre[2], J.C. Whittaker[1], S.L. Chissoe[1], V.E. Mooser[1].* 1) Genetics, GlaxoSmithKline, Res Triangle Park, NC; 2) Biostatistics, University of Michigan, Ann Arbor, MI; 3) Ecology & Evolutionary Biology, University of California-Los Angeles.

We sequenced the exons of 202 genes encoding known or potential drug targets in more than 14,000 individuals for the study of 12 diseases. We found 38,535 variants: 95% with Minor Allele Frequency (MAF) of (0,0.005], 2% with (0.005, 0.05] and 3% with (0.05, 0.5]. Genotype-phenotype analyses of disease phenotypes included 300 to 1700 cases and 1000 to 7000 genetically-matched controls.

Within the common variants, 14% of them were coding. These variants were analyzed using single marker tests. Using at least 3,348 subjects from our sequencing study with GWAS data, we were able to impute 75% of variants with MAF>0.005 including 49% of variants with MAFs of (0.005, 0.05]. Therefore, we were able to increase our sample size by 30 to 100% by imputing variants in additional cases and population controls with GWAS data. This highlights the potential for sequence data from large reference panels such as this to be coupled with GWAS consortium resources. This will enable the discovery of disease associations for variants in the (0.005, 0.05] range for which individual sequence studies are currently underpowered.

We observe an excess of nonsynonymous over sysnonymous variants for rare variants in constrast to common variants. The ratio of damaging to benign variants, as predicted by SIFT and PolyPhen, increases substantially as the MAF decreases and they are located at highly conserved sites more often than for common variants. Therefore, we analyzed all functionally damaging or conserved rare variants with MAF<0.005 using aggregation tests. The power of these tests is dependent on the aggregate MAF observed which is highly correlated with the coding sequence length. Despite this challenge, the observation of rare functionally damaging variants in individuals with outlying phenotypic values leads to extreme p-values (an association of *PSEN1* and age of onset for Alzheimer's Disease, $p<10^{-9}$, is an example confirmed in these data). In summary, this study provides an early view into the functionality of variation uncovered and the nature of gene-phenotype associations (rare and common) identified in deep re-sequencing studies for common diseases.

206

Multivariate genetic analysis of phenotypic sets identifies 10 novel loci for systemic metabolism. *M. Inouye[1], P.I.W. de Bakker[2], J. Kettunen[3], A.J. Kangas[4], P. Soininen[4,5], M.J. Savolainen[4], J. Viikari[7], M. Kahonen[8], T. Lehtimaki[6], S. Ripatti[3], O. Raitakari[11], M.R. Jarvelin[9], A. Palotie[10], M. Ala-Korpela[4,5].* 1) Walter & Eliza Hall Institute of Medical Research, Melbourne, Victoria, Australia; 2) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 3) Institute of Molecular Medicine (FIMM), University of Helsinki, Finland; 4) Computational Medicine Research Group, Institute of Clinical Medicine, University of Oulu and Biocenter Oulu, Oulu, Finland; 5) NMR Metabonomics Laboratory, Laboratory of Chemistry, Department of Biosciences, University of Eastern Finland, Kuopio, Finland; 6) Department of Clinical Chemistry, University of Tampere and Tampere University Hospital, Tampere 33521, Finland; 7) Department of Medicine, University of Turku and Turku University Hospital, Turku 20521, Finland; 8) Department of Clinical Physiology, University of Tampere and Tampere University Hospital, Tampere 33521, Finland; 9) Department of Epidemiology & Biostatistics, Imperial College London, St. Mary's Campus, Norfolk Place, London W2 1PG, United Kingdom; 10) Department of Human Genetics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom; 11) Research Centre of Applied and Preventive Cardiovascular Medicine, Department of Clinical Physiology, University of Turku and Turku University Hospital, Turku 20521, Finland.

Recent advances in our knowledge of genomic variation and susceptibility to disease have shown a need for more comprehensive understanding of phenotypic variation. Technological advancements now allow for profiling of individuals and individual tissues in much greater detail. For genetic association studies, the structure of phenotypic variation has received less attention than that of genomic variation. Complex phenotypes like human height are agglomerations of multiple sub-phenotypes (e.g. bone lengths), and many complex phenotypes are themselves correlated. Therefore, a proportion of the variance of a phenotype is attributable to the levels of other related phenotypes. In terms of statistical power the simultaneous treatment of a multiplicity of phenotypes offers potential advantages over univariate analyses. To investigate this empirically, we assessed 7,268 individuals from two population-based cohorts (Cardiovascular Risk in Young Finns and Northern Finland Birth Cohort 1966) with both genome-wide genotype data and serum metabolomic profiles. The high dimensionality of systemic metabolism was reduced by identifying 11 correlated clusters of biologically relevant metabolite sets. Multivariate testing of these metabolite sets uncovered 33 genomic loci overall reaching genome-wide significance. Of these, 10 were novel loci, harbouring candidate genes SERPINA1 (P= 5x10-48), AQP9 (P=3x10-27), GC (P=1x10-14), EREG (6x10-14), and GLTPD2/TM4SF5 (P=1x10-11). In comparison to univariate testing, the multivariate approach identified nearly twice as many genome-wide significant genetic associations. Taken together, these findings strongly support the greater power of multivariate testing and its utility as an alternative to univariate testing in genome-wide screening of correlated quantitative phenotypes.

207

Multi-ethnic fine-mapping reveals rare and potential causal variants for a complex disease. *R.C. Almeida[1,2], G. Trynka[1], J. Gutierrez-Achury[1], J. Romanos[1], J.R. Bilbao[3], D. Barisani[4], L. Greco[5], M.C. Mazilli[6], A. Sood[7], B. Cukrowska[8], K. Hunt[10], E. Urcelay[9], D. Van Heel[10], C. Wijmenga[1], Wellcome Trust Case Control Consortium.* 1) Department of Genetics, University Medical Centre Groningen and University of Groningen, Groningen, Netherlands; 2) University of Brasilia School of Health Sciences, Brasilia, DF,Brazil; 3) Immunogenetics Research Laboratory, Hospital de Cruces, Barakaldo 48903 Bizkaia, Spain; 4) Department of Experimental Medicine, Faculty of Medicine University of Milano-Bicocca,Monza, Italy; 5) European Laboratory for Food Induced Disease, University of Naples Federico II, Naples, Italy; 6) Department of Molecular Medicine, Sapienza University of Rome, Rome, Italy; 7) Dayanand Medical College and Hospital, Ludhiana, Punjab, India; 8) Department of Pathology, Children's Memorial Health Institute, Warsaw, Poland; 9) Depatment of Immunology, H. Clínico S. Carlos, Instituto de Investigación Sanitaria San Carlos (IdISSC), Madrid, Spain; 10) Blizard Institute of Cell and Molecular Science, Barts and The London School of Medicine and Dentistry, London E1 2AT, United Kingdom.

Genome Wide Association Studies (GWAS) have identified thousands of new susceptibility loci for common diseases suggesting that susceptibility to these traits is attributable to variants present in more than 5% of the population. Nevertheless, most of these common variants have a modest effect on risk and do not fully explain the heritability of complex diseases. The current genotype arrays do not include low frequency and rare variants, which may confer higher risk to disease and more likely to be the true causative one. We used data from the Immunochip, a custom-made array, which contains _50,000 rare SNPs from 183 distinct loci associated to 12 immune-mediated diseases. We genotyped this chip in 7 different populations comprising 24,262 individuals. We performed in-depth analysis of the *LPP* locus that is strongly associated to celiac disease. In a meta-analysis across all populations we observed strong association to a common SNP (MAF=0.4, p $3x10^{-49}$). Upon conditioning on this SNP the association within this locus disappeared, indicating association to a single common haplotype. We then investigated each population separately. Only the UK population, which had the largest number of samples (16,002 individuals), showed the same top meta-analysis SNP, while all other populations showed association to a different SNP; two populations (Poland and India) were not associated. We further investigated low frequency variants and found two variants (MAF= 0.02) with a large effect (OR >2) in three populations (p-value per population < $9x10^{-3}$). Haplotype analysis showed that these low frequency variants had accumulated on the most frequent risk haplotype. None of the associated SNPs mapped to the coding sequence of LPP. To investigate the presence of functional regulatory elements in this region, we used ENCODE and the UCSC Genome Browser. We investigated all population-specific top signals together with all SNPs in high LD (r2>0.8) with those and showed that four of them fall into transcription factor binding sites. We suggest that these are the most likely causal variants and will be studied in a functional manner. We conclude that disease associated regions capture low frequency variants with high effect sizes, however these will only be detected when analyzing dense genotype data across different populations. We suggest that dysregulation of transcription factor binding properties might be a causal mechanism underlying the association at *LPP* region.

208

Custom array genotyping identifies new loci for type-2 diabetes, providing insight into the biological and genetic architecture of disease.

B.F. Voight[1], A.P. Morris[2], T.M. Teslovich[3], T. Ferreira[2], E.J. Rossin[1,4,5,6,7], R.J. Strawbridge[8], A.V. Segrè[1,4], A. Mahajan[2], I. Prokopenko[2,9], H. Grallert[10], M. Boehnke[3], M.I. McCarthy[2,9] on behalf of the DIAGRAM Consortium. 1) Medical Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 3) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 4) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 5) Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, MA, USA; 6) Health Science and Technology MD Program, Harvard University and Massachusetts Institute of Technology, Boston, MA, USA; 7) Harvard Biological and Biomedical Sciences Program, Harvard University, Boston, MA, USA; 8) Cardiovascular Genetics and Genomics Group, Karolinska Institutet, Sweden; 9) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK; 10) Research Unit Molecular Epidemiology, Helmholtz Zentrum Muenchen, Neuherberg, Germany.

With over 40 loci contributing susceptibility to type-2 diabetes (T2D), an important next step is to translate knowledge of these loci and variants into insights regarding the genetic and biological architecture of disease. Toward this aim, the DIAGRAM consortium used a custom array, the *Metabochip*, designed specifically for replication and fine-mapping studies of metabolic traits. We undertook a two-stage design, evaluating 12,057 T2D cases and 56,071 controls in stage 1 using standard GWAS arrays, and stage 2 consisting of 24,204 cases and 60,858 controls genotyped with the *Metabochip*, with 95% of these samples from European descent (5% in stage 2 were of Pakistani origin). In addition to reproducing associations at 42 established loci, 14 new loci achieved genome-wide significance ($P < 5 \times 10^{-8}$) in combined meta- or sex-stratified analysis; the closest genes for these novel signals were *ZMIZ1*, *ANK1*, *HMG20A*, *TLE1*, *SF4*, *ANKRD55*, *CTRB1*, *KLHDC5*, *TMEM154*, *CCND2*, *GIPR*, *SSR1*, *MC4R*, and *GRB14*. Four additional loci - near *GLIS3*, *POU5F1*, *PLKHA1*, and *ETS1* - were borderline significant ($P < 5 \times 10^{-7}$).

The depth of replication allowed us to test hypotheses related to the genetic and biological architecture of disease. Statistical analysis of 3,435 LD-independent sites targeted for T2D replication revealed an excess of directionally consistent effects between stage 1 and stage 2 (63.6% observed, 50% expected, $P < 8 \times 10^{-36}$). The observed data are consistent with ~14% of the 3,435 sites influencing T2D susceptibility, supporting a model of an extensive polygenic component. We tested the hypothesis that genes mapping to T2D associated loci were more likely to physically interact with each other than expected by chance. Using a high confidence protein-protein interaction map (Rossin et al, 2011), a significant enrichment of direct and indirect connection among genes in these associated regions was identified ($P < 0.0001$). Examination of this network revealed (a) genes important for anti-diabetic drug therapy (e.g., *CREBBP*), (b) mouse knockouts of genes in the network demonstrating metabolic and beta-cell dysfunction, (e.g., *NCR2C2*, *NCOR2*, *FGFR1*), (c) genes important for adipocyte function (e.g., *ESR1*, *PRKCZ*), and (d) highly connected genes with no previously established relationship to T2D. Our analysis of these data show promise for improving our understanding of the biological underpinnings of disease.

209

Search for novel susceptibility loci for type 2 diabetes using genome-wide association studies imputed from a 1000 Genomes reference panel.

I. Prokopenko[1,2], H. Chen[3], C. Ma[4], R. Mägi[1], B.F. Voight[5,6,7], Q. Lu[8,9,10], N. Van Zuydam[11,12], H. Grallert[13], L. Yengo[14,15,16], C. Dina[17,18,19], G. Thorleifsson[20], C. Fucshberger[4], L. Liang[10,21], M. Mueller[22], S.M. Willems[23], K.J. Gaulton[1], V. Steinthorsdottir[20], C.M. Van Duijn[23], M. Boehnke[4], J. Dupuis[3,24], M.I. McCarthy[1,2,25], L.J. Scott[4] on behalf of DIAGRAM+ consortium. 1) University of Oxford, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford OX3 7LJ, UK; 3) Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, MA 02118, USA; 4) Center for Statistical Genetics and Dept of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, USA; 5) Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts 02142, USA; 6) Center for Human Genetic Research, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, USA; 7) Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA; 8) Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA; 9) Channing Laboratory, Department of Medicine, Brigham and Womens Hospital and Harvard Medical School, Boston, Massachusetts, USA; 10) Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA; 11) Diabetes Research Centre, Biomedical Research Institute, University of Dundee, Ninewells Hospital, Dundee DD1 9SY, UK; 12) Pharmacogenomics Centre, Biomedical Research Institute, University of Dundee, Ninewells Hospital, Dundee DD1 9SY, UK; 13) Research Unit Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 14) Centre National de la Recherche Scientifique (CNRS)-Unité mixte de recherche (UMR) 8199, Lille, France; 15) Pasteur Institute, F-59019 Lille, France; 16) Nord de France University, Lille, France; 17) INSERM, UMR915, L'Institut du Thorax, Nantes, France; 18) Université de Nantes, Nantes, France; 19) Centre National de la Recherche Scientifique (CNRS), ERL3147, Nantes, France; 20) deCODE genetics, Reykjavik, Iceland; 21) Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA; 22) Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 23) Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands; 24) National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts 01702, USA; 25) Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford OX3 7LJ, UK.

To date 43 common variant associations have been described for type 2 diabetes (T2D); however these account for <10% of inherited susceptibility. Almost all maximally-associated T2D variants have minor allele frequency (MAF) >10%. Our goal is to identify causal variants at known T2D loci and to uncover low allele frequency risk variants at novel sites by performing imputation from successively larger and more complete reference panels generated by genome resequencing. To that end, we imputed SNP variants using the 1000 Genomes European reference panel (566 CEU haplotypes, 12/2010 release) within each of 7 cohorts of European descent with genome-wide association (GWA) data. Following T2D association analysis, we performed a fixed-effects inverse variance meta-analysis for 10,234 individuals with T2D and 47,066 controls. 7.7M SNPs were present in at least 4 studies, passed genotype/imputation quality control, and did not show significant between-study heterogeneity. 2.3M of these SNPs had MAF<0.05. Imputation using the 1000 Genomes reference increased the number of SNPs tested from the previous DIAGRAM+ analysis by >3 fold. We observed genome-wide significant (p-value<5×10^{-8}) associations at 7 and suggestive (p-value<1×10^{-5}) associations at 16 known T2D loci. Within 3 of these loci (*TCF7L2*, *CDKN2A/2B* and *CDKAL1*), we identified common variants (rs35198068, rs7451008, rs10965246, respectively) which on imputation demonstrated more significant association than previously detected variants. In *TCF7L2*, we also observed association (OR=1.5[1.3-1.8], p-value=1×10^{-9}) at a low frequency (rs116369954; MAF=0.03) variant, 35kb from the known signal (p-value=1×10^{-42}): conditional tests of independence are not yet available. None of the lower frequency variants outside of known T2D loci was significant at 5×10^{-8}. We observed suggestive association at two novel common variant loci near *MSRA* (OR=1.11[1.07-1.16], p-value=6×10^{-7}), not well-tagged in previous imputations, and within HLA-DRB1 (OR=1.12[1.06-1.15], p-value=4×10^{-7}). The latter may well reflect subtle population stratification or misclassification involving cases with autoimmune diabetes. Imputation with the current 1000 Genomes reference panel has not revealed novel association signals driven by low-MAF variants. However, imputation for many low-MAF variants remains suboptimal given the size of the current reference panel and the density of much existing GWAs data.

210

Exome sequencing to identify genes harboring rare variants that determine lung function decline in COPD. *R.A. Mathias[1], Y. Kim[2], M. Taub[1], L. Huang[1], N. Rafaels[1], T. Murray[1], C. Vergara[1], M. Bamshad[3], M. Emond[3], T.H. Beaty[1], I. Ruczinski[1], N. Hansel[1], K.C. Barnes[1] on behalf of the NHLBI's Exome Sequencing Program and Lung Project Team.* 1) Johns Hopkins University, Baltimore, MD; 2) NHGRI, NIH, Baltimore, MD; 3) University of Washington, Seattle, WA.

Rationale: A recent genome-wide association study (GWAS) has revealed several loci determining rate of lung function decline in European American individuals with COPD from the NHLBI-supported Lung Health Study (LHS). Here, we use an exome-sequencing approach to identify additional rare coding variants not adequately tagged by the genomewide marker array. Methods: A GWAS was conducted using the Illumina 660W chip on 4,102 subjects with a minimum of 3 time-points where lung function was measured over a 5-year period. As part of the NHLBI-supported Exome Sequencing Program, a subset of these samples with extreme quartiles of lung function decline have been exome sequenced using Nimblegen capture arrays on the Illumina platform. Sequences were aligned using BWA, variants called using the Genome Analysis Toolkit, and annotated using Seattle-Seq. Standard sequencing quality control filters were applied. A case-control analysis was performed restricting several alternative burden tests to functional (nonsynonymous and nonsense) variants comparing the rapid decliners (N=130) and the slow decliners (N=119). Results: 161,774 variants were noted, of which 80,059 (49%) were novel. 25,743 annotated transcripts had at least one functional coding variant. Integration of the GWAS and sequence data revealed >99% concordance over ~7,500 variants sites. The four most significant genes for the CMC method collapsing variants with a minor allele frequency <5% included *CTH*, *GOLGA4* and *DCAF5* with protective effects and *HK2* with risk effects for lung function decline. The strongest signals under the Madsen-Browning weighted sum test (weighting by frequency in the slow-decliner group) included *ANKRD6, TRIM47, HLA-B* and again *HK2*. The number of variants for each test ranged from 3 (*CTH*) to 46 (*HLA-B*). We examined capture rates for these rare variants by the GWAS platform with a tagging approach. Conclusions: We identified several loci not detected in our GWAS that appear to harbor rare variants associated with lung function decline in COPD. While sampling only extreme subjects for sequencing optimizes statistical power to detect rare variants controlling lung function decline, we show that our genomewide marker panel did a poor job of tagging rare variants and an integrative approach of a GWAS in a larger sample (4,102) with the exome sequence in a smaller, highly selected sample (N=251) can be important in identifying genetic control of complex quantitative phenotypes.

211

Two new genes for dominant and recessive pseudohypoaldosteronism type II provide novel insights into regulation of serum potassium, pH, and blood pressure. *L.M. Boyden, K.A. Choate, M. Choi, C.J. Nelson-Williams, A. Farhi, R.P. Lifton.* Yale University School of Medicine, New Haven, CT.

Pseudohypoaldosteronism type II (PHAII) is a Mendelian disorder characterized by impaired potassium and acid excretion, increased renal salt reabsorption, and hypertension. These abnormalities are ameliorated by thiazide diuretics, which inhibit salt reabsorption by the Na-Cl cotransporter (NCCT) in the distal nephron of the kidney. Previous work demonstrated that mutations in *WNK1* and *WNK4*, genes encoding kinases that regulate NCCT and other mediators of ion flux in the distal nephron, cause dominant PHAII; however, mutations in these genes explained only a small fraction of the reported PHAII cases. Via whole-exome sequencing of 11 PHAII index cases, we have now discovered that mutations in a third gene cause both dominant and a novel recessive form of PHAII. Resequencing of this gene in kindreds and our remaining PHAII cohort revealed mutations cosegregating with disease in 8 recessive and 16 dominant PHAII kindreds. Recessive mutations are loss-of-function (including nonsense, frameshift, and canonical splice site mutations) and dominant mutations are missense and presumed dominant-negative; most of the dominant mutations occur in specific regions of a domain implicated in intermolecular interactions by crystal structure studies of related proteins. Via a candidate gene approach, we have also discovered that mutations in a fourth gene cause dominant disease in an additional 17 kindreds. All of the mutations in this gene are in and around a single exon, and occur in canonical splice sites, the splice branch point, and an exonic splice enhancer. We show that these mutations lead to production of a unique transcript with an in-frame deletion. These two new genes account for the majority of disease in our PHAII cohort. We demonstate a spectrum of PHAII disease severity, with correlations of phenotype and genotype in affected patients. Via immunofluorescence we show that the two new genes, like *WNK1* and *WNK4*, are expressed in the distal nephron of the kidney. Consideration of all four genes now implicated in PHAII lends new insights into a model whereby a molecular switch toggles the kidney's reaction to states of serum potassium excess and volume depletion, affecting electrolyte homeostasis and blood pressure.

212

Identification of a susceptibility gene for Moyamoya disease, *RNF213* by a genome-wide association study. *S. Kure[1,2], F. Kamada[1,2], Y. Aoki[2], Y. Abe[1,2], S. Komatsuzaki[2], A. Kikuchi[1], J. Kanno[1,2], T. Niihori[2], M. Fujimura[3], Y. Mashimo[4], M. Ono[5], N. Ishii[6], Y. Owada[7], Y. Suzuki[4], A. Hata[4], T. Tominaga[3], Y. Matsubara[2].* 1) Dept Pediatrics, Tohoku Univ Sch Med, Sendai, Japan; 2) Dept Medical Genetics, Tohoku Univ Sch Med, Sendai, Japan; 3) Dept Neurosurgery, Tohoku Univ Sch Med, Sendai, Japan; 4) Dept Public Health, Graduate Sch Med, Chiba Univ, Chiba, Japan; 5) Dept Pathology, Tohoku Univ Sch Med, Sendai, Japan; 6) Dept Immunology, Tohoku Univ Sch Med, Sendai, Japan; 7) Dept Organ Anatomy, Yamaguchi Univ Sch Med, Sendai, Japan.

Moyamoya disease (MMD) shows progressive cerebral angiopathy characterized by bilateral internal carotid artery stenosis and abnormal collateral vessels. Although 10-15% of MMD cases are familial, the MMD gene(s) remains unknown. A genome-wide association study of 785,720 single nucleotide polymorphisms (SNPs) was performed, comparing 72 Japanese MMD patients to 45 Japanese controls and resulting in a strong association of chromosome 17q25-ter with MMD risk. This result was further confirmed by a locus-specific association study using 335 SNPs in the 17q25-ter region. A single haplotype consisting of seven SNPs at the *RNF213* locus was tightly associated with MMD ($P=5.3x10^{-10}$). *RNF213* encodes a RING finger protein with an AAA ATPase domain and is abundantly expressed in spleen and leukocytes. An RNA in situ hybridization analysis of mouse tissues indicated that mature lymphocytes express higher levels of *Rnf213* mRNA than their immature counterparts . Mutational analysis of *RNF213* revealed a founder mutation, p.R4859K, in 95% of MMD families, 73% of non-familial MMD cases and 1.4% of controls; this mutation greatly increases the risk of MMD ($P=1.2x10^{-43}$, odds ratio=190.8, 95% confidence interval=71.7 to 507.9). Three additional missense mutations were identified in the p.R4859K-negative patients. These results indicate that *RNF213* is the first identified susceptibility gene for MMD.

213

***FBN1* susceptibility loci associated with thoracic aortic aneurysm and aortic dissection disease spectrum.** *M.N. McDonald[1], S.A. Lemaire[2,4], D. Guo[15], L. Russell[2,4], C.C Miller, III[16], R.J. Johnson[15], M.R. Bekheirnia[1], L.M. Franco[1], M. Nguyen[2,4], R.E. Pyeritz[6], J.E. Bavaria[7], R. Devereux[8], C. Maslen[9], K.W. Holmes[10], K. Eagle[11], S.C. Body[12], C. Seidman[13], J.G. Seidman[13], E.M. Isselbacher[14], M.S. Bray[3], J.S. Coselli[2,4], A.L. Estrera[16], H.J. Safi[16], J.W. Belmont[1], S.M. Leal[1], D.M. Milewicz[5,15].* 1) Department of Human and Molecular Genetics, Baylor College Med, Houston, TX; 2) Division of Cardiothoracic Surgery, Michael E. DeBakey Department of Surgery, Baylor College of Medicine, Houston, TX; 3) Department of Pediatrics, Baylor College Med, Houston, TX; 4) Cardiovascular Surgery, St. Luke's Episcopal Hospital, Houston, Texas; 5) Medicine Services of the Texas Heart Institute at St. Luke's Episcopal Hospital, Houston, Texas; 6) Division of Medical Genetics, Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania; 7) Division of Cardiovascular Surgery, Department of Surgery, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania; 8) Greenberg Division of Cardiology, Department of Medicine, Weill Cornell Medical College, New York, New York; 9) Division of Cardiovascular Medicine, Department of Molecular & Medical Genetics, Oregon Health & Science University, Portland, Oregon; 10) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland; 11) Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan; 12) Department of Anesthesiology, Harvard Medical School, Boston, Massachusetts; 13) Department of Genetics, Harvard Medical School, Boston, Massachusetts; 14) Department of Medicine, Harvard Medical School, Boston, Massachusetts; 15) Department of Internal Medicine, University of Texas Health Science Center at Houston, Houston, Texas; 16) Department of Cardiothoracic and Vascular Surgery, University of Texas Health Science Center at Houston, Houston, Texas.

Thoracic aortic aneurysm and dissection (TAAD) is a common cause of premature death, having ranked as high as the 15th leading cause of death in the United States. Ascending aortic aneurysms asymptomatically and progressively enlarge over time and ultimately lead to life-threatening acute aortic dissection or rupture. Acute ascending aortic dissection is termed type A dissection and descending aortic dissection is termed type B dissection. A subset of TAAD patients have bicuspid aortic valve (BAV), the most common congenital heart defect, found in 1-2% of the population. There is indication of a strong association between patients with BAV and risk for TAAD. Among patients referred for surgical treatment of BAV, 20% have concurrent ascending aortic aneurysms and approximately 15% of patients with acute aortic dissections have BAV. All of these conditions are part of the TAAD disease spectrum. In Stage 1 of a multistage genome-wide association study (GWAS), we compared 765 patients who had sporadic TAAD (STAAD) without Marfan syndrome (MFS) with 874 controls and identified common SNPs at a 15q21.1 locus that were associated with STAAD with odds ratios (ORs) of 1.6-1.8 that achieved genome-wide significance (GWS). The associated SNPs fall into a large region of linkage disequilibrium (LD) encompassing *FBN1*, which encodes fibrillin-1. Mutations in this gene cause MFS, whose major cardiovascular complication is TAAD. We followed up 107 SNPs associated with STAAD with $P < 1 \times 10^{-5}$ in the region in two separate STAAD cohorts. We further investigated this region in STAAD subphenotype groups, specifically STAAD with and without BAV, aneurysm, Type A and Type B dissection, to determine if the associations were present within STAAD subphenotype presentations. The meta-analysis of the three STAAD cohorts demonstrated that *FBN1* region SNPs achieved GWS for STAAD with and without BAV, with aneurysm and with dissection, specifically type A, in addition to the STAAD overall finding. The most highly associated meta-analysis finding was with Type A dissection (rs9806323, OR:2.4 and $P=4.9x10^{-13}$). Our findings demonstrate the associations are present with all investigated STAAD subphenotypes. This study shows that common genetic variants (minor allele frequency range: 0.08-0.36) in the *FBN1* gene region, distinct from the missense and nonsense mutations that cause MFS, are associated with the STAAD disease spectrum.

214

A major modifier locus for vascular disease in Marfan syndrome. *A. Doyle[1,2,3], K. Kent[1,2], H. Dietz[1,2].* 1) Institute of Genetic Medicine, Johns Hopkins Medical Institute, Baltimore, MD; 2) Howard Hughes Medical Institute, Baltimore, MD; 3) The William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, London, UK.

Marfan syndrome (MFS) is a connective tissue disorder caused by mutations in the FBN1 gene. The major cause of mortality for patients with MFS is aortic aneurysm, dissection and rupture. There is a growing body of evidence that promiscuous activation of TGF) is responsible for multiple manifestations of MFS including aortic aneurysm and that blocking the angiotensin II type 1 receptor (AT1R) can ameliorate the disease. While MFS is known to show clinical variability, this tends to be greatest between families and has been attributed to allelic heterogeneity. Complete penetrance has been suggested for this condition. We recruited 5 exceptional MFS families with discrete intrafamilial variation in the severity of vascular disease among age- and gender-matched FBN1 mutation carriers; complete nonpenetrance was also documented. Individuals were assigned an unprotected status (n=18) with an aortic root Z-score>3.0 or a history of aortic root dissection/surgery and protected status (n=20) with an aortic root Z-score<2.0 or an age>60 without prior dissection/surgery; those with a Z-score of 2.0-3.0 (n=7) were considered indeterminate. In total, 51 individuals were genotyped (including 6 informative but unaffected relatives) and analyses were carried out using 11936 autosomal and 278 sex chromosome SNPs. Non-parametric linkage analysis (NPL) was performed with an exponential model. Parametric linkage analysis (PL) was performed following a maximized maximum LOD score procedure. Both NPL and PL focused on protection from vascular disease in FBN1 mutation carriers. The results indicated a single region of linkage on chromosome 6 (LOD=3.0 and 4.0 respectively) present in all 5 families. While the protective haplotype varied between families, all patients with mild disease (20/20) shared a 3.9Mb familial haplotype between markers rs676017 and rs6455736, while only 2/18 patients assigned unprotected status harboured the familial protective haplotype (p<0.0000001). We are currently evaluating 3 strong candidate genes in this region that encode a functional antagonist of the AT1R (MAS1), the insulin-like growth factor 2 receptor (IGF2R) that contributes to TGF) activation and mitogen-activated protein kinase kinase kinase 4 (MAP3K4), a downstream effector of noncanonical TGF) signalling. In summary, we have identified a major modifier locus for vascular disease in MFS. A refined understanding of the mechanism of protection should reveal novel therapeutic strategies.

215

Exome sequencing identifies SMAD3 mutations causing familial thoracic aortic aneurysm and dissection with intracranial and other arterial aneurysms. *E.S. Regalado[1], L. Gong[1], C. Villamizar[1], D.C. Guo[1], N. Avidan[1], D. Gilchrist[2], B. McGillivray[3], L. Clarke[3], F. Bernier[4], R.L. Santos-Cortez[5], S.M. Leal[5], A.M. Bertoli-Avella[6], J. Shendure[7], M.J. Rieder[7], D.A. Nickerson[7], D.M. Milewicz[1], NHLBI GO Exome Sequencing Project.* 1) Internal Medicine, University of Texas Health Science Center Houston, Houston, TX, USA; 2) Department of Medical Genetics, University of Alberta, Edmonton, AB, Canada; 3) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 4) Department of Medical Genetics, University of Calgary, Calgary, AB, Canada; 5) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 6) Department of Clinical Genetics, Erasmus Medical Centre, Rotterdam, The Netherlands; 7) Department of Genome Sciences, University of Washington, Seattle, WA, USA.

Thoracic aortic aneurysms leading to acute aortic dissections (TAAD) can be inherited in families in an autosomal dominant manner with reduced penetrance and variable expression. Recently we reported that a subset of TAAD families with members at risk for TAAD who presented with either intracranial aneurysms (ICA) or abdominal aortic aneurysms (AAA). To identify the causative mutation in a large family with autosomal dominant inheritance of TAAD with ICA and AAA, we performed exome sequencing of relatives with TAAD and 1/16th relatedness. A novel frame shift mutation, p.N218fs (c.652delA), was identified in SMAD3 that segregated with the vascular diseases in this family (LOD score 2.52). Sequencing of 181 affected probands from unrelated TAAD families identified three additional SMAD3 mutations in 4 families, p.R279K, p.E239K, and p.A112V (combined LOD score of 5.21). Sequencing of selected TAAD families identified two additional SMAD3 mutations, p.W406X and p.P263A. These variants are absent in 4500 control chromosomes and predicted to disrupt protein function based on Polyphen analysis. Of the 50 individuals who carry a SMAD3 mutation, 25 presented with TAAD, 5 with ICA, 2 with AAA, and 2 with bilateral iliac artery aneurysms. The features of Aneurysms Osteoarthritis Syndrome were notably absent in the majority of mutation carriers. Fibroblasts explanted from the Smad3-/- mice were previously shown to have attenuated TGF-) -induced expression of (-smooth muscle (SM) actin and connective tissue growth factor (CTGF) compared with wildtype fibroblasts, indicating that trans-differentiation of fibroblasts to myofibroblast is dependent on Smad3. With exposure to TGF-) 1, fibroblasts explanted from SMAD3 mutation carriers also showed decreased expression and protein levels of SM (-actin, SM) -myosin and CTGF when compared to matched control fibroblasts, along with decreased phosphorylation of SMAD3. Similar results were obtained from SM cells explanted from an aorta of a SMAD3 mutation carrier when compared to SM cells from a matched control aorta. In summary, SMAD3 mutations are responsible for 3% of familial TAAD, are identified in families with TAAD alone and families with TAAD, ICA, AAA and iliac artery aneurysms, and lead to loss of SMAD3 signaling associated with decreased contractile protein expression in cells.

216

Bone morphogenic protein receptor 2 (BMPR2) alternative splicing has a role in the reduced penetrance seen in autosomal dominant heritable pulmonary arterial hypertension (HPAH). *R. Hamid, L. Hedges, B. Womack, J. Phillips, J. Cogan.* Dept Pediatrics, Div Med Gen, Vanderbilt Univ Sch Med, Nashville, TN.

Only 20% of BMPR2 mutation carriers develop HPAH. The molecular reasons of this striking reduced penetrance are unknown. Here we present data that show, for the first time, that this reduced penetrance is likely due to BMPR2 alternative splicing. A preliminary analysis of 537 BMPR2 EST sequences showed two major alternatively spliced variants; isoform-A and isoform-B. Both variants are expressed in all tissues analyzed. Isoform A consists of the full-length BMPR2 transcript while the isoform B lacks exon 12 of the BMPR2 gene. Interestingly deletion of exon 12 is also a common BMPR2 mutation found in HPAH patients and studies have shown that it can disrupt BMPR2 function in a dominant negative fashion. Furthermore studies in mice have shown that overexpression of exon 12-deleted BMPR2 transcript results in PAH phenotypically similar to human HPAH. We hypothesized that relative amount (ratio) of isoform-A compared to isoform-B in HAPH patients is associated with disease penetrance. To investigate this hypothesis we analyzed the relative amounts of these isoforms by isoform specific real-time PCR analysis in PBMCs of 46 BMPR2 mutation positive HPAH patients and 31 BMPR2 mutation positive unaffected carriers. Our data show that affected individuals had relatively higher isoform-B expression compared to isoform-A expression (B/A ratio) than unaffected individuals ($p<0.002$). We then proceeded to determine the molecular mechanism behind this differential isoform expression. Analysis of exon 12 sequence identified a putative exonic splice enhancer whose mutations result in exon 12 skipping leading to increased isoform-B mRNA. Using siRNA and western blot analysis we identified three splicing factors that bind to this splice enhancer and regulate BMPR2 exon 12 skipping. We then conducted a focused drug screen and identified FDA approved pharmacological agents, which alter the activity of this enhancer resulting in alteration of the splice isoform ratio in either direction. CONCLUSIONS: 1) Penetrance of BMPR2 related HPAH can be predicted by BMPR2 isoform ratio; 2) this isoform ratio is controlled by an exonic splice enhancer and its associated splicing factors; 3) this isoform ratio can be manipulated in either direction by FDA approved pharmacological agents.

217

Intracranial Aneurysm risk locus 5q23.2 is associated with elevated systolic blood pressure: evidence for pathomechanism of a complex disease locus. *E.I. Gaál[1, 2, 3], P. Salo[1], M. Niemelä[3], K. Kristiansson[1, 2], M. Günel[4], M.S. Nieminen[5], J. Sinisalo[5], M-L. Lokki[6], A. Jula[1], O. Raitakari[7,8], T. Lehtimäki[9], J. Eriksson[1, 10, 11, 12, 13], E. Widen[2], M. Kurki[14], M. Fraunberg[14], K. Rehnström[2,15], J.E. Jääskeläinen[14], J. Hernesniemi[3], I.C.B.P.-GWAS[18], A. Palotie[15], V. Salomaa[16], M. Perola[1,2,17].* 1) Public Health Genomics Unit, Dept. of Chronic Disease Prevention, the National Institute for Health and Welfare, Helsinki, Finland; 2) Institute for Molecular Medicine Finland FIMM, University of Helsinki, Finland; 3) Department of Neurosurgery, Helsinki University Central Hospital, Helsinki, Finland; 4) Department of Neurosurgery, Yale University School of Medicine, New Haven, Connecticut, USA; 5) Division of Cardiology, Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland; 6) Transplantation Laboratory, Haartman Institute, University of Helsinki, Helsinki, Finland; 7) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Finland; 8) Clinical Physiology, University of Turku and Turku University Hospital, Finland; 9) Department of Clinical Chemistry, University of Tampere and Tampere University Hospital, Tampere, Finland; 10) Department of General Practice and Primary Health Care, Institute of Clinical Medicine, University of Helsinki, Helsinki, Finland; 11) Vasa Central Hospital, Vasa, Finland; 12) Folkhälsan Research Centre, Helsinki, Finland; 13) Helsinki University Central Hospital, Unit of General Practice, Helsinki, Finland; 14) Neurosurgery of Kuopio University Hospital, NeuroCenter, Kuopio, Finland; 15) The Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 16) Chronic Disease Epidemiology and Prevention Unit, Dept. of Chronic Disease Prevention, the National Institute for Health and Welfare, Helsinki, Finland; 17) University of Tartu, Estonian Genome Centre, Tartu, Estonia; 18) The International Consortium for Blood Pressure Genome-wide Association Studies.

Pathomechanism of most complex disease loci or genes remains elusive. Studying the genetics of risk factors predisposing to disease is an attractive approach for identifying function. GWA studies have identified five loci with strong association (posterior probability of association -PPA>0.5) and further fourteen loci with suggestive association (0.1<PPA<0.5) to intracranial aneurysm (IA). We tested whether there are IA loci that convey their effect through elevating the blood pressure (BP). We hypothesized that complex disease loci are more likely identifiable in a genetically more homogenous population.

We performed a meta-analysis of four Finnish cohorts consisting of altogether 10,104 individuals to assess the association of previously identified IA candidate loci (n=19) with BP. We used a two tier approach; one discovery cohort and three replication cohorts. We defined systolic BP (SBP), diastolic BP, mean arterial pressure, and pulse pressure as quantitative outcome variables. Significant results were further tested for association in the ICBP-GWAS cohort of 200,000 individuals.

We found that the suggestive IA locus at 5q23.2 in *PRDM6* was strongly associated with SBP in all individuals of European descent ($p_{finns}= 0.00016$, $p_{icbp\text{-}gwas}= 0.00078$). The risk allele of IA was associated with higher SBP. *PRDM6* encodes a protein predominantly expressed in vascular smooth muscle cells (VSMC). We hypothesized that common variants in *PRDM6* may contribute to altered vascular wall structure hence increasing SBP, and predisposing to IA.

Our study connects a complex disease locus (IA), with its common risk factor (SBP) and implies *PRDM6* in the pathogenesis of both. PRDM6 protein activation leads to VSMC proliferation. Excessive VSMC proliferation exacerbates hypertension and is a putative pathomechanism of the cardiovascular risk locus at 9p21.3, the strongest known IA risk locus. Our results show that traditional risk factors are useful intermediate phenotypes in complex disease mapping. True positive associations often fail to reach genome-wide significance in GWA studies. Our findings show that a meta-analysis of intermediate phenotypes is an effective tool for identifying type II errors, hence helping to decipher hidden heritability. Further, we demonstrate that common disease loci identified in a population isolate may bear general significance.

218

The 22q11.2 Deletion Syndrome: Second Hit Copy Number Variants and Conotruncal Heart Defects. *M.B. Sheridan[1], T. Guo[2], D. McDonald-McGinn[1], M. Bowser[1], M. Xie[3], X. Gai[3], J.C. Perin[3], A. Bassett[4], E. Chow[4], A. Blonska[2], A. Shanske[2], F. Beemer[5], K. Devriendt[6], M.C. Digilio[7], B. Marino[8], B. Dallapiccola[7], A-M. Higgins[9], N. Philip[10], T.J. Simon[11], K. Coleman[12], W. Kates[13], M. Devoto[1,14], E. Zackai[1,14], J. Ott[15], D. Heine-Suñer[16], T. Shaikh[17], R. Shprintzen[2,9], B. Morrow[2], B.E. Emanuel[1,14], The International Chromosome 22q11.2 Consortium.* 1) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Genetics, Albert Einstein College of Medicine, Bronx, NY; 3) Bioinformatics Core, Children's Hospital of Philadelphia, Philadelphia, PA; 4) Clinical Genetics Research Program, Centre for Addiction and Mental Health and Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; 5) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 6) Center for Human Genetics, University of Leuven, Leuven, Belgium; 7) Medical Genetics, Bambino Gesù Hospital, Rome, Italy; 8) Department of Pediatrics, La Sapienza University of Rome, Rome, Italy; 9) Velo-Cardio-Facial Syndrome International Center, Department of Otolaryngology and Communication Science, SUNY Upstate Medical University, Syracuse, NY; 10) Department of Medical Genetics, AP-HM and University of Mediterranee, Timone Children's Hospital, Marseille, France; 11) M.I.N.D. Institute & Department of Psychiatry and Behavioral Sciences, University of California, Davis, CA; 12) Children's Healthcare of Atlanta, Atlanta, GA; 13) Department of Psychiatry and Behavioral Sciences, and Program in Neuroscience, SUNY Upstate Medical University, Syracuse, NY; 14) Department of Pediatrics, The University of Pennsylvania School of Medicine, Philadelphia, PA; 15) Institute of Psychology, Chinese Academy of Sciences, Beijing, China; 16) Genetics Department, Hospital Universitari Son Espases, Palma de Mallorca, Spain; 17) Department of Pediatrics, University of Colorado, Denver, CO.

The 22q11.2 deletion syndrome (22q11DS) is the most common microdeletion disorder, occurring in ~1:4,000 births. It is associated with a wide variety of phenotypes including conotruncal heart defects (CTDs). Approximately 70% of 22q11DS patients have a CTD. We hypothesize that 22q11DS patients with CTDs have copy number variants (CNVs), other than the 22q11.2 deletion, that affect genes critical to heart development. We used the Affymetrix SNP 6.0 platform to identify CNVs in a large cohort of 22q11DS patients (399 with CTDs and 227 with a normal heart). Analysis focused on CNVs detected by two software packages (CNV Workshop and Partek Genomics Suite (PGS)). These CNVs are unlikely to be false positives as they were called by independent methods. Indeed, we validated 100% (50/50) of the CNVs predicted by both methods that were tested using quantitative PCR. The validation rate was much lower for CNVs predicted by only one algorithm [9/29(31%) CNV workshop; 6/17(35.3%) PGS]. There was no difference in the genomic location of known common CNVs (projects.tcag.ca/variation) in patients with and without CTDs with the exception of one CNV. A duplication of *SLC2A3* on chr. 12p13.31 was detected in 13 CTD patients and was not identified in any 22q11DS patients with a normal heart (p=0.0025). *SLC2A3* encodes a glucose transporter that is expressed in cells with high energy needs. This gene has not previously been associated with CTDs, but is expressed in human adult and fetal myocardium. Additionally, several large rare CNVs were identified in the 626 22q11DS samples that were not seen in healthy controls. To determine the relevance of rare copy number altered genes to cardiac development, we used phenotype data from ~5000 mouse knockout experiments from the Mouse Genome Informatics Resource (www.informatics.jax.org). Among the rare CNVs enriched in patients with CTDs versus patients with a normal heart are those affecting genes that are implicated in biological processes including “abnormal embryogenesis/development” and “abnormal cardiovascular system morphology”. These include deletions in cadherin 2 (*CDH2*) and semaphorin 3E (*SEMA3E*) and duplications of methylenetetrahydrofolate dehydrogenase 1 (*MTHFD1*) and paired box 8 (*PAX8*). Each was detected in one 22q11DS patient with a CTD. Collectively, these data support the hypothesis that “second hit” CNVs may affect genes important in heart development and alter the risk for CTDs in some 22q11DS patients.

219

Autophagy failure in Alzheimer's disease. *R. Nixon[1, 2, 3], J.H. Lee[2, 3], D.S. Yang[2, 3], S. Lee[3].* 1) Cell Biology Dept, NYU Langone Medical Cntr, New York, NY; 2) Psychiatry Dept, NYU Langone Medical Cntr, New York, NY; 3) Center for Dementia Research, Nathan Kline Institute, Orangeburg, NY.

In the extensive neuritic dystrophy of Alzheimer's disease (AD), a pathological hallmark of the disease, autophagic vacuoles containing incompletely digested proteins selectively accumulate in focal axonal swellings, reflecting defects in both autophagy and axonal transport. Recently, we investigated the possibility that impaired lysosomal proteolysis could be a basis for defects in both autophagy and axonal transport leading to axonal dystrophy. In primary cortical neurons expressing fluorescence-tagged markers, disrupting lysosomal proteolysis by either inhibiting cathepsins directly or suppressing lysosomal acidification slowed the axonal transport of autophagy-related organelles but not other organelles and caused their accumulation within AD-like dystrophic axonal swellings. These studies identify AD-associated defects in neuronal lysosomal proteolysis as a possible basis for the selectively disrupted transport of autophagy-related organelles and characteristic pattern of neuritic dystrophy in AD. We also found that the AD-related gene presenilin1 (PS1) is essential for lysosomal proteolysis and autophagy and plays a novel role in lysosome acidification required for protease activation. PS1 mutations causing familial AD (FAD) confer partial loss of these same functions in fibroblasts from PS-FAD patients and in neurons of PS1/APP mutant mice. Presenilin 2 has effects on the lysosomal system quite distinct from those of PS1. Lysosomal and autophagy dysfunction also develops in sporadic AD and in AD mouse models driven in part by other AD-related genes, including amyloid precursor protein and apolipoprotein E. Supporting the pathogenic significance of lysosomal system dysfunction in AD, we found that partially restoring deficient autophagy in the CRND8 mouse model of AD by genetically manipulating lysosomal protease activities substantially ameliorates lysosomal pathology, amyloid burden, neuritic dystrophy, and memory deficits. Support: National Institute on Aging.

220

Resequencing of the complement gene pathway to identify rare variants that affect risk of Alzheimer's disease. *M.K. Lupton[1], S. Newhouse[1], M. Weale[2], S. Lovestone[1], J.F. Powell[1].* 1) MRC Centre for Neurodegeneration Research, Institute of Psychiatry, Kings College London, United Kingdom; 2) Division of Genetics and Molecular Medicine, King's College London, London, United Kingdom.

There is increasing evidence implicating complement activation, a critical part of the innate immune system, in Alzheimer's disease (AD). Inflammation is an early event in AD where A) is a powerful trigger of the complement cascade, suggesting a primary role in AD pathogenesis. Recent GWAS studies have identified associations in genes involved in complement activation; CR1 and CLU, although with small effect sizes. Rare variation in the complement pathway may explain some of the missing heritability, where several rare variants may accumulate in a single individual contributing to AD risk. To test this hypothesis we are resequencing all exons, UTRs and promoter regions of 55 genes in the complement pathway using next generation sequencing. These regions are enriched using custom sure select in-solution capture (Agilent). Molecular barcoding allows multiple individuals to be pooled in a single sequencing lane of the Illumina genome analyser. Alignments were carried out using Novoalign and subsequent filtering and the production of quality metrics by PICARD tools and GATK. Twelve individuals have been sequenced thus far in one lane of sequencing with an average depth of 412 and over 95% of the captured regions having at least 40X coverage. After quality control filtering 1184 SNPs were called. To validate the method an individual of Yoruba ethnicity with publically available data was included. Of 190 common SNPs called in HapMap data 99% had matching genotype calls. Of the 450 SNPs identified by the 1000 genome project (depth of 2-4X) 96% were in agreement, with an additional 175 SNPs called in our high coverage data. Using this method 1000 individuals will be sequenced including AD cases, controls and individuals with mild cognitive impairment from the Addneuromed study (a multicentre European collaboration). Association of identified variants will be tested with AD risk both individually and cumulatively based on functional predictions. This large cohort has whole genome SNP data, MRI scans and plasma biased biomarkers, allowing testing for association of identified variants with quantitative traits related to disease progression and complement activation. This study validates the use of in-solution capture and DNA bar-coding, allowing affordable sequencing of candidate regions in complex disease genetics. The final data will examine the importance of both common and rare variation in the complement pathway in AD pathogenesis.

221

Identification of Rare Variants for Late-Onset Alzheimer Disease in a Family-Based Genome-Wide Association Study. *G. Jun[1,2,3], J. Buros[1], K.L. Lunetta[3], T.M. Foroud[6], R. Mayeux[7,8], L.A. Farrer[1,2,3,4,5].* 1) Medicine (Biomedical Genetics Section), Boston University School of Medicine, Boston, MA; 2) Ophthalmology, Boston University School of Medicine, Boston, MA; 3) Biostatistics, Boston University School of Public Health, Boston, MA; 4) Epidemiology, Boston University School of Public Health, Boston, MA; 5) Neurology, Boston University School of Medicine, Boston, MA; 6) Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN; 7) The Taub Institute on Alzheimer's Disease and the Aging Brain and The Gertrude H. Sergievsky Center, College of Physicians and Surgeons, Columbia University, New York, NY; 8) Epidemiology, Mailman School of Public Health, Columbia University, New York, NY.

Including recent genome wide association studies (GWAS), approximately 10 loci are robustly associated with late-onset Alzheimer disease (LOAD). However, about one-half of the genetic component of AD is unexplained. Arguably, much of the “missing heritability” is accounted by rare variants (minor allele frequency [MAF] < 5%) which even large GWAS are not powered to detect. We conducted a GWAS for LOAD using two large family-based cohorts. The discovery set was the Framingham Study (FS) cohort (2364 subjects from 750 families including 55 incident cases and 2530 elderly cognitively normal controls) and top SNPs were genotyped in a replication set, the NIA-LOAD cohort (3828 subjects from 2265 families including 1840 cases and 1988 controls). Analyses were performed using an approach that took into account family structure and an outcome variable based on residuals from a model for LOAD adjusted for age, sex and principal components (PCs). To address potential confounding based on differences between the cohorts in age of onset (mean = 82 in FS and 73 in NIA-LOAD) and *APOE* (frequency of ε4 = 12% in FS and 33% in NIA-LOAD), we analyzed imputed SNPs from the top-ranked genes in two additional models with further adjustment for dose of ε4 and for age of onset conditional on parental affection status. Analysis of the discovery dataset with the base model revealed significant association ($p<10^{-6}$) with five rare SNPs ranging MAF from 0.01 to 0.05 (rs4920448 in *IGSF21*, $p = 2.3 \times 10^{-7}$; rs9311482 in *ITIH3*, $p = 4.6 \times 10^{-9}$; rs16824295 near *MME*, $p = 4.1 \times 10^{-7}$; rs277484 in *PLXNA4*, $p = 9.0 \times 10^{-10}$; rs13057714 in *MYO18B*, $p = 8.9 \times 10^{-9}$). While none of these SNPs replicated in NIA-LOAD using the base model as well as the extended model with dose of ε4, adjustment for age of onset yielded two very significant *IGSF21* SNPs (rs17435018, $p = 7 \times 10^{-7}$; rs1869806, $p = 3 \times 10^{-13}$), both near rs4920448, suggesting that multiple variants in *IGSF21* influence risk of LOAD. Replication analysis in a third family cohort is ongoing. This study demonstrates the power of a family-based design for detecting association in GWAS, particularly for variants with frequencies of less than 5%.

222

Alzheimer's risk variants in the CLU gene modify the use of alternative transcription start sites. *M. Szymanski[1], R. Wang[2], S. Bassett[2], D. Avramopoulos[1,2].* 1) McKusick Nathans Institute of Genetic Medicine, School of Medicine, Johns Hopkins University Baltimore MD, 21205; 2) Department of Psychiatry, School of Medicine, Johns Hopkins University Baltimore MD, 21205.

Two recent GWAS have implicated the CLU gene in Alzheimer's disease (AD) followed by many replications, however studies of the coding sequence and microarray expression data have not identified a connection to coding variation or to the overall transcription levels of the gene. CLU has four distinct transcription start sites and undergoes alternative splicing which is suggested to regulate the production of secreted and nuclear forms with different and opposing functions. We hypothesized that the biological impact of the variants associated with AD or other variation in CLU may act through disruption of the gene's splicing, transcription start site regulation or specific transcript expression. We tested this hypothesis by genotyping single nucleotide polymorphisms (SNPs) in and around CLU and quantifying each transcript by specific real time PCR in 190 temporal lobe samples without pathology. We found the risk allele of the AD associated SNP rs9331888 to increase the relative abundance of transcript NM_203339 ($p= 4.3 \times 10\text{-}12$). We replicated this result in an independent set of 115 AD and control samples ($p=0.0014$) where we also observed that multiple CLU transcripts are at higher levels in AD compared to controls. We proceeded to test whether rs9331888 is the functional variant as it is located in the first exon of NM_203339. In vitro dual luciferase assays using SK-N-SH cells and mouse primary cortical neurons showed significant allelic differences of enhancer function in a direction consistent with the post mortem brain transcript measurements. The SNP rs9331888 has been associated with AD, however another SNP, rs11136000, whose risk allele is more frequent and in positive linkage disequilibrium with the rs9331888 risk allele, shows more consistent results. Our results suggest a biological mechanism for the genetic association of CLU with AD risk and indicate rs9331888 as one of the underlying functional DNA variants. Additional functional variation and the path from splicing variation to disease remain to be examined.

223

Investigating Alzheimer's Disease using knock-in mouse models. *Q. Guo[1,2], N. Justice[1,3], H. Zheng[1,2,3].* 1) Huffington Center on Aging, Baylor College of Medicine, Houston, TX., US; 2) Translational Biology & Molecular Medicine Graduate Program, Baylor College of Medicine, Houston, TX., US; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX., US.

Alzheimer's Disease (AD) is the most common cause of dementia in the elderly. There are three pathological hallmarks of AD: 1) plaques formed by A) peptide deposition; 2) neurofibrillary tangles (NFT) caused by aggregation of hyperphosphorylated tau protein; 3) loss of neurons and synapses. AD patients not only suffer from impaired learning and memory ability, but also mood disorders, like anxiety and depression, and behavioral changes. A) is one of the proteolytic cleavage products of Amyloid Precursor Protein (APP). PS1 and PS2 are components of .-secretase which mediates APP cleavage and A) production. Genetic mutations in *APP*, *PS1* or *PS2* affect APP cleavage, leading to early onset of Familial Alzheimer's Disease (FAD). While mouse tau does not develop NFT pathology, transgenic mice expressing genomic wild-type human tau (*htau-tg*) in the absence of mouse tau (*mtau -/-*) was shown to display AD-like tau pathology. To better mimic human AD in mice and to avoid potential complications from transgene overexpression, we have generated three knock-in (KI) mouse models based on previously established KI alleles, including $APP^{sI/A)}$*/sI/A)* single knock-in (SKI) which expresses the humanized A) sequence and carries the Swedish and London mutations, $APP^{sI/A)}$*/sI/A) ; PS1*$^{M146V/M146V}$ double knock-in (DKI), the latter contains the *PS1 M146V* FAD mutation, and $APP^{sI/A)}$*/sI/A) ; PS1*$^{M146V/M146V}$*; htau-tg; mtau -/-* pseudo triple knock-in (pTKI). The SKI line serves as a pre-plaque model, DKI line serves as a plaque model, and pTKI line serves as a model to study both plaques and tangles. Behavioral tests revealed that all three lines exhibit increased anxiety levels at an early age (3-4 months) on elevated plus maze test. DKI and pTKI animals also showed increased freezing frequency in conditioned fear test. The Corticotropin Releasing Factor (CRF) levels and resting glucocorticoid levels in DKI animals were found to be higher than controls, suggesting that enhanced Hypothalamic-Pituitary-Adrenal (HPA) axis activity may contribute to the increased anxiety and freezing. We are currently investigating the possible changes in cognition, synaptic plasticity and evidence of neuronal and/or synaptic loss and correlate these changes as a function of age and neuropathology.

224

Genome-wide association of Progressive Supranuclear Palsy (PSP) and the role of *MAPT* locus in PSP and Parkinson's disease. *N. Melhem[1], G. Hoglinger[2], D. Dickson[3], P. Sleiman[4], A. Singleton[5], B. Devlin[1], U. Muller[6], G. Schellenberg[7], PSP Genetics Study Group, International Parkinson Disease Genomics Consortium.* 1) Psychiatry, University of Pittsburgh, Pittsburgh, PA; 2) Neurology, Philipps-Universitat, Marburg, Germany; 3) Neuroscience, Mayo Clinic, Jacksonville, FL; 4) Center for Applied Genomics, Childrens hospital of Philadelphia, Philadelphia, PA; 5) Laboratory of Neurogenetics, National Institute on Aging, NIH, Bethesda, Maryland; 6) Institut for Humangenetik, Justus-Liebig-Universitat, Germany; 7) Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA.

Objective: Progressive supranuclear palsy (PSP) is a movement disorder with prominent tau neuropathology and the second most common cause of degenerative Parkinsonism, after Parkinson's Disease (PD). Brain diseases with abnormal tau deposits are called tauopathies, the most common being Alzheimer's disease. Environmental causes of tauopathies include repetitive head trauma associated with some sports. We carried out a genome-wide association (GWA) study of PSP to identify common genetic variation contributing to risk for tauopathies. We followed up this work with additional analyses to fine-map the *MAPT* region because variation therein has been implicated not only in risk for PSP but also for PD. In the fine-mapping process, we seek shared and unique genetic risk variants for PSP and PD. Methods: A two-stage GWA study of 1,114 PSP cases and 3,247 controls (Stage 1); and 1,051 cases and 3,560 controls (Stage 2) genotyped for Stage 1 SNPs that yielded $P \leq 10^{-3}$, was conducted. For the fine mapping of the *MAPT* locus in PSP and PD, we analyzed 1,713 PD Cases and 7,560 controls (Stage 1) and 1,811 PD cases and 6,500 controls (Stage 2). We also examine whether the identified SNPs are expression Quantitative Trait Loci (eQTLs) for gene expression in the *MAPT* region. Results: We found significant novel signals ($P < 5 \times 10^{-8}$) associated with PSP risk at *STX6*, *EIF2AK3*, and *MOBP*. We confirmed two independent variants in *MAPT* affecting risk for PSP, one of which influences *MAPT* brain expression. We also find a highly significant interaction between the *MAPT* inversion SNP and another SNP, rs11079738, associated with increased risk of PD [OR= 1.55, 95% CI (1.34-1.79), $P=2.2\times10^{-9}$] but not PSP [OR=0.78, 95% CI (0.57-1.06), $P=1.2\times10^{-1}$]. The SNP, rs11079738, has not been detected in PD GWA studies and we find it to be associated with *ARL17A* brain expression. Conclusions: The genes implicated in PSP encode proteins for vesicle-membrane fusion at the Golgi-endosomal interface, for the endoplasmic reticulum unfolded protein response, and for a myelin structural component. Risk variants in the *MAPT* region exert impact on risk for PD and PSP, either independently or in interaction with the inversion, and these risk variants are different for both diseases.

225

A common variant in Myosin-18B contributes to mathematical performance in children with dyslexia and intraparietal sulcus variability in adults. *K.U. Ludwig[1,2], P. Sämann[3], M. Alexander[1,2], J. Becker[1,2], J. Bruder[4], K. Moll[5], S. Streiftau[4], D. Spieler[3,6], M. Czisch[3], S.J. Docherty[7], O.S.P. Davis[7], R. Plomin[7], M.M. Nöthen[1,2], K. Landerl[8], B. Müller-Myhsok[3], P. Hoffmann[1,2], J. Schumacher[1], G. Schulte-Körne[4], D. Czamara[3].* 1) Institute of Human Genetics, University of Bonn, Bonn, Germany; 2) Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany; 3) NMR Research Group, Max Planck Institute of Psychiatry, Munich, Germany; 4) Department of Child- and Adolescent Psychiatry, University of Munich, Munich, Germany; 5) Department of Psychology, University of Salzburg, Salzburg, Austria; 6) Institute of Human Genetics, Helmholtz Zentrum Munich, Neuherberg, Germany; 7) King's College London, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, London, United Kingdom; 8) Department of Psychology, University of Graz, Graz, Austria.

Mathematical ability is a quantitative trait for which substantial heritability estimates have been reported. To date, only one genome-wide association study (GWAS) for mathematical performance has been published. This study by Docherty and colleagues was performed on individuals drawn from the general population, however, it failed to detect any genome-wide significant association. As large genetic heterogeneity might have contributed to this, we assumed that the investigation of a more homogeneous sample would overcome this limitation. Notably, the prevalence of mathematical disabilities among children with dyslexia is clearly higher than in the general population. We thus hypothesized that distinct cognitive processes underlying mathematical disability are preferentially affected in dyslexia patients, and that investigation of such sample would thus increase statistical power. We performed a genome-wide association study on 200 German dyslexics, using different dimensions of mathematical ability as quantitative traits. Replication of the three top markers in an independent sample of children with dyslexia (n = 183) revealed that rs133885 on chromosome 22q12.1 is a genome-wide significant marker for mathematical abilities (Pcombined= 8.81 x 10-10). We replicated this finding in a second independent dyslexia sample (n = 316, P = 0.0446), resulting in an overall P-value of 7.71 x 10-10. Our analyses show that rs133885 accounts for 4.87% of the variance in mathematical performance in samples with dyslexia. Investigation of rs133885 in two independent data sets of non-dyslexics (n = 319) and individuals that were not screened for dyslexia status (n = 1,081), respectively, confirmed the association, albeit with a smaller effect size. The variant rs133885 encodes an amino-acid substitution in MYO18B, a member of the family of unconventional myosins. To gain insight into its biological function, we investigated whether rs133885 is associated with structural variability of the right intraparietal sulcus (IPS), a key area of the brain in terms of numerical processing. This genotype-phenotype correlation in healthy individuals revealed the presence of a structural MRI-endophenotype, since carriers of the MYO18B risk-genotype displayed a reduced depth of the right IPS. Our study identified rs133885 as the first genome-wide significant marker to contribute to human mathematical ability and implicates MYO18B in cognitive processes.

226

Genome-wide consequences of compromised RNA surveillance and its relevance for normal brain function. *L.S. Nguyen[1,2], L. Jolly[2], C. Shoubridge[1,2], W.K. Chan[3], L. Huang[4], F. Laumonnier[5,6,7], M. Raynaud[5,7,8], A. Hackett[9], M. Filed[9], J. Rodriguez[10], A.K. Srivastava[10], Y. Lee[11], A.M. Addington[11], J.L. Rapoport[11], S. Suren[12], C. Hahn[13], J. Gamble[14], M.F. Wilkinson[4], M.A. Corbett[2], J. Gecz[1,2].* 1) Dept of Paediatrics, University of Adelaide, North Adelaide, South Australia, Australia; 2) Dept of Genetics Medicine, SA Pathology, South Australia, Australia; 3) Dept of Bioinformatics and Computational Biology, University of Texas M.D. Anderson Cancer Center, TX, USA; 4) Dept of Reproductive Medicine, University of California, San Diego, CA, USA; 5) INSERM, U930, Tours, France; 6) CNRS, ERL3106, Tours, France; 7) University Francois-Rabelais, UMR “Imaging and Brain”, Tours, France; 8) CHRU de Tours, Service de Genetique, Tours, France; 9) GOLD Service, Hunter Genetics, Newcastle, Australia; 10) J.C. Self Research Insittute, Greenwood Genetic Centre, Greenwood, SC, USA; 11) Child Psychiatry Branch, National Institute of Mental Health, Bethesda, MD, USA; 12) Human Developmental Biology Resource, Neural Development Unit, UCL Institute of Child Health, 30 Guilford Street, London WC1N 1EH; 13) Department of Molecular Pathology, Centre for Cancer Biology, SA Pathology, South Australia, Australia; 14) Centenary Institute of Cancer Medicine & Cell Bioloy, University of Sydney, NSW, Australia.

Nonsense mediated mRNA decay (NMD) is universally conserved mechanism best known for its ability to rapidly degrade transcripts bearing premature termination codons. In man, mutations in UPF3B, a core member of NMD, cause intellectual disability (ID) with or without autism, schizophrenia and other dysmorphic features, suggesting that NMD is essential for brain development and function. To assess the impact of UPF3B-NMD deficiency on the transcriptome and to identify pathways relevant to the patients' neuronal phenotypes, we used RNA-SEQ to analyse polyA+ RNA extracted from lymphoblastoid cell lines of four patients and two controls (Illumina GAII). We complemented our RNA-SEQ results using microarray profiling of the same samples on the Affymetrix Human Exon 1.0 ST array platform. We found strong concordance in the absolute expression between the two platforms (R=0.78) and that 85% of genes have the same trend of up or down regulation as determined by either platform. The results were further validated using RT-qPCR (10/10 genes), including UPF3B itself. We identified 526 genes, 5% of the transcriptome, de-regulated in the patients by at least 2-fold. Among these genes, ARHGAP24, a negative regulator of Rho-GTPase, stood out as a strong candidate for further analysis as it is consistently upregulated in all patients, and is a member of Rac/Cdc42 pathway previously implicated in ID. Mouse primary hippocampal neurons transiently transfected with ARHGAP24 showed a significant reduction in axonal length and increased cell death, suggesting that deregulation of ARHGAP24 contributes directly to the ID phenotypes seen in the patients. Moreover, we found that UPF3A, a paralog of UPF3B, is up-regulated in the patients at varying levels and its level of up-regulation is inversely correlated to the extent of transcriptome de-regulation and the patient's phenotype severity. This suggests that UPF3A partially rescues UPF3B loss of function and that UPF3A is a major modifier of UPF3B associated phenotype. Our results support an intrinsic role of NMD in brain development and point out the need to exercise caution when manipulating NMD for therapeutic purposes.

227

New findings in phenotype-genotype correlations in Holoprosencephaly: about a large European series of 645 probands. *S. Odent[1,2], C. Dubourg[2,3], L. Pasquier[1], C. Bendavid[2,3], P. Loget[4], N. Garcelon[5], B. Campillon-Gimenez[5], S. Jaillard[3], C. Quelin[1], L. Rochard[2], V. Dupe[2], V. David[2, 3], S. Mercier[1,2].* 1) Genetique Clinique, CHU de Rennes Hosp Sud, Rennes CDX 2, France; 2) Institut de genetique et developpement, UMR 6061 CNRS, Rennes1 university, Rennes, France; 3) Laboratoire de genetique moleculaire, CHU Pontchaillou, Rennes, france; 4) Service anatomie pathologique, CHU Pontchaillou, Rennes, France; 5) Laboratoire informatique medicale, CHU Pontchaillou, Rennes, France.

Holoprosencephaly (HPE) is the most common malformation of the human forebrain resulting from incomplete midline division of the prosencephalon. We report on the largest European series of 645 HPE probands (and 699 relatives) that encompasses 51% of foetuses and 49% of children with a sex ratio (F:M) of 1,2. Mutations in the four main genes implicated in HPE (SHH, ZIC2, SIX3, TGIF) are identified in 25% and are inherited in more than 70% of cases in SHH, SIX3, and TGIF whereas 70% of mutations occurred de novo in ZIC2. Moreover 22% of rearrangements are detected in 260 array-CGH screened patients. Interestingly 8 probands have both a microrearrangement and a mutation in another HPE gene, which adds additional support to “multiple-hit process” in HPE. Based on statistical analyses, a positive correlation is found between the severity of the brain malformation and the facial features in SHH, SIX3, TGIF, but no correlation was shown in ZIC2. We particularly focus on specific findings in ZIC2: (i) probands are rather to have very mild facial features even with alobar or semilobar HPE; (ii) Middle Interhemispheric variant (MIH) is only described with ZIC2 mutations in our series; (iii) neuronal migration troubles are frequently found in ZIC2 probands; (iv) rachischisis are exclusively described with ZIC2 mutations; (iv) the highest rate of extra-brain malformations (40%) is reported in ZIC2 compared to the global series (27%). Finally we propose an algorithm based on these new phenotype-genotype correlations that helps define molecular analysis strategy and genetic counselling in HPE.

228

Rhombencephalosynapsis: an under-recognized disorder associated with aqueductal stenosis and a wide spectrum of severity. *D. Doherty[1], G.E. Ishak[2], K.J. Millen[1], D.W. Shaw[2], W.B. Dobyns[1].* 1) Dept Pediatrics, Univ Washington, Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA; 2) Dept Radiology, Seattle Children's Hospital and Univ Washington, Seattle, WA.

Rhombencephalosynapsis (RES) is a unique brain malformation characterized by loss of the cerebellar vermis with apparent fusion of the cerebellar hemispheres. RES can be seen in isolation or together with other central nervous system (CNS) and extra-CNS malformations. Gomez-Lopez-Hernandez syndrome (GLH) is the best-recognized syndrome that combines RES with parietal/temporal alopecia, and sometimes, trigeminal anesthesia, towering skull shape and dysmorphic features. RES can also be seen in patients with features of VACTERL association. Despite these distinctive clinical presentations, the genetic causes of RES remain unknown. To date, neither recurrences nor vertical transmission have been reported, indicating that disorders associated with RES are likely due to de novo dominant mutations. To aid in the categorization of patients with RES for gene identification studies, we performed a comprehensive evaluation of neuroimaging findings in 42 patients, the largest series in the literature to date. Based on the imaging findings, we propose a spectrum of severity, ranging from mild (loss of posterior vermis) to moderate (loss of posterior and anterior vermis) to severe (loss of posterior and anterior vermis as well as nodulus) to complete (loss of entire vermis and nodulus). The severity of RES correlates with other hindbrain malformations including fusion of the tonsils and cerebellar herniation, as well as midbrain and forebrain abnormalities including aqueductal stenosis, midline fusion of the tectum, absent septum pellucidum, and abnormalities of the corpus callosum, fornices and mammillary bodies. Patients with isolated RES, GLH and RES + VACTERL features could not be distinguished from each other based on brain imaging findings, and severe ventriculomegaly was present in patients from each group. These findings prompted us to evaluate the brain MRIs from all other patients with a radiologic diagnosis of aqueductal stenosis at our institution, and remarkably, we identified RES in 9%. These data indicate that RES is more common than previously appreciated and necessitates close examination of the cerebellum in patients with aqueductal stenosis. Careful phenotyping using clinical and imaging data is essential for the design and interpretation of genomic approaches to identify the genetic causes of RES.

229

Phenotypes and genetics of polymicrogyria: copy number variations and evidence for a novel locus for bilateral perisylvian polymicrogyria narrowed to 2p16.1-p16.3. *D. Amrom[1,4,5,9], A. Poduri[10,11], B. Dan[15,17,19], N. Deconinck[15,17,19], C. Christophe[16,17,19], B. Pichon[18,19], F. Dubeau[2,4,5,9], D. Tampieri[3,4,8,9], G. Kuchukhidze[21], W. Dobyns[20], C. Walsh[12,13,14], F. Andermann[2,4,5,6,9], E. Andermann[1,4,5,7,9].* 1) Neurogenetics Unit; 2) Epilepsy Service and Seizure Clinic; 3) Diagnostic and Interventional Neuroradiology Divisions; 4) Department of Neurology, Montreal Neurological Hospital & Institute, Montreal, PQ, Canada; 5) Departments of Neurology & Neurosurgery; 6) Pediatrics; 7) Human Genetics; 8) Radiology; 9) McGill University, Montreal, Quebec, Canada; 10) Division of Epilepsy and Clinical Neurophysiology; 11) Department of Neurology; 12) Division of Genetics and Manton Center for Orphan Disease Research; 13) Children's Hospital Boston, Boston, MA; 14) Howard Hughes Medical Institute, Beth Israel Deaconess Medical Center; and Harvard Medical School, Boston, MA; 15) Pediatric Neurology Unit; 16) Department of Neuroimaging; 17) Hôpital Universitaire des Enfants Reine Fabiola (HUDERF); 18) Department of Medical Genetics, Hôpital Erasme; 19) Université Libre de Bruxelles, Brussels, Belgium; 20) Departments of Pediatrics and Neurology, University of Washington; and Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA; 21) Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria.

BACKGROUND: Polymicrogyria (PMG) is a malformation of cortical development. Common clinical features include seizures, developmental delay, oromotor dysfunction and motor disabilities. PMG is a clinically and etiologically heterogenous condition. It can be attributable to environmental causes, single gene disorders with various patterns of inheritance, or chromosomal rearrangements, usually sporadic. PURPOSE: To report the phenotypes and copy number variation (CNV) data of a cohort of PMG patients and establish genotype-phenotype correlations. METHODS: Search of our brain malformation databases and inclusion of all types of PMG, except those associated with schizencephaly and confirmed congenital CMV/toxoplasmosis infections; detailed review of medical records; karyotype and FISH 22q11; CGH and/or SNP microarray of genomic DNA. RESULTS: We enrolled 29 patients: 13 symmetric bilateral perisylvian polymicrogyria (BPP), 2 asymmetrical BPP, 6 unilateral right PMG, 4 unilateral left PMG, 2 bilateral temporo-occipital, and 2 bilateral parieto-occipital PMG. Associated brain malformations were found: microcephaly in 2 patients, nodular heterotopia in 4 patients. A 22q11 deletion was found in one patient with unilateral right PMG. The brain MRI showed an associated large contralateral frontal heterotopias. He presented with congenital mitral valve stenosis, congenital left hemiparesis, facial dysmorphism, and moderate mental retardation. A 2p13.3-p16.3 duplication was found in a patient with symmetric BPP. He presented with neonatal global hypotonia, feeding difficulties, and delayed psychomotor development. At the age of 10 years, physical examination revealed mild facial dysmorphic signs, mental retardation, severe language delay, attention deficit helped by methylphenidate, and growth deficiency treated with growth hormone. Phenotypic comparison with previously published patients harboring other types of proximal 2p duplications shows a subgroup of patients with BPP sharing a common locus narrowed to 2p16.1-p16.3. CONCLUSIONS: The inheritance of PMG is heterogeneous. Our observations further illustrate that PMG can be due to CNV, here 2 out of 30 patients (6.6%) had a chromosomal rearrangement. Unilateral PMG may be associated with deletion 22q11 (DiGeorge syndrome). This suggests asymmetrical gene(s) expression between the hemispheres. Phenotypic and CNV comparison with previously published patients allowed us to narrow a locus for BPP to 2p16.1-p16.3.

230

Posterior fossa anomalies diagnosed with fetal MRI: Associated anomalies and neurodevelopmental outcomes. *K.J. Patek[1], B.M. Kline-Fath[3], V.V. Pilipenko[1], C.G. Spaeth[1,2], T.M. Crombleholme[2], R.J. Hopkin[1,2].* 1) Human Genetics, Cincinnati Children's Hospital, Cincinnati, OH; 2) The Fetal Care Center of Cincinnati, Cincinnati Children's Hospital , Cincinnati, OH; 3) Department of Radiology, Cincinnati Children's Hospital Cincinnati, OH.

Posterior fossa anomalies (PFA) have been challenging to provide prenatal counseling for since the advent of high quality ultrasound. While the risk for developmental disability has been reported to range from 20-80%, up to 80% of pregnancies with these malformations are terminated due to the perceived risk of poor neurologic or developmental outcome. There is a clear need to define and understand those features which will help differentiate fetuses at high risk from those with a better prognosis. **Objective**: The purpose of this study was to investigate associated imaging features which may increase the ability to predict prognosis when a PFA has been prenatally identified. We specifically assessed the relationship between intra-cranial anomalies and extra-cranial anomalies on neurodevelopmental outcome for fetuses diagnosed with a PFA on fetal MRI. **Methods**: Cases of Dandy-Walker malformation (DWM), vermian hypogenesis/hypoplasia (VH), and mega cisterna magna (MCM) were identified through the Fetal Care Center of Cincinnati between January 2004 and December 2010. Imaging studies were reviewed for additional abnormal findings. Parental interview and retrospective chart review were used to assess neurodevelopmental outcome. Statistical analysis was used to determine the impact of associated intracranial malformations including separate analysis of brainstem malformations and extra-cranial malformations on long term outcomes. **Results**: PFAs were identified in 59 fetuses; 9 with DWM, 36 with VH, and 14 with MCM. Cases with isolated PFAs (14/59) had better outcomes than those with additional anomalies (p=0.00016). Isolated cases of MCM (n=7) were all neurodevelopmentally normal for age. Cases with additional intra-cranial anomalies were more likely to have neurodevelopmental deficits than those without intra-cranial anomalies (p=0.00085). The presence of extra-cranial anomalies increased the likelihood of developmental disability (OR=10.2, 95% CI, 2.6-40.3, p=0.0009). The identification of an abnormal brainstem (n=15)was universally associated with abnormal cognitive development (OR=12.97, 95% CI, 2.9-58.0, p=0.0008). **Conclusion**: Intra- and extra-cranial anomalies predicted abnormal neurodevelopmental outcome in this study. The prognosis was poor for individuals with an abnormal brainstem while those with isolated MCM had normal neurodevelopmental outcome.

231

Distal arthrogryposis type 5D: a new autosomal recessive syndrome. *M.J. McMillin[1,3], A.E. Beck[1,3], J. Pinner[4], S.G. Mehta[5], D.K. Grange[6], G.R. Gogola[7], J.T. Hecht[8], D.J. Harris[9], S. Jagadeesh[10], L. Garavelli[11], D.L. Earl[1], M.J. Bamshad[1,2,3].* 1) Department of Pediatrics, University of Washington, Seattle, WA, USA; 2) Department of Genome Sciences, University of Washington, Seattle, WA, USA; 3) Department of Genetic Medicine, Seattle Children's Hospital, Seattle, WA, USA; 4) Department of Molecular and Clinical Genetics, Royal Prince Alfred Hospital, Camperdown, NSW, Australia; 5) East Anglian Medical Genetics Service, Addenbrooke's Hospital, Cambridge, UK; 6) Department of Pediatrics, Washington University School of Medicine, St. Louis, MO, USA; 7) Shriners Hospitals for Children-Houston, Houston, TX, USA; 8) Department of Pediatrics, University of Texas Medical School, Houston, TX, USA; 9) Department of Genetics and Metabolism, Division of Genetics, Children's Hospital Boston, Boston, MA, USA; 10) Mediscan, Chennai, India; 11) Department of Pediatrics, Maria Nuova Hospital, Reggio Emilia, Italy.

The distal arthrogryposis (DA) syndromes are a group of ten multiple congenital contracture syndromes characterized primarily by contractures of the hands and feet. In 1996, we revised the classification of DA syndromes and proposed strict diagnostic criteria to delineate discrete DA conditions and facilitate gene discovery. To this end, mutations that cause different types of DA have been identified in six genes, each of which encodes a component of the contractile apparatus of skeletal myofibers. The most phenotypically heterogeneous of the DA syndromes is DA with ophthalmoplegia, or DA5, and to date causal mutations have been found in less than ten percent of DA5 families. Over the last several years we have, based on linkage results and screening of candidate genes, tentatively divided DA5 into (1) DA5A caused by mutations in *MYH2*; (2) DA5B caused by mutations in *MYH13*; and (3) DA5C, which maps to chromosome 11. We now report the phenotype of a subset of DA5 cases, which we designate DA5D, that appears to segregate in an autosomal recessive pattern with clinical characteristics that distinguish this subset from other DA5 types. Specifically, we describe the phenotypic features of 17 affected individuals in 9 families. Affected individuals in two of these families are the result of consanguineous matings. In four families, multiple affected offspring were born to unaffected parents. In the three remaining families, a single affected individual was born to unaffected parents. The major clinical features of this new syndrome include a distinctive facial appearance with ptosis, arched eyebrows, a bulbous nasal tip, and a small jaw. The limb contractures include camptodactyly in 16/17 (94%), clubfoot or vertical talus in 16/17 (94%), fixed extension of the knees in 6/17 (35%), and a short neck with mild cervical pterygium in 11/17 (65%). None of the affected individuals had ophthalmoplegia or pulmonary disease, both of which are frequent causes of morbidity in DA5. Collectively, the clinical features, natural history, and mode of inheritance of DA5D clearly distinguish it from other forms of DA. This result has important consequences both for clinical care of individuals with arthrogryposis syndromes and for efforts to find genes underlying DA syndromes.

232

A frameshift mutation in *MYH3*, encoding embryonic myosin heavy chain, causes clubfoot. *A.E. Beck[1,2], K.M. Bofferding[1], H.I. Gildersleeve[1], M.J. McMillin[1,2], J.C. Carey[3], M.J. Bamshad[1,2,4].* 1) Pediatrics/Genetic Med, Univ Washington Sch Med, Seattle, WA, USA; 2) Seattle Children's Hospital, Seattle, WA, USA; 3) Pediatrics, University of Utah, Salt Lake City, UT, USA; 4) Dept Genome Sciences, Univ Washington, Seattle, WA, USA.

Congenital contractures such as clubfoot affect ~1 of every 200-500 live births in the United States and are caused by a combination of genetic and environmental risk factors. The distal arthrogryposis (DA) syndromes are a group of ten autosomal dominant disorders characterized by congenital contractures of the hands and feet, most notably clubfoot and camptodactyly. DA syndromes can be caused by mutations in at least nine genes that encode proteins of the skeletal muscle contractile complex including: troponin I, troponin T, tropomyosin, myosin binding proteins and several myosin heavy chains. Mutations in *MYH3* explain >90% of DA2A (Freeman-Sheldon syndrome), ~1/3 of DA2B (Sheldon-Hall syndrome), and a few DA1 cases. To date, all of the causative *MYH3* mutations reported are either missense or in-frame deletion mutations. To further understand the relationship between *MYH3* mutations and the pattern of contractures observed in different DA syndromes, we sequenced *MYH3* in 94 families with multiple different DA syndromes. In a large, multiplex family with DA1, we discovered a two base pair duplication, c.4438_4439dupAC, in exon 31 of *MYH3* predicted to cause a frameshift, p.Glu1481Leufs*6. This mutation is predicted to lead to premature truncation removing >50% of the tail domain of the embryonic myosin heavy chain protein. The c.4438_4439dupAC mutation segregated with 14 of 14 affected individuals in three generations and was not found in nearly 400 control chromosomes. All affected individuals had contractures of the feet including clubfoot (n=8) or metatarsus adductus (n=5), and one individual had a congenital hip dislocation with normal feet. However, 13/14 affected individuals had either normal hands or only mildly hypoplastic flexion creases. The results suggest that (1) frameshift mutations in *MYH3* may cause a less severe pattern of congenital contractures than is typically observed in DA2A or DA2B; (2) individuals with *MYH3* mutations might easily be misdiagnosed with idiopathic clubfoot; and (3) variants in *MYH3* might directly influence risk for idiopathic clubfoot.

233

***VAX1* mutation associated with microphthalmia, corpus callosum agenesis and orofacial clefting - the first description of a *VAX1* phenotype in humans.** *A.M. Slavotinek[1], R. Chao[1], M. Yahyavi[1], H. Abouzeid[2], T. Bardakjian[3], A. Schneider[3], E.H. Sherr[4], M. Youseff[5], G. Lemke[6], D.F. Schorderet[7].* 1) Dept Pediatrics, U585P, Univ California, San Francisco, San Francisco, CA; 2) Jules-Gonin Eye Hospital, Lausanne, Switzerland; 3) Clinical Genetics, Albert Einstein Med. Center, Philadelphia, PA; 4) Dept. Neurology, UCSF, San Francisco, CA; 5) Dept. Pediatrics, University of Alexandria, Egypt; 6) Molecular Neurobiology, The Salk Institute, La Jolla, CA; 7) Institut de Recherche en Ophtalmologie, Sion, Switzerland.

The ventral anterior homeobox genes *Vax1* and *Vax2* are implicated in eye morphogenesis in mice and zebrafish. During embryogenesis, *Vax1* is expressed in the murine optic nerve and chiasm, choroid fissure and rostral diencephalon, whereas *Vax2* is expressed in the retina. *Vax1* null mice have moderately severe colobomas, cleft palate and corpus callosum defects and *Vax2* null mutants have milder colobomas. *Vax1* and *Vax2* are part of a signaling pathway and are induced by *Shh*; they ventralize the eye and inhibit retinal development by repressing *Pax6* transcription and maintaining *Pax2* expression. We sequenced the coding exons of *VAX1* and *VAX2* in 70 patients with anophthalmia/microphthalmia (A/M). In *VAX1*, we observed homozygosity for two adjacent nucleotide substitutions, c.453G>A and c.454C>A that predict p.Arg152Ser in isoform 1, in an Egyptian patient with bilateral microphthalmia, cleft lip and palate, small optic nerves, hippocampal malformations, corpus callosum agenesis and absence of the pineal gland. This mutation affects an invariant residue in the homeodomain and was not present in 96 Egyptian controls; the clinically normal, consanguineous parents were heterozygotes. We hypothesized that p.Arg152Ser would alter *Pax6* expression, as corpus callosum defects have been seen in patients with *PAX6* mutations (Abouzeid et al. 2009). The early retinal expression of *Pax6* is controlled by the (-enhancer located in a region with binding sites for *Vax1* and *Pax6* autostimulation (Kammandel et al. 1999). Wildtype *Vax1* can repress transcription driven by the *Pax6* enhancer 3 to 4 fold in-vitro (Mui et al. 2005). A mutant p.Arg152Ser *VAX1* construct retained some ability to repress *Pax6* using the same system and thus it is unclear if altered *Pax6* is the only mechanism for pathogenesis. We found only one other exonic *VAX1* alteration: c.945C>T, a silent substitution not predicted to alter splicing, in a patient with bilateral anophthalmia and cleft lip and palate. In *VAX2*, we identified four coding SNPs: c.75C>A, c.102C>G, c.793C>G and c.857A>G, all with similar allele frequencies to controls. We conclude that the clinical findings associated with *VAX1* mutations can be similar to the murine phenotype seen with loss of *Vax1* function. This is the first description of a *VAX1* mutation and establishes *VAX1* as a new causative gene for A/M in humans. References: Abouzeid et al., Mol Vis 2009;15:2074. Kammandel et al., Dev Biol 1999;205:79. Mui et al., Genes Dev 2005;19:1249.

234

Expanding the Clinical Spectrum Associated with Defects in CNTNAP2 and NRXN1. *C. Zweier[1], A. Gregor[1], B. Albrecht[2], I. Bader[3], E.K. Bijlsma[4], A.B. Ekici[1], H. Engels[5], K. Hackmann[6], D. Horn[7], J. Hoyer[1], J. Klapecki[8], J. Kohlhase[9], I. Maystadt[10], S. Nagl[11], E. Prott[2], S. Tinschert[6], R. Ullmann[12], E. Wohlleber[5], G. Woods[13], A. Rauch[14], A. Reis[1].* 1) Institute of Human Genetics, University Erlangen-Nuremberg, Erlangen, Germany; 2) Institut für Humangenetik, Universitätsklinikum, Universität Duisburg-Essen, Essen, Germany; 3) Department of Medical Genetics, Kinderzentrum, Munich, Germany; 4) Department of Clinical Genetics, Leiden University Medical Centre, Leiden, The Netherlands; 5) Institute of Human Genetics, Rheinische Friedrich-Wilhelms-University, Bonn, Germany; 6) Institut für Klinische Genetik, Medizinische Fakultät Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany; 7) Institute of Medical Genetics and Human Genetics, Charité - Universitätsmedizin Berlin, Berlin, Germany; 8) Department of Medical Genetics, Institute of Mother and Child, Warsaw, Poland; 9) Center for Human Genetics, Freiburg, Germany; 10) Centre de Genetique Humaine, Institut de Pathologie et de Genetique, Gosselies (Charleroi), Belgium; 11) Synlab Medizinisches Versorgungszentrum Humane Genetik Munich GmbH, Munich, Germany; 12) Department of Human Molecular Genetics, Max Planck Institute for Molecular Genetics, Berlin, Germany; 13) Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Addenbrooke's Hospital, Cambridge, UK; 14) Institute of Medical Genetics, University of Zurich, Zurich-Schwerzenbach, Switzerland.

Heterozygous copy-number and missense variants in CNTNAP2 and NRXN1 have repeatedly been associated with a wide spectrum of neuropsychiatric disorders such as developmental language delay and autism spectrum disorders, epilepsy and schizophrenia. Recently, homozygous or compound heterozygous defects in either gene were reported as causative for severe intellectual disability disorders. By molecular karyotyping and mutational screening of CNTNAP2 and NRXN1 in a group of severely intellectually disabled patients we identified a heterozygous deletion in NRXN1 in one patient and heterozygous splice-site, frameshift and stop mutations in CNTNAP2 in four patients, respectively. Neither in these patients nor in eight further patients with heterozygous deletions within NRXN1 or CNTNAP2 we could identify a defect on the second allele. One deletion in NRXN1 and one deletion in CNTNAP2 occurred de novo, in another family the deletion was also identified in the mother who had learning difficulties, and in all other tested families one parent was shown to be healthy carrier of the respective deletion or mutation. We therefore show that heterozygous defects in NRXN1 or CNTNAP2 can also be seen in association with severe intellectual disability. These results expand the spectrum of phenotypic severity in patients with heterozygous defects in either gene. The large variability between severely affected patients and mildly affected or asymptomatic carrier parents might suggest the presence of major modifiers, not necessarily located in the same gene.

235

The Ethical, Legal and Social Implications of the Convergence of Cell-Free Fetal DNA (cffDNA) with Genomic Sequencing. *J.S. King[1], M.E. Nunes[2], S.E. Kelly[3], L.C. Sayres[4], M. Allyse[4], M.K. Cho[4].* 1) University of California, Hastings College of Law, San Francisco, CA; 2) Pediatrics and Medical Genetics, Kaiser Permanente, San Diego, CA; 3) ESRC Centre for Genomics and Society (Egenis), University of Exeter, Exeter, UK; 4) Stanford Center for Biomedical Ethics, Stanford University, Palo Alto, CA.

The discovery of cell-free fetal DNA (cffDNA) in maternal blood has the potential to dramatically change the current practice of prenatal diagnosis. Reliable genetic tests that offer non-invasive prenatal diagnosis (NIPD) as early as seven weeks gestation have recently been developed. In December 2010, two independent laboratories have demonstrated that a full genome-wide analysis of the fetus could be performed from a sample of maternal blood, making fetal diagnostic testing possible for any genetic condition for which the gene is known. The convergence of cffDNA testing with low cost genomic sequencing will enable prospective parents to have inexpensive access to a wide range of genetic information about their fetus very early in a pregnancy. While a major medical advance, cffDNA testing raises significant challenges. Initially, it will strain existing consent, timing, and counseling protocols. It has the potential to expand the number of routine prenatally diagnosed conditions from 5-10 to over 500. Because cffDNA testing can be performed at or before the first prenatal visit, many providers may have difficulty explaining the expanded range of risks and benefits to patients prior to testing. Many patients may not appreciate the difference between a blood test screening their genome from one testing their fetus', which could have significant consequences. Introducing a vastly expanded NIPD panel will also tax counseling resources recently re-geared for prenatal integrated screening. In the longer term, widespread cffDNA testing also raises significant regulatory questions: 1) how should these tests enter the market?; 2) should all tests be provided by a physician?; and 3) should all tests be available to all individuals at all times? Since states may also wish to limit access to certain tests, cffDNA testing may challenge our current understandings of constitutionally protected reproductive autonomy. While our goal is to examine a number of ethical, legal and social issues raised by cffDNA testing, we also propose that it should be offered through a two-step informed consent process and performed via NIPD panels of validated conditions with similar salient characteristics. The validation process, involving input from a range of stakeholders, would be modeled on that used for introducing new tests to newborn screening protocols. Finally, we propose significant stakeholder involvement in discussions surrounding regulation of the technology.

236

Ethical Approaches to Genotype-Driven Research Recruitment. *L.M. Beskow.* Duke Institute for Genome Sciences & Policy, Durham, NC.

Purpose: Translating advances in genomics into knowledge about the function of the human genome represents an unprecedented opportunity to understand health and disease. Genotype-driven research recruitment is a potentially powerful tool for facilitating this translation. With this approach, investigators use an existing study population for which genetic analyses have been conducted to identify individuals with a gene variant of interest. They then recontact those individuals about participation in further research involving in-depth phenotyping. This kind of recruitment, however, presents ethical challenges. Concerns about the use and disclosure of genetic information—more commonly associated with research participation—are shifted to the recruitment phase when information that is generated in one study is used as the basis for identifying and recontacting participants about further research. **Methods:** To inform policy development on these issues, we conducted qualitative interviews with 78 participants in a diverse set of studies where genotype-driven recruitment occurred. We also conducted a survey of U.S. IRB chairs (n=201) to assess their perspectives. Finally, we convened a multidisciplinary workshop to review our data and develop recommendations. **Results:** More than 80% of research participants expressed favorable views concerning recontact about taking part in additional research compared to 37% of IRB chairs. About two-thirds of research participants favored return of individual genetic results from the first study in the context of genotype-driven recruitment, compared to 42% of IRB chairs. Both groups were less positive about disclosure of results with uncertain clinical validity, but both had fewer concerns about results with uncertain clinical utility. These and related findings were discussed at the stakeholder workshop as the foundation for addressing the following questions: When is it acceptable to recontact participants for recruitment into additional genetic research? When is it acceptable to disclose individual genetic results from the first study during the process of genotype-driven recruitment for further research? What approaches to genotype-driven recruitment are optimal to respect and protect participants, minimize ethical dilemmas, and facilitate beneficial research? The recommendations from this workshop provide important guidance for researchers conducting genotype-driven recruitment and for IRBs reviewing such studies.

237

The attitudes and opinions of genetic researchers and clinicians towards direct-to-consumer genetic testing. *P.M. Kaushik, S.M. O'Neill, M.G. Hayes, C.A. Wicklund.* Northwestern University, Center for Genetic Medicine, Chicago, IL.

Currently consumers are able to purchase genetic tests directly from commercial companies (DTC genetic tests). Several of the tests available are based on genetic variants identified by genome-wide association studies for common complex conditions, but the clinical validity of these tests is questionable. Though several studies have assessed the public's attitudes towards DTC testing to date, no studies have been carried out to assess researchers' attitudes towards DTC genetic testing, and limited studies have been carried out to assess clinicians' attitudes. Therefore the purpose of this study was to assess the attitudes of genetic researchers and clinicians towards DTC genetic testing. Electronic surveys were sent to all members of the American Society of Human Genetics (~7000). The survey assessed attitudes towards the different types of DTC genetic tests and their perceived value and importance. The overall response rate was 23.3% (n=1631). The majority of respondents (96.1%) were aware of DTC genetic tests. 1167 respondents (74.7%) felt that DTC genetic testing should not be available for all purposes. DTC genetic testing for personal ancestry was the only group of tests that a little over the majority agreed should be available (59.6%). Of the 1539 participants who answered this section, 65% disagreed or strongly disagreed that DTC genetic tests are valuable for consumer's personal knowledge, 66.2% disagreed or strongly disagreed that DTC genetic tests are valuable for clinical knowledge, and 64.9% disagreed or strongly disagreed that it will help improve risk assessment for common complex diseases. About half (50.9%) agreed or strongly agreed that DTC genetic testing is okay as long as the consumer understands the benefits, limitations and results. The majority (84.9%) disagreed or strongly disagreed that they themselves are indifferent to DTC genetic testing. Clinical experience, involvement in GWAS, gender and age were shown to be independent predictors of the participants' opinions. Individuals without clinical experience, who had been involved in GWAS, who were male, and who were younger were shown to be less cautious about DTC genetic testing. This study suggests that while both genetic researchers and clinicians largely disagree with the use of DTC genetic testing, clinicians are more cautious about this form of testing than researchers.

238

Feedback of individual genetic results to research participants: in favor of a qualified disclosure policy. *A.L. Bredenoord[1], N.C. Onland-Moret[2], J.J.M. van Delden[1].* 1) Medical Ethics, University Medical Center Utrecht, Utrecht, Netherlands; 2) Epidemiology, University Medical Center Utrecht, Utrecht, Netherlands.

In recent years, a debate evolved regarding the question whether researchers have a duty to return individual genetic research results to research participants. This unsettled debate has rapidly gained in urgency in view of the emergence of biobanks and the advances in next-generation sequencing technology, which has the potential to generate unequalled amounts of genetic data. This implies that the generation of many known and unknown genetic variants in individual participants of genetics/genomics research as intentionally or collaterally obtained by-products is unavoidable, including genetic variants that were outside the focus of the study. In both the scientific debate and in international guidelines, the extreme positions of full disclosure and no disclosure whatsoever are seldom defended. A duty to warn when this may save the life of a research participant is apparently recognized widely. As no disclosure is unethical because it fails to adhere to the so-called rule of rescue and full disclosure nonsensical (at best) as it could imply disclosure of all raw sequencing data, any disclosure policy in between means we have to consider how and by whom a selection should be made about which results are eligible for disclosure. In this paper, a qualified disclosure policy is proposed. This policy contains a standard default package, possibly supplemented with (one or more of) three additional packages. Whereas the default package, containing life-saving information of immediate clinical utility, should be offered routinely and mandatory to all research participants, offering (one of) the three additional packages is context-specific. Such a qualified disclosure policy in our opinion best balances the potential benefits of disclosure with the potential risks for research participants and the harms of unduly hindering biomedical research. We appeal to the genetics community to make a joint effort to further refine the packages and set thresholds for result selection.

239

Using legislation to facilitate disclosure of relevant genetic information to genetic relatives: the Australian experience. *M.FA. Otlowski.* Law, University of Tasmania, Hobart, Tasmania, Australia.

The familial nature of genetic information means that information about one family member may be of significant interest to others eg if a woman tests positive to the BRCA1 or BRCA2 gene mutations, this would be of great relevance to any daughters, indicating heightened risk of breast and ovarian cancer. Awareness of this risk enables strategies to be put in place for regular screening or even prophylactic measures. Whilst generally, family members can be encouraged to share relevant information with others for whom it is relevant, this is not always the case. This paper outlines groundbreaking legislative developments in Australia which now create a lawful pathway for health practitioners to make disclosure of genetic information to genetic relatives notwithstanding that the index patient has refused to consent to such disclosure. It traces the background of this reform, and the recommendations made by the Australian Law Reform Commission/ Australian Health Ethics Committee Inquiry into the protection of human genetic information which had identified legal obstacles to disclosure due to privacy and confidentiality obligations owed to the index patient. The Inquiry's recommendation for disclosure to be permitted in circumstances where there is reasonable belief that doing so is necessary to lessen or prevent a serious threat to the life, health or safety of their genetic relatives, has since been implemented through the new s95AA: Privacy Legislation Amendment Act 2006 (Cth) which governs the use and disclosure of personal information in the private sector. Enabling guidelines have been developed by the National Health and Medical Research Council and endorsed by the Federal Privacy Commissioner: Use and Disclosure of Genetic Information to a Patient's Genetic Relative Under s 95AA of the Privacy Act 1988: Guidelines for Health Practitioners in the Private Sector (2009). The paper examines the circumstances in which disclosure will be permitted and the protocols established under these guidelines. Notably, these changes do not legally oblige health practitioners to make disclosure but create a lawful means for them to do so, aiming to encourage this as best practice in appropriate circumstances. This raises some interesting issues regarding the interplay of law and ethics: if it is ethical to disclose, but not legally mandated, what recourse do persons have to whom disclosure had not been made and for whom this has had adverse health consequences?

240

Legal Provisions Related to Genetic Testing: The German Act on Genetic Testing (GenDG; 2010). *H. Tonnies.* Administrative Office of the Commission on Genetic Testing (GEKO), Robert Koch-Institute, Berlin, Germany.

The German Act on Genetic Testing (Genetic Diagnosis Act, GenDG) became effective in 2010. The purpose of this new Act is to determine the requirements for genetic examinations and genetic analyses conducted in the framework of genetic examinations and to prevent any discrimination and disadvantage based upon genetic characteristics, especially in regard to the duty of the state to protect human dignity and to ensure the individual right to self-determination via sufficient information. The right of individuals to determine the use of their own data is one of the Act's central principles. This includes the right to know as well as the right not to know own genetic data. This Act applies to genetic examinations and genetic analyses conducted within the framework of genetic examinations involving born natural persons as well as embryos and foetuses during pregnancy and the handling of genetic data and genetic samples gained thereby for medical purposes, for purposes of determining descent as well as in the insurance and employment sectors. The GenDG determines when and under what circumstances genetic testing is allowed, who may carry this out, requirements for a good genetic testing practice and who is allowed to use the sensitive results. This Act does not apply to genetic examinations and genetic analyses or the handling of genetic samples or genetic data conducted (1) for research purposes, (2) on the basis of specific applicable regulations (e.g. relating to criminal procedure the Infection Protection Act). At the time the law came into force, an independent and interdisciplinary commission was established by the German Ministry of Health in order to give final guidelines on certain aspects of the law as professional and qualitative requirements for the performance of genetic tests in the medical field and genetic tests for ancestry disclosure purposes in reference to the generally accepted status of science and technology. The Robert Koch-Institute is the hosting institution for this Commission, which is made up of 18 representatives. The members are drawn from different medical, legal, ethical, scientific and social disciplines. The Commission has set up several working groups, each of which deals with one of the specific directives detailed in § 23 (2) of the GenDG.

241

Science regulation and the complexities of genomics. *J.M. Siqueiros[1], A. Arellano[1], E. Schwartz[2], G. Saruwatari[1], P.F. Oliva[1].* 1) ELSI, National Institute of Genomic Medicine, Periférico Sur No. 4809, P.ostal Code 14610, Mexico City, D. F., Mexico; 2) EGENIS, Exeter University, ESRC Genomics Network. Egenis Office University of Exeter Byrne House St. German's Road Exeter Devon EX4 4PJ.

In 2008, an amendment regarding what has been called Genomic Sovereignty was discussed by the Mexican Senate and it was added to the National Health Act (NHA) in that same year. One of the purposes of the amendment was to regulate the research on the Mexican genome and to shield it from `external' research agents by the regulation of the access to human biological samples, all of this under the premise that the Mexican genome had to be studied by and for Mexicans (primarily). As the amendment is written, the privileges of all research and control over the research that could be done in different institutions about the genome are allocated at the INMEGEN. Since the creation of the INMEGEN and the amendment to the NHA, little has improved in enabling the development of the field in terms of genomic medicine done by and for Mexicans. The idea that we are presenting here is that policymakers didn't have a very clear idea of how science is done, but most of all they didn't have the right picture of the complexities of the genome. Genomics was understood as a new version of classical genetics and that the genome can be defined and delimited by one structure-one function (not only as in the Central Dogma, but as a `monolithic object').As part of a biological system, the genome has many different expressions according to the variety of contexts, i. e., it cannot be defined out of its context, and each of its possible contexts is a possible research area about the genome. We propose that due to the complexity of the genome and its contexts, regulation and normativity shouldn't be centered on the access to the object (Mexican genome) but on the scientific policies that, on ethical groundings, can help develop the field further. That is, a regulation for the creation of a research system on genomics based on openness and governance instead of pretending to protect an object such as the Mexican genome through the limitation of access to human biological samples. We also propose to open a new debate agenda on three priority areas: 1) How to make medical genomics a public good, translating it in concrete actions, in both regulation and scientific practice, 2) how to foster cooperation and knowledge exchange in order to build trust and a robust research infrastructure in the service of public interest, 3) develop an autonomous and inclusive ELSI analysis, in order to bring emerging matters of concern to the public light and gradually strengthen democratic engagement.

242

Long-Term Follow-Up Data Collection After Newborn Screening: Development of a Consensus Core Data Set. *S.A. Berry[1], A.M. Brower[2] for the Jt Committee, NICHD-NBSTRN Clin Ctrs Wkgrp & HRSA-NCC/RC Long-Term F/U Wkgrp.* 1) Dept Pediatrics, Univ Minnesota, Minneapolis, MN; 2) American College of Medical Genetics, Bethesda MD.

Background: Newborn screening (NBS) has emerged as a public health imperative. Screening technologies and treatment strategies continue to advance with new conditions added to the recommended NBS panel. Despite these achievements, plans are lacking for long-term care and treatment strategies for individually rare NBS conditions; only national/international data will yield sufficient numbers to determine patient outcomes and improve scientific understanding. To provide the clinical history necessary for translational research and program improvements for improving health outcomes and to identify populations robust enough for research trials needed to generate novel treatments, a uniform minimum data set with accompanying information collection, management and analysis tools is needed. Methods: A workgroup comprised of content experts from lab-based specialties, departments of health, and clinical activities was convened. National meetings were held to define a uniform data set common to all NBS conditions; additional meetings were held with content experts to define condition-specific elements and to specify which elements of the data set are of interest for public health activities. Workgroup members also assisted with the development of tools and infrastructure for the NBS Translational Research Network (NBSTRN) including model consent forms, pilots of new disorders, and a virtual repository of specimens. Results: A uniform minimum data set that comprises ~80%; of desired information in common across all NBS disorders has been defined by consensus; the input of public health in definition of the minimum data set has been essential. Additional workgroups to define condition-specific elements for metabolic conditions, endocrine conditions, and hemoglobinopathies detected by NBS are defining condition-specific elements for initial encounter and follow-up measures for each NBS condition. New groups have been convened to define elements for follow-up of lysosomal storage diseases, severe combined immunodeficiency, hearing loss, and cystic fibrosis. Conclusions: A national community of specialty providers residing in public health, clinical centers and academic research centers can reach consensus regarding priorities for data collection for long-term follow-up. This effort lays an effective foundation for a uniform minimum consensus data set to ascertain the natural history of NBS disorders for both public health- and research-related activities.

243

The Genetic History of Native Americans. *D. Reich[1], N. Patterson[2], D. Campbell[3], A. Tandon[1,2], S. Mazieres[3,4], N. Ray[5], C.M. Bravi[3,6], M.-C. Bortolini[7], F. Salzano[7], M.L. Letzl-Erler[8], V. Acuña-Alonzo[9], S. Canizales-Quinteros[10,11], T. Tusié-Luna[10], J. Molina[12], A. Carracedo[13], C. Gallo[14], G. Alkorta-Aranburu[15], D. Labuda[16], R. Barrantes[17], L. Excoffier[18], G. Bedoya[19], F. Rothhammer[20], W. Klitz[21], J. Kidd[22], K. Kidd[22], A. Di Rienzo[15], N. Freimer[23], A. Price[2,24], A. Ruiz-Linares[3].* 1) Genetics, Harvard University, Boston, MA; 2) Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 3) Department of Genetics, Evolution and Environment. University College London, UK; 4) Anthropologie Bioculturelle, UMR 6578, Université de la Méditerranée/CNRS/EFS, Marseille, France; 5) EnviroSPACE Laboratory, Climate Change and Climatic Impacts, Institute for Environmental Sciences, University of Geneva, Carouge, Switzerland; 6) IMBICE, La Plata, Argentina; 7) 7 Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; 8) Departamento de Genética, Universidade Federal do Paraná, Curitiba Brazil; 9) National Institute of Anthropology and History, Mexico City, México; 10) Unit of Molecular Biology and Genomic Medicine, Instituto Nacional de Ciencias Médicas y Nutrición, México City, México; 11) Department of Biology, Facultad de Química, Universidad Nacional Autónoma de México, Mexico City, México; 12) Centro de Investigaciones Biomédicas de Guatemala, Ciudad de Guatemala, Guatemala; 13) Unidade de Xenética, Instituto de Medicina Legal, Facultade de Medicina, Universidade de Santiago de Compostela, Santiago de Compostela, Galicia, Spain; 14) 14 Laboratorios de Investigación y Desarrollo, Facultad de Ciencias y Filosofía, Universidad Peruana Cayetano Heredia, Lima, Perú; 15) Department of Human Genetics, University of Chicago, Chicago, USA; 16) Département de Pédiatrie, Centre de Recherche du CHU Sainte-Justine, Université de Montréal, Montréal, Quebec, Canada; 17) Escuela de Biología, Universidad de Costa Rica, San José, Costa Rica; 18) Computational and Molecular Population Genetics Lab, Institute of Ecology and Evolution, University of Bern, Switzerland; 19) Laboratorio de Genética Molecular, Universidad de Antioquia, Medellín, Colombia,; 20) Facultad de Medicina, Universidad de Chile, Santiago and Instituto de Alta Investigación, Universidad de Tarapacá, Arica, Chile; 21) School of Public Health, University of California Berkeley, and Public Health Institute, Oakland, California, USA; 22) Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA; 23) Center for Neurobehavioral Genetics, University of California Los Angeles, Los Angeles, California, USA; 24) Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA.

We assembled genetic data on more than 450,000 SNPs genotyped in 55 Native American and 16 Siberian populations. Our data are consistent with at least three pre-Colombian genetic interchanges with Eurasians: the initial peopling of the Americas; a migration involving relatives of Na-Dene speakers; and a migration involving ancestors of the Inuit. We also provide evidence of gene exchange between diverse Native North Americans and relatives of present-day Inuit. The genetic relationships among most Native Americans are well fit by a model of southward expansion, facilitated by the coast and with little gene flow after population separation. A striking exception is a Chibchan population cluster around Panama, that arose from a 2,400-9,300 year old mixture of a group related to eastern South Americans and an ancient lineage that branched prior to the separation of the ancestors of present-day Mexicans and South Americans.

244

A gene involved in metabolism modulates natural variation in sleep duration: from genome wide association studies to function in *Drosophila*. *K. Allebrandt[1], N. Amin[2], B. Müller-Myhsok[3], T. Esko[4], M. Teder-Laving[4], R. V.D.M. Azevedo[5], C. Hayward[6], J. van Mill[7], N. Vogelzangs[7], E. Green[5], S. Melville[8], P. Lichtner[9], H. Wichmann[10], B. Oostra[11], C. Janssens[2], H. Campbell[11], J. Wilson[11], A. Hicks[8], P.P. Pramstaller[8], Z. Dogas[12], I. Rudan[11,12], M. Merrow[13], B. Penninx[7], C. Kyriacou[5], A. Metspalu[4], C. van Duijn[2,14], T. Meitinger[9], T. Roenneberg[1].* 1) Chronobiology, Medical Psychology, Munich, Bayern, Germany; 2) Genetic epidemiology unit, Department of Epidemiology and Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands; 3) Max-Planck-Institute of Psychiatry, Munich, Germany; 4) Estonian Genome Center and Institute of Molecular and Cell Biology of University of Tartu, Estonian Biocentre, Tartu, Estonia; 5) Department of Genetics, University of Leicester, Leicester, UK; 6) Medical Research Council, Human Genetics Unit, IGMM, Edinburgh, Scotland; 7) VU University Medical Center Amsterdam, Amsterdam, The Netherlands; 8) Institute of Genetic Medicine, European Academy of Bolzano, Bozen, Italy - affiliated Institute of the University of Lübeck, Germany; 9) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany; 10) Institute of Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany - Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-University, Munich, Germany, and Klinikum Gro) hadern, Munich Germany; 11) Centre for Population Health Sciences, University of Edinburgh, Community Health Sciences, Edinburgh, Scotland; 12) Croatian Centre for Global Health, University of Split Medical School, Split, Croatia; 13) Department of Chronobiology, University of Groningen, Haren, The Netherlands; 14) Centre of Medical Systems Biology, Netherlands Genomics Initiative (NGI), the Netherlands.

Extremes of sleep duration and timing have been associated with adverse health outcomes that characterise the metabolic syndrome. Although sleep duration can be influenced by photoperiod (season) and phase of entrainment (chronotype), human familial sleep disorders indicate that there is a strong genetic modulation of sleep. We thus investigated 7 European populations in relation to their sleep duration habits and genome-wide variability. Meta-analysing the independent genome-wide association results, we identified a variant (P = 3.99 x 10^{-8}) in the *ABCC9* gene that explains _ 5% of the variation in sleep duration. We found supportive evidence for this association in a subgroup of an independent de novo replication cohort, which we selected based on season of entry and chronotype. To investigate the functional relevance of our findings, we knocked down a homologue of the gene in *Drosophila*, which resulted in a night-sleep duration reduction of 3 h. Our study shows that *ABCC9* modulates epidemiological variation in human sleep duration, which is also influenced by inter-individual differences in sleep timing and seasonality (seasonal differences in entrainment of the biological clock). Therefore, scanning only for main effects on sleep duration, one might miss important genetic variants specific to subgroups of the population. *ABCC9* is involved with energy homeostasis, and the susceptibility to overweight and cardiovascular disease, which correlate with sleep duration. The relation of this gene to metabolism and disease indicates a possible common mechanism for the regulation of these phenotypes and sleep duration.

245

Non-Mendelian inheritance in split-hand/foot malformation associated with CNVs on chromosome 17p. *E. Klopocki[1,2], S. Lohan[1,2], S.C. Doelken[1], S. Stricker[2], C.W. Ockeloen[3], R. Soares Thiele de Aguiar[4], K. Lezirovitz[4,5], R.C. Mingroni Netto[4], A. Jamsheer[6,7], H. Shah[8], I. Kurth[9], R. Habenicht[10], M. Hempel[11], M. Warman[12], K. Devriendt[13], U. Kordass[14], A. Rajab[15], O. Mäkitie[16], M. Naveed[17], U. Radhakrishna[18], S.E. Antonarakis[18,19], D. Horn[1], S. Mundlos[1,2].* 1) Institute for Medical Genetics and Human Genetics, Charité - Universitätsmedizin Berlin, Berlin, Germany; 2) Max-Planck-Institute for Molecular Genetics, Berlin, Germany; 3) Department of Human Genetics, Radboud University Nijmegen Medical Centre, The Netherlands; 4) Centro de Estudos do Genoma Humano, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, Brazil; 5) Laboratório de Otorrinolaringologia/LIM32, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, Brazil; 6) Department of Medical Genetics, University of Medical Sciences, Poznan, Poland; 7) NZOZ Center for Medical Genetics GENESIS, Poznan, Poland; 8) Department of Orthopaedics, Paediatric Orthopaedic Service, Kasturba Medical College, Manipal, Karnataka, India; 9) Institute for Human Genetics, Universitätsklinikum Jena, Germany; 10) Kath. Kinderkrankenhaus Wilhemstift, Hamburg, Germany; 11) Helmholtz Zentrum München, Institut für Humangenetik, Neuherberg Germany; 12) Orthopaedic Research Laboratories, Department of Orthopaedic Surgery, The Howard Hughes Medical Institute, Children's Hospital, Boston, Massachusetts, USA; 13) Department of Medical Genetics, Leuven University Hospital, Leuven, Belgium; 14) Institute for Human Genetics, Universitätsmedizin Greifswald, Germany; 15) Department of Genetics, Directorate General of Health Affairs, Ministry of Health, Muscat, Sultanate of Oman; 16) Children's Hospital, Helsinki University Central Hospital, University of Helsinki, and Folkhälsan Institute of Genetics, Helsinki, Finland; 17) Center for Arab Genomic Studies (CAGS), Dubai, United Arab Emirates; 18) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 19) Division of Medical Genetics, Geneva University Hospitals, Geneva, Switzerland.

Split-hand/foot malformation (SHFM) also known as ectrodactyly is a congenital disorder characterized by malformations affecting the central rays of hands and/or feet. SHFM is extremely variable in its phenotypic expression between families, within families, and even within individuals, ranging from syndactyly and oligodactyly to monodactyly. SHFM can be associated with additional long bone deficiency in particular involving tibia and fibula which is referred to as SHFM with long bone deficiency (SHFLD). Although the underlying genetic cause has been identified in some cases a large number of SHFM/SHFLD cases remain unsolved. Reduced penetrance and variable expressivity hamper a precise interpretation of inheritance patterns and have made this condition less accessible for mapping approaches with subsequent disease gene discovery. Based on our previous results indicating that copy-number variations (CNVs) may cause congenital malformations we performed high-resolution array CGH in a SHFM/SHFLD cohort. Thereby we identified microduplications on chromosome 17p13.3 in 17 families including a large Brazilian family previously mapped to this region. The breakpoints were non-recurrent; duplication sizes varied from 69 to 594 kb. The smallest region of overlap corresponds to ~12 kb and encompasses a single gene, *BHLHA9*. Whole mount in situ hybridization revealed expression of this gene in the distal limb bud mesenchyme underlying the apical ectodermal ridge (AER) suggesting a role during limb development. Central AER defects have been shown to underlie ectrodactyly. Interestingly upon studying further family members by quantitative RT-PCR we identified a high degree of non-penetrance. In total, duplications were observed in 82 individuals of whom 42 were affected but also in 40 unaffected carriers. Furthermore, a clear sex-bias was observed that is more male (30/42) than female (12/42) affected. Together, these effects result in an unusual non-Mendelian inheritance pattern characterized by reduced penetrance, particularly in females. We postulate contribution of yet to be identified modifiers. In summary, we describe the association of a severe limb malformation with duplications encompassing a putative basic loop helix transcription factor. Our finding shows that rare CNVs can serve as a susceptibility factor for congenital disease, a mechanism which may explain increased recurrence risk in conditions otherwise considered to be sporadic.

246

17p13.3 Microduplications are associated with Split-Hand/Foot Malformation and Long Bone Deficiency (SHFLD). *C. Armour[1], D.E. Bulman[2,3], O. Jarinova[4], R.C. Rogers[5], K.B. Clarkson[5], B.R. DuPont[5], A. Dwivedi[5], F.O. Bartel[5], L. McDonell[2,6], C.E. Schwartz[5,7], K.M. Boycott[2,8], D.B. Everman[5], G.E. Graham[2,8].* 1) Clinical Genetics Unit, Queen's University/Kingston General Hospital, Kingston, ON, Canada; 2) University of Ottawa, 451 Smyth Rd. Ottawa, Ontario Canada K1H 8M5; 3) Ottawa Hospital Research Institute 501 Smyth Road Ottawa, Ontario K1H 8L6; 4) Ottawa Heart Institute 40 Ruskin St Ottawa, ON K1Y 4W7; 5) Greenwood Genetics Center 101 Gregor Mendel Circle, Greenwood, SC 29646; 6) CHEO Research Institute 401 Smyth Road, Rm. 139 Ottawa, Ontario Canada K1H 8L1; 7) Dept of Genetics and Biochemistry Clemson University,Clemson University Clemson, SC 29634-0318; 8) Genetics Clinic, Children's Hospital of Eastern Ontario 401 Smyth Road Ottawa, ON, K1H 8L1.

The Split-Hand/Foot Malformations (SHFMs) are a heterogeneous group of malformations in which the hand and/or foot findings may occur in isolation or with other anomalies. Split-Hand/Foot malformation with Long Bone Deficiency (SHFLD) is an autosomal dominantly inherited form of one of the SHFMs that manifests with variable expressivity and incomplete penetrance; the clinical presentation may range from isolated hypoplastic halluces, to classic SHFM, to monodactyly with tibial aplasia. While several chromosomal loci for this malformation (designated SHFLD1 [OMIM# 119100], SHFLD2 [OMIM# 610685], and SHFLD3 [OMIM #612576]) have been identified, the molecular basis and pathogenesis of most SHFLD cases remain unknown. SHFLD3 has been previously mapped to an 861 kb interval at 17p13.1-17p13.3 in a Brazilian family. Here we describe three unrelated kindreds in which SHFLD segregated with distinct but overlapping duplications in 17p13.3, the region previously linked to SHFLD3. In a large three generation family, the disorder was found to segregate with a 254 kb microduplication; a second microduplication of 527 kb was identified in an affected female and her unaffected mother, and a 430 kb microduplication versus microtriplication was identified in three affected members of a multi-generational family. These findings, along with previously published data, suggest that one locus responsible for this form of SHFLD is located within a 173 kb overlapping critical region, and that the copy gains are incompletely penetrant. This region contains exons 1-3 of ABR, BHLHA9, and a putative processed transcript AC016292.1. The phenotype resulting from this genomic rearrangement may arise directly from the duplication or alteration of one of these genes. Alternatively, it may arise through the direct duplication of a conserved regulatory element (CRE) located within the region, or by changing the interaction of a CRE with its target gene via another mechanism.

247

A Copy Number Variation Morbidity Map of Developmental Delay. *B.P. Coe[1], G.M. Cooper[1], S. Girirajan[1], J.A. Rosenfeld[2], T. Vu[1], C. Baker[1], C. Williams[3], H. Stalker[3], R. Hamid[4], V. Hannig[4], H. Abdel-Hamid[5], P. Bader[6], E. McCracken[7], D. Niyazov[8], K. Leppig[9], H. Thiese[9], M. Hummel[10], N. Alexander[10], J. Gorski[11], J. Kussmann[11], V. Shashi[12], K. Johnson[12], C. Rehder[13], B. Ballif[2], L.G. Shaffer[2], E.E. Eichler[1,14].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Signature Genomic Laboratories, LLC, Spokane, WA; 3) Department of Pediatrics, Division of Genetics, University of Florida, Gainesville, FL; 4) Vanderbilt University Medical Center, Nashville, TN; 5) Department of Pediatrics, Division of Child Neurology, University of Pittsburgh, Pittsburgh, PA; 6) Northeast Indiana Genetic Counseling Center, Ft. Wayne, IN; 7) Children's Hospital Pittsburgh, Pittsburgh, PA; 8) Ochsner Clinic, New Orleans, LA; 9) Group Health Cooperative, Seattle, WA; 10) West Virginia University, Morgantown, WV; 11) University of Missouri, Columbia, MO; 12) Departments of Pediatrics and Pathology, Duke University Medical Center, Durham, NC; 13) Clinical Molecular Diagnostic Laboratory, Duke University Health System, Durham, NC; 14) Howard Hughes Medical Institute, University of Washington, Seattle, WA.

Genetic screens of larger numbers of cases and controls are needed in order to discover pathogenic copy number variants (CNVs) and to refine the genes responsible for neurocognitive disease and developmental delay. Here, we compare the CNV landscape of 15,767 children with intellectual disability and developmental delay to 8,329 adult controls. We estimate that 14.2% of developmental delay and congenital birth defects are due to the presence of CNVs greater than 400 kbp in length (OR = 2.7, p = 5.86E-158). We show that the CNV burden is significantly greater for craniofacial anomalies and cardiovascular defects compared to epilepsy and autism, suggesting a differential impact of large CNVs on development. We identify 60 likely pathogenic regions enriched for CNVs within affected individuals; these include 45 known genomic disorders, CNVs of previously unknown significance or deemed benign, and 15 new candidate regions. We refine the smallest regions of overlap for 17q21.31 microdeletion syndrome, identify 940 candidate dosage-sensitive genes and associations—in some cases with specific subsets of developmental defects. Additionally, we ascertained the inheritance status for 2058 CNVs of uncertain clinical significance and observed a linear correlation between increasing CNV size and the proportion of calls classified as de novo (r2 = 0.97). Such genome-wide CNV morbidity maps of developmental delay combined with the specificity of exome sequencing will be critical for deciphering the genetic heterogeneity associated with these complex genetic diseases and their clinical impact.

248

Single gene copy number abnormalities in syndromic cardiovascular malformations. *J. Belmont[1,2,4], Q. Tian[2], J. Flores[1], M. Azamian[2], P.M. Boone[1], C. Shaw[1], S. Ware[3], J.R. Lupski[1,4], A. Ester[1], L. Patterson[4], S.W. Cheung[1], D. Penny[4], S. Lalani[1].* 1) Dept Molec & Human Gen, Baylor Col Med, Houston, TX; 2) Children's Nutrition Research Center, Baylor Col Med, Houston, TX; 3) Cincinnati Children's Hospital, Cincinnati, OH; 4) Department of Pediatrics, Baylor Col Med, Houston, TX.

Congenital cardiovascular malformations (CVM) are among the most common of all medically significant birth defects and are a leading contributor to infant mortality in the United States. Copy number variations (CNVs) resulting from instability of regional genomic architecture are an important cause of CVM. These genomic disorders most likely affect several contiguous genes, but often as exemplified by the TBX1 gene in 22q11del syndrome and ELN1 in the Williams syndrome, a single gene is thought to be the major factor in causing the cardiovascular developmental defect. We reviewed the phenotype data on 858 individuals with syndromic CVM from Texas Children's Hospital (TCH), who had 105k or 180k genome-wide oligonucleotide based array-comparative genomic hybridization analyses (aCGH) through the Baylor Cytogenetics Laboratory. The 180k customized array has exon by exon coverage of 1,700 genes, including 200 genes known to be important for cardiac patterning. The array additionally has coverage for 700 microRNAs. All echocardiograms in this cohort were reviewed through the TCH CardioIMS database and the cardiac lesions were categorized into eight major classifications. We identified several rare CNVs affecting single genes, involving one or several exons: These include DTWD2 loss observed with complex CVM phenotype, ANKRD11 loss with septal and atrioventricular septal defects, WDR27 gain seen with conotruncal malformation, and DPP6 loss with septal defects. These exonic changes were not observed in 5,868 individuals with intellectual disability, studied by the customized array. The 139 kb deletion corresponding to ANKRD11 gene on 16q24.3 was found to be de novo in an individual with developmental delay and ventriculoseptal defect. Several others with larger de novo deletions of 16q24.3 were found to have syndromic CVM. Familial inheritance was observed for several of these rare exonic deletions and duplications, suggestive of incomplete penetrance, and consistent with the inheritance pattern of CVM observed in the population. Our data implicate novel genes that are likely to be important for human cardiac patterning and support the hypothesis that rare CNV contribute to complex traits such as CVM.

249

De novo and rare inherited genomic alterations reveal genes responsible for isolated congenital diaphragmatic hernia and pentalogy of Cantrell phenotypes. *D.A. Scott[1], M.J. Wat[1], D. Veenma[2,3], J. Hogue[4], A.M. Holder[5], Z. Yu[1], J.J. Wat[6], N. Hanchard[1], O.A. Shchelochkov[7], C.J. Fernandes[8], A. Johnson[1,9,10], K.P. Lally[11], A. Slavotinek[4], O. Danhaive[12], T. Schaible[13], S.W. Cheung[1], K.A. Rauen[4], V.S. Tonk[14], D. Tibboel[3], A. de Klein[2].* 1) Dept Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Dept Clinical Genetics, Erasmus Medical Center, Rotterdam, the Netherlands; 3) Dept Paediatric Surgery, Erasmus Medical Center, Rotterdam, the Netherlands; 4) Dept Pediatrics, University of California San Francisco, San Francisco, CA; 5) Dept Surgery, Washington University School of Medicine, Saint Louis, MO; 6) Dept Biochemistry & Cell Biology, Rice University, Houston, TX; 7) Dept Pediatrics, University of Iowa Hospitals and Clinics, Iowa City, IA; 8) Dept Pediatrics, Baylor College of Medicine, Houston, Texas, USA; 9) Dept Obstetrics & Gynecology, Baylor College of Medicine, Houston, TX; 10) Dept Pediatric Surgery, Baylor College of Medicine, Houston, TX; 11) Dept Pediatric Surgery, University of Texas Medical School, Houston, TX; 12) Dept Neonatology, Bambino Gesu Children's Hospital, Rome, Italy; 13) Dept Pediatrics, University Hospital Mannheim, University of Heidelberg, Mannheim, Germany; 14) Dept Pediatrics, Texas Tech University Health Sciences Center, School of Medicine, Lubbock, TX.

Congenital diaphragmatic hernia (CDH) affects approximately 1 in 4000 newborns, accounts for 8% of all major congenital birth defects, and has a mortality rate of 30-60%. Most cases of isolated CDH are sporadic and are thought have a multifactorial etiology in which several genetic and/or environmental factors contribute to the development of CDH in an individual. Multifactorial inheritance makes identifying the genes that contribute to isolated CDH inherently difficult and has cast doubt on the clinical utility of screening individuals with isolated CDH for causative genomic alterations. Although efforts to identify the genes responsible for non-isolated and syndromic forms of CDH have met with greater success, the genetic factors that contribute to pentalogy of Cantrell—a constellation of defects that includes CDH, omphalocele, cardiac defects, sternal clefts, and pericardial defects—have yet to be fully identified. In a screen for genomic alterations in patients with CDH or diaphragmatic eventrations, we identified two patients with isolated diaphragmatic defects who inherited deletions of *ZFPM2* form unaffected parents and a large *de novo* 8q deletion overlapping the same gene was found in a patient with non-isolated CDH. Our screen also identified an individual with a partial pentalogy of Cantrell phenotype—CDH, omphalocele and cardiac defects—who carried a *de novo* deletion of *FZD2,* a gene whose protein product interacts with WNT3A and WNT5A to signal through both the canonical and non-canonical Wnt pathways. This finding was particularly interesting since mutations in *PORCN*, a gene that encodes a protein that modifies Wnt proteins—including WNT3A—for membrane targeting and secretion, have been shown to cause CDH and omphalocele as part of focal dermal hypoplasia (Goltz syndrome). Data from this study allow us to conclude that hapolinsuficiency of *ZFPM2* can cause autosomal dominantly inherited isolated CDH with incomplete penetrance and suggest that alterations in this gene may be responsible for a portion of families with what appears to be autosomal recessively inherited CDH. These results also suggest that dysregulation of Wnt signalling can contribute to the development of pentalogy of Cantrell phenotypes and demonstrate the clinical utility of screening for genomic alterations in individuals with both isolated and non-isolated diaphragmatic defects.

250

A DNA copy number variant conferring increased susceptibility to expressive speech delay in a distinct population. *W. Wiszniewski[1], J.V. Hunter[2], N.A. Hanchard[1], Q. Tian[1], S.W. Cheung[1], P. Stankiewicz[1], A. Ester[1], M.K. York[3], P. Bader[4], G. Scharer[5], H. Crawford[6], A. Mutirangura[7], P. Yanatatsaneejit[7], J. Kerr[8], M. Hurles[9], R. Goin-Kochel[1], G. Zapata[1], G. Simpson[10], L. Immken[10], M.E. Haque[3], M. Stosisc[3], N. Van Vink Chau[11], S. Dunstan[11], C. Simmons[11], M. Hibberd[12], C.C. Khor[12], M. Maletic-Savatic[3], J.W. Belmont[1], J.R. Lupski[1], S.R. Lalani[1].* 1) Department Molecular & Human Genetics, Baylor College of Medicine, Houston, TX, USA; 2) Department of Pediatric Radiology, Texas Children's Hospital, Houston, Texas, USA; 3) Department of Neurology, Baylor College of Medicine, Houston, Texas, USA; 4) Northeast Indiana Genetic Counseling Center, Fort Wayne, Indiana, USA; 5) Department of Pediatrics, the Children's Hospital, Aurora, Colorado, USA; 6) Department of Genetics, Children's Memorial Hermann Hospital, Houston, Texas, USA; 7) Center of Excellence in Molecular Genetics of Cancer and Human Diseases, Chulalongkorn University, Thailand; 8) Department of Neurology, Dell's Children's Medical Center, Austin, Texas, USA; 9) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK; 10) Dell's Children's Medical Center, Austin, Texas, USA; 11) Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam; 12) Genome Institute of Singapore, Singapore, Thailand.

DNA copy number variation (CNV) is recognized as an important basis of human genetic variation. Relatively rare human genome, but population specific CNV, may underlie many apparent common disease traits inherent to different world populations. In a study of 5,970 individuals with intellectual disability, using a high resolution customized genome-wide array-CGH with 180,000 oligonucleotides, we detected an identical 4 kb complex deletion in nine unrelated individuals. The affected children presented with expressive speech delay, autism spectrum disorder or global developmental delay. All of them were found to be of Southeast Asian descent. Large data sets of human genetic variation in this population demonstrate the allele frequency of this CNV to be ~1%. This deletion has not been observed in the European, West African or other Asian populations. The complex deletion CNV occurs on a common genetic background with shared haplotype in the unrelated individuals, suggesting that this is a founder deletion CNV. Our findings suggest that this deletion is frequently associated with expressive speech delay in the Southeast Asian pediatric population. Our observations provide a framework for larger population based studies investigating frequency and haplotype estimates for this CNV associated with a distinct phenotype with variable expression.

251

Further clinical and molecular delineation of the 15q24 microdeletion syndrome: fifteen newly reported patients and two atypical deletions. *H. Mefford[1,2], J. Rosenfeld[3], N. Shur[4], A. Slavotinek[5], V. Cox[5], R. Hennekam[6], H. Firth[7], L. Willatt[7], P. Wheeler[8], E. Morrow[9], J. Cook[1], R. Sullivan[10], A. Oh[11], J. Zonana[12], K. Keller[12], M. Hannibal[1,2], S. Ball[13], J. Kussmann[14], J. Gorski[14], S. Zelewski[15], V. Banks[16], W. Smith[16], R. Smith[16], L. Paull[11], K. Rosenbaum[11], D. Amor[17], J. Silva[18], A. Lamb[3], E. Eichler[19].* 1) Pediatrics, University of Washington, Seattle, WA; 2) Seattle Children's Hospital, Seattle, WA; 3) Signature Genomics, Spokane, WA; 4) Pediatrics, Rhode Island Hospital, Providence, RI; 5) Pediatrics, UCSF, San Francisco, CA; 6) Pediatrics, AMC, University of Amsterdam, Amsterdam, The Netherlands; 7) Medical Genetics, Cambridge University Hospitals Foundation Trust, Cambridge, UK; 8) Genetics, Nemours Children's Clinic, Orlando, FL; 9) Bio Med Molecular, Cellular Biology Biochemistry, Brown University, Providence, RI; 10) Plastic Surgery, Rhode Island Hospital, Providence, RI; 11) Children's National Medical Center, Washington, DC; 12) Oregon Health & Sciences University, Portland, OR; 13) Children's Village, Yakima, WA; 14) University of Missouri Healthcare System, Columbia, MO; 15) Altru Health System, Grand Forks, ND; 16) Maine Medical Partners, Portland, ME; 17) Murdoch Children's Research Institute, Royal Children's Hospital, Victoria, Australia; 18) Institute Nacioanl Saude Dr. Ricardo Jorge, Porto, Portugal; 19) Genome Sciences and Howard Hughes Medical Institute, University of Washington, Seattle, WA.

Chromosome 15q24 microdeletion syndrome is a rare, novel genomic disorder characterized by growth retardation, unusual facial morphology, genital anomalies and digital anomalies. To date, twenty patients with 15q24 deletion have been reported; seventeen have had detailed breakpoint analysis, and a 1.1-Mb critical region has been proposed. We report clinical and molecular characterization of fifteen additional patients with deletions in the 15q24 region, nearly doubling the number of reported patients. Among these are nine distinct deletions with different breakpoints ranging in size from 266 kb to 3.75 Mb. The two smallest deletions in our series are both *de novo* and lie outside of the proposed critical region, with the smallest encompassing only five unique genes (*COMMD4, NEIL1, MAN2C1, SIN3A, PTPN9*). The majority of deletion breakpoints lie within segmental duplication (SD) blocks. Relatively low sequence identity and large intervals of unique sequence between SD blocks likely contribute to the rarity of 15q24 deletions, which occur 8-10 times less frequently than 1q21 or 15q13 microdeletions in our series. Common features in our patients include prominent forehead, high anterior hairline, ocular abnormalities, ear abnormalities, severe speech delay or absence of speech and mild to moderate intellectual disability. The molecular characterization of our patients suggests that the core cognitive features of the 15q24 microdeletion syndrome, including developmental delays and severe speech problems, are largely due to deletion of genes in the proposed 1.1-Mb critical region. However, the two small, atypical deletions in our series suggest that genes outside of this critical region also play an important role in cognition and in the development of characteristic facial features associated with 15q24 deletions. In summary, deletions within the 15q24 region are not only rare but are variable in size and gene content. Knowledge of the breakpoints and size of deletion combined with the natural history and medical problems of our patients provide insights that will inform management guidelines.

252

Characterization of a 8q21.11 Microdeletion Syndrome associated with Intellectual Disability and a Recognizable Phenotype. *M. Palomares Bralo[1, 4], A. Delicado[2, 4], E. Mansilla[2, 4], M.L. de Torres[2, 4], E. Vallespín[1, 4], L. Fernández[2, 4], V. Martinez- Glez[1, 4], S. García-Miñaur[3, 4], J. Nevado[1, 4], F. Santos Simarro[3, 4], V.L. Ruiz-Perez[4, 5], S.A. Lynch[6], F.H. Sharkey[7], A.C. Thuresson[8], G. Annerén[8], E.F. Belligni[9], M.L. Martínez-Fernández[4, 10], E. Bermejo[4, 10, 11], B. Nowakowska[12, 13], A. Kutkowska-Kazmierczak[12], E. Bocian[12], E. Obersztyn[12], M.L. Martínez-Frías[4, 10, 11], R.C.M. Hennekam[14], P. Lapunzina[3, 4].* 1) Section of Functional and Structural Genomics Instituto de Genética Médica y Molecular (INGEMM), Madrid, Madrid, Spain; 2) Section of Cytogenetics. Instituto de Genética Médica y Molecular (INGEMM), IdiPAZ, Hospital Universitario La Paz. Madrid, Spain;; 3) Section of Clinical Genetics. Instituto de Genética Médica y Molecular (INGEMM), IdiPAZ, Hospital Universitario La Paz. Madrid, Spain;; 4) CIBERER, Centro de Investigación Biomédica en Red de Enfermedades Raras, ISCIII, Madrid, Spain; 5) IIB, Instituto de Investigaciones Biomédicas de Madrid (Centro Superior Investigaciones Científicas-Universidad Autónoma de Madrid; 6) National Centre for Medical Genetics, Our Lady's Childrens Hospital, Crumlin, Dublin, Ireland; 7) Microarray Unit, Cytogenetics Laboratory, Western General Hospital Edinburgh, Edinburgh, UK; 8) Department of Immunology, Genetics and Pathology, Uppsala University Rudbeck Laboratory. Dag Hammarskjöldsv, Uppsala, Sweden; 9) Department of Pediatrics, University of Torino, Torino, Italy; 10) ECEMC, Estudio Colaborativo Español de Malformaciones Congénitas, ISCIII, Madrid, Spain; 11) IIER, Instituto de Investigación de Enfermedades Raras, ISCIII, Madrid, Spain; 12) Department of Medical Genetics, Institute of Mother and Child, Warsaw, Poland; 13) Center for Human Genetics, K.U Leuven, University Hospital Gasthuisberg, Leuven, Belgium; 14) Department of Pediatrics, Academic Medical Center, UVA, Amsterdam, The Netherlands.

We report eight unrelated individuals with intellectual disability and overlapping submicroscopic deletions of 8q21.11 (0.66-13.55 Mb in size). The deletion was familial in one and simplex in seven individuals. The phenotype was remarkably similar and consisted of a round face with full cheeks, high forehead, ptosis, cornea opacities, underdeveloped alae, short philtrum, cupid's bow of the upper lip, down-turned corners of the mouth, micrognathia, low-set and prominent ears, and mild finger and toe anomalies (camptodactyly; syndactyly; broadening of first rays). Intellectual disability, hypotonia, decreased balance, sensorineural hearing loss, and unusual behavior were frequently observed. High resolution oligonucleotide array showed different proximal and distal breakpoints in all of them. Sequencing studies in three of the individuals revealed that proximal and distal breakpoints were located in unique sequences with no apparent homology. The smallest region of overlap was a 539.7 kb interval encompassing three genes: a Zinc Finger Homeobox 4 (ZFHX4), one micro RNA of unknown function and one nonfunctional pseudogen. ZFHX4 encodes a transcription factor expressed in adult human brain, skeletal muscle and liver. It has been suggested to be a candidate gene for congenital bilateral isolated ptosis. Our results suggest that the 8q21.11 submicroscopic deletion represents a clinically recognizable entity and that a haploinsufficient gene or genes within the minimal deletion region could underlie this syndrome.

253

"Peeking near the peaks" for large-effect rare variants. *A. Coventry[1], L. Bull-Otterson[2], A. Keinan[1], X. Liu[3], A. Clark[1], T. Maxwell[3], J. Hixson[3], T. Rea[4], A. Templeton[5], D. Muzny[2], L. Lewis[2], D. Villasana[2], E. Boerwinkle[3], R. Gibbs[2], C. Sing[4].* 1) Molec Biol & Gen, Cornell Univ, Ithaca, NY; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, 77030; 3) Human Genetics Center, UT Houston Health Science Center, Houston, TX, 77030; 4) Department of Human Genetics, University of Michigan School of Medicne, Ann Arbor, MI 48109; 5) Department of Biology, Washington University, St Louis, MO 63130.

We tested a novel resequencing assay targeting marker-based GWAS hits. It addresses concerns that some GWAS peaks may be caused by "synthetic associations," i.e., multiple distant rare variants of large phenotypic effect which are all in linkage disequilibrium (LD) with a GWAS-flagged marker. Goldstein *et al*.(*PLoS Biology*, Jan 2010 doi:10.1371/journal.pbio.1000294), for example, note that the LD blocks targeted by most resequencing assays would miss the distant phenotypically-relevant rare variants hypothesized by a synthetic association scenario. Such variants would lie on long, rare haplotypes associated with large phenotypic effects, so for each GWAS marker we investigated, we resequenced haplotypes which a) carry the marker and b) are associated with the largest phenotypic effect among haplotypes carrying the marker. We estimated haplotypes from Affymetrix SNP Array 6.0 genotypes measured in the ARIC cohort (13,422 people). When recombinants of the targeted haplotypes also had informative phenotype distributions, we used them to narrow down the resequencing loci. We resequenced 10 regions, choosing 50 samples carrying the targeted haplotypes for each region (total of 500 samples.) For each haplotype we targeted, we chose the haplotype carriers from the upper and lower tails of the phenotype distribution, with an equal number from each tail. All samples were resequenced at all targeted loci. Variants found only in haplotype carriers were regenotyped in the entire ARIC cohort, and checked for large phenotypic impact. We found such large-effect variants, showing that a GWAS resequencing assay based on combined phenotype/haplotype analysis can be very effective compared to earlier methods, because it provides a clear signal of potential synthetic associations.

254

A Unified strategy for rare variant testing: combined burden based and similarity based testing. *E. Urrutia, M. Wu.* Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

Rare variants are thought to have a strong effect on complex traits and can now be detected using sequencing technology. Since testing strategies for common variants are underpowered for rare variants, region based testing for the cumulative effect of a group of rare variants on a trait has been established as an effective strategy and several tests have been developed. Two important practical challenges have emerged for researchers interested in rare variant analysis. First, it is difficult to choose which test to use in practice since each of the developed methods is powerful under different structures of the genotype/phenotype relationship. Second, it is unclear which group of variants within a region should be tested, e.g. all rare variants or only the nonsynonymous variants. The answer to both problems depends on knowing the underlying trait architecture and which variants are causal. This information is unavailable and would preclude need for analysis. Therefore, we propose a new pragmatic testing approach that searches across a range of particular tests and variant grouping strategies. We demonstrate that several popular rare variant tests including burden based tests, the C-alpha test, and others are all special cases of the sequence kernel association test (SKAT). SKAT is a similarity based test wherein pair-wise similarity in trait value between subjects is compared to pair-wise similarity in the rare variant genotypes between subjects as measured through a kernel function. Choosing a particular test is equivalent to choosing a particular kernel function. Similarly, we can show that the choice of which group of variants to test also reduces to a kernel choice problem. Thus, we develop the Multi-Kernel SKAT (MK-SKAT), a statistical framework based on perturbation, which tests across a range of kernel choices that correspond to a range of rare variant tests and variant groupings. We demonstrate through simulations that our testing framework controls type I error. Both simulations and real data analyses show that MK-SKAT tends to have high power across a virtually all settings, losing a little bit of power when compared to the optimal test for a particular scenario but having much greater power than poor choices of test or grouping strategy. Additional features of the method are its ability to effectively control for confounding variables, as well as its computational efficiency.

255

A Powerful Association Test of Rare Variants Using a Random Effect Model. *K.F. Cheng[1], J.Y. Lee[2], C. Li[3].* 1) Biostatistics Center, China Medical University, Taichung, Taiwan; 2) Graduate Institute of Statistics, National Central University, Chungli, Taiwan; 3) Department of Biostatistics and Center for Human Genetics Research, Vanderbilt University, Nashville, USA.

There is an emerging interest in association studies of multiple rare variants. Most association tests suggested in the literature involve collapsing rare variants with or without weighting. These tests often have good power performance when there are many functional rare variants in the analysis. However, the power of these tests can be dramatically reduced when many non-functional rare variants are included or when the range of variant frequency is wide with varied levels of effect. We propose a random effect model for describing the heterogeneous effects of rare variants. The model uses a mixture distribution of two components. The first component assumes that the effects of the functional variants follow a uniform distribution. The second component assumes that effects of the non-functional variants follow a degenerate distribution. Under this model, testing for association is equivalent to testing whether the fraction of the first component equals zero. We propose a likelihood ratio (LR) statistic and a permutation procedure for p-value calculation. Simulations showed that this test has controlled Type I error rate. It has very similar power as that for the combined multivariate and collapsing (CMC) and weighted sum (WS) tests when all the rare variants are functional. As the fraction of non-functional rare variants increases, the new test has much better performance in resisting power loss than the CMC and MS tests. This effect is consistent under a wide range of scenarios with different variant allele frequencies, sample sizes, and numbers of functional and non-functional variants. For example, when there were 10 functional variants with variant allele frequencies between 0.001 and 0.01 and odds ratios around 2, with 500 cases and 500 controls, the power of the new test dropped from 93% to 74% as the proportion of non-functional variants increased from 0% to 95%, while the CMC test dropped from 94% to 14% and the WS test dropped from 95% to 33%. An example from a breast cancer study also demonstrated that the LR test was able to detect association signal(p-value=0.0275)for variants near SNP rs10510102 at the FGFR2 locus on 10q26.13 and with minor allele frequency <0.03,while the CMC and WS tests failed to detect any signal(p-values=0.3019 and 0.1690,respectively).

256

A new Bayesian graphical model for detecting rare variant disease associations. *Y. Zhang[1], K. Wang[2], H. Hakonarson[3], S. Ghosh[4].* 1) Statistics, Pennsylvania State University, State College, PA; 2) Psychiatry and Preventive Medicine, Zilkha Neurogenetic Institute, University of Southern California, Los Angeles, CA; 3) Human Genetics and Molecular Biology, University of Pennsylvania School of Medicine, Philadelphia, PA; 4) Quantitative Genetics, GlaxoSmithKline, King of Prussia, PA.

We present a novel Bayesian model to detect rare variants associated with common disease in case-control studies. We consider various disease models, including (1) multiple rare variants demonstrating a collective frequency difference; (2) haplotype frequency difference; and/or (3) region-specific frequency difference (i.e., not all rare variants are disease related). We use simulated data sets to demonstrate the power of our method compared to existing rare variant methods. We also demonstrate the successful usage of our method in two real resequencing data sets: type 1 diabetes low coverage whole genome data, and inflammatory bowel disease high coverage data at known loci from 600 patients. Compared to most existing rare variant methods, our approach has several desired features. First, rather than testing on predefined sets of rare variants, or collapsing all rare variants within a region, our method dynamically selects combinations of subsets of rare variants to optimize power. Second, the method uses graphs to account for linkage disequilibrium (LD) amongst genetic variants, which is more flexible and effective in this problem than standard Markov chains (or saturated models) that are often used for LD. Only the most disease-relevant variants will be selected by our method, while secondary associations through LD are filtered out. Third, our method works simultaneously for both common and rare variants without arbitrarily choosing between the two. Our method thus compares common variant association (including synthetic associations) to rare variant association using the likelihood principle. Our method is computationally efficient on large data sets with many thousands of variants. It is also straightforward to incorporate into our model any prior knowledge on function to facilitate identification of the true disease mutations. Given that this is a high-dimensional complex problem, rather than reporting an "optimal" set of disease-associated variants, we output posterior probability of disease association for each variant along with a corresponding disease graph. In summary, we provide a Bayesian, graphical approach for rare variant association that is flexible and powerful across different models, parameters and data types. Finally, the output provides a realistic ranking scheme for further simultaneous follow-up of disease-causing rare and common variants.

257

A new analytical approach to prove the involvement of a rare variant in disease susceptibility. *F. Clerget-Darpoux[1], H. Perdry[1,3], P. Broet[1,3,4], B. Muller-Myhsok[2].* 1) INSERM U669, Batiment INSERM 15/16, Villejuif Cedex, France; 2) Max Planck Institute of Psychiatry, Munich, Germany; 3) University Paris-Sud, France; 4) Assistance Publique-Hopitaux de Paris, Paul Brousse Hospital, France.

Large-scale sequencing projects such as whole-genome sequencing, whole-exome sequencing or targeted resequencing of specific areas of interest are increasingly commonly used to demonstrate the involvement of rare variants with minor allele frequencies below 5% in disease susceptibility. The effects of such rare variants as measured in genotypic relative risks are believed to be considerably higher than those of the common variants found through GWAS. Genotypic relative risks as high as 2 or even 4 may be realistic. This leads to two different but coinciding consequences for the demonstration of the effect of a rare variant in a complex disease. Firstly, we can demonstrate that in index patients recruited from affected sib-pairs the frequency of the rare alleles is greatly increased compared to samples composed of unrelated cases. To give an example, for an allele with a population allele frequency of 1% and a genotypic relative risk of 4 the expected frequency of unrelated cases carrying the risk variant is around 8%; whilst in patients with an affected sibling this frequency is more than doubled to roughly 17%. As a consequence the number of patients that to have to be sequenced in order to establish a role of this type of variant at a genome-wide level of significance (5e-08) would drop from more than 1100 to less than 400. Secondly in affected sib pairs an additional aspect of genetic information, orthogonal to the information on association, can be used. For the example aforementioned we can also compute the distribution of IBD conditional on the genotype of the index patients. In a hypothetical sample of 400 affected sib pairs this test will have a power of more than 90 % again at a genome wide level (assuming genetic parameter values as specified above). Alternatively these observations may be used to better model the role of rare variants in complex diseases. Thus, clearly, the information provided by the affected sibs in rare variants when analyzed stratified on the genotype of the sequenced index yields more efficient designs both with respect to sequencing cost and sample size requirements. In addition it offers an orthogonal aspect of information absent from case-control studies. Furthermore it offers a handle to model the genetic effect of a given variant. We would like to point out that these concepts are readily transferred also to approaches considering ensembles of variants in a given genetic region rather than single variants.

258

FINDING RARE VARIANTS IN GWAS AND EXOME SEQUENCING DATA BY MAKING USE OF RECENT COMMON FOUNDER INFORMATION. *W. Yang, L. Zhang, D. Ying, PC. Sham, YL. Lau.* Paediatrics & Adolescent Med, The Univ Hong Kong, Hong Kong, Hong Kong.

Two issues hinder the discovery of rare genetic variants using the current methodology. First, the power of association study is directly related to allele frequency, thus the current sample size of GWAS means that they are all underpowered for rare variants discovery. Second, unlike common alleles, rare alleles are more likely to be population-specific and therefore poorly-represented in current genotyping platforms. Some of the rare variants contributing to diseases are recent founder mutations and they have a few features. 1. They are each likely to only affect very small number of patients. 2. They reside on founder haplotypes that can be detected from SNP data. 3. They may relate to severe or particular clinical manifestations. We have developed a series of algorithms to effectively distinguish recent founder haplotypes shared by patients, making use of the fact that they are younger, longer and very low in frequency compared to other haplotypes in the general population that are shaped by many more generations of recombination events. We also developed an evaluation process to apply the same procedure to control samples, so that the program is specifically designed to detect founder haplotypes in cases that are associated with disease. To our knowledge, this is the first comprehensive program to identify recent founder haplotypes and evaluate their role in disease susceptibility. Next generation sequencing technologies are capable of detecting these rare variants. The problem is that, usually without statistical means and functional characterization, it is very hard to distinguish rare variants related to disease from many other variants. In this regard, the recent founder haplotype method provides the badly needed statistical means to help establish links between genetic variants identified and the underlying disease phenotype. The founder haplotype detection method is applicable to both complex diseases and single gene disorders when combined with exome sequencing or targeted sequencing. For free download, please visit http://paed.hku.hk/uploadarea/yangwl/html/index.html (HRRA for recessive mutations and HaploShare for dominant mutations, yangwl@hku.hk). We are currently developing the method to make it directly applicable to exome sequencing data.

259

Rare variant analysis in genetic association studies under quantitative trait-dependent sampling designs. *Y.E. Yilmaz[1,2], J.F. Lawless[2,3], S.B. Bull[1,2].* 1) Samuel Lunenfeld Research Institute, Toronto, Ontario, Canada; 2) Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 3) Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada.

We develop and evaluate sampling designs that sequence some fraction of an entire GWAS cohort, and we describe and compare corresponding methods for rare variant analyses, assuming an existing cohort of unrelated individuals, phenotyped for a quantitative trait (QT). In addition to simple random sampling (SRS), we examine QT-dependent sampling designs, including extreme phenotype selection, which are of increasing interest in genetic association analysis, especially in studies concerned with detection of rare genetic variants. We develop a novel inverse probability selection (IPS) sampling design, in which all individuals have a non-zero probability of being selected, but those with extreme phenotypes have a proportionately higher probability. In QT-dependent sampling designs, because we select observations according to QT value, standard linear regression methods that treat the rare variant score as a covariate and the QT as the dependent variable are not valid. We propose and evaluate three approaches to solve the matter. In the first, we use likelihood methods for linear regression models with expensive variables missing by design (Zhao et al., Biometrical J 2009), apply semiparametric maximum likelihood (SML) estimation and use the likelihood ratio (LR) for testing the association. For the IPS design, another approach involves application of inverse probability weighting (IPW) in the estimation of linear regression model parameters. This serves to account for unequal sampling probabilities and yields estimates generalizable to the source population. Thirdly, in a potentially robust approach for QT-dependent sampling we reverse the direction of the regression by modelling the gene-specific rare allele count conditional on the QT value using Poisson or mixed Poisson regression. In evaluations of sampling designs and analytic methods by simulation, we found that QT-dependent sampling designs, including IPS, are generally more efficient than SRS of the same number of individuals, with efficiency depending on the sampling fraction and sample allocation. For most designs examined, SML provides relatively efficient estimates, although LR test type I error tends to be less than nominal, with higher power than IPW. In addition, in many designs Poisson or mixed Poisson regression of allele counts appears useful as a robust approach for rare variant analysis.

260

Disease risk prediction to prioritize individuals and families for whole-genome sequencing experiments. *L. Jostins[1], A.P Levine[2], L.B. Lovat[3], A.P. Walker[2], A.W. Segal[2], J.C. Barrett[1].* 1) Human Gen, Wellcome Trust Sanger Inst, Cambridge, United Kingdom; 2) Division of Medicine, University College London, WC1E 6JJ, UK; 3) Department of Gastroenterology, University College London Hospitals NHS Foundation Trust, London, NW1 2BU, United Kingdom.

For many complex traits a significant minority of variance in disease risk has been explained by loci discovered in genome-wide associations studies. These variants can be used in risk prediction, but the utility of such prediction in a clinical or personal setting is hotly debated. However, risk prediction can also be used to inform the design and implementation of sequencing experiments to discover novel low-frequency risk alleles. We discuss two such applications of risk prediction: the selection of subsets of large case-control cohorts, and the selection of families for combined linkage-resequencing studies. While case-control sequencing is the most reliable way of detecting low-frequency variants, sequencing entire cohorts is prohibitively expensive and often only a subset can be sequenced. Conditioning on low risk scores creates a higher load of undiscovered risk variants. As a result, case individuals with a low risk score at known loci have a higher risk allele frequency at unknown loci, and conversely disease-free controls with high risk have a lower risk allele frequency. We demonstrate that this results in a significant increase in power, and that this effect increases with the size of the risk variant under study and the size of the cohort the cases are taken from. Combining sequencing and genotyping of multiplex families can also be used to discover rare associated variants, by identifying functional mutations on haplotypes shared by affected individuals. However, a pedigree-wide high risk at known associated loci can indicate multiplex families that are unlikely to have rare risk variants. Likewise, high risk scores can indicate affected individuals that lack a rare variant present in other family members: in effect these are phenocopies. We generalize the risk prediction approach to apply to a partially genotyped family with missing data, using an application of the inside-outside algorithm to pass genotype likelihoods around the pedigree. We show examples of multiplex inflammatory bowel disease families that illustrate both scenarios, and demonstrate how we have used risk prediction to prioritize families and family members who are more likely to harbor rare risk variants for whole-genome sequencing. We present a software package, Mangrove, for performing risk prediction either for individuals or families with sporadic or systematic missing data. The package includes tools to select high-priority targets for sequencing follow-up.

261

Indexing and deep sequencing of a point mutation in mosaic samples from Proteus syndrome: Alternative detection strategies. *M.J. Lindhurst[1], J.K. Teer[1,2], J.J. Johnston[1], E.M. Finn[1], J.C. Sapp[1], J.C. Mullikin[2], L.G. Biesecker[1].* 1) GDRB, NHGRI/NIH, Bethesda, MD; 2) NIH Intramural Sequencing Center, Rockville, MD.

Detection of mutations in mosaic samples is challenging due to the variability in the level of the mutant allele that can occur in various tissues. Recently, we have shown that Proteus syndrome is caused by a somatic mutation that results in constitutive activation of a key signaling pathway in affected tissues. The variant was initially identified using next gen sequencing of 17 samples. To confirm and extend these findings, we screened 158 samples by Sanger sequencing. Seventy-one of these samples had clear evidence of the mutant allele on the electropherograms, 17 had electropherograms that suggested a low level of the mutation, and 70 were apparently negative. However, it was challenging to distinguish low levels (<1%) of mosaicism from background signals or noise in the electropherogram traces. We next developed a PCR/restriction endonuclease assay (MboII) that could be separated on the ABI3130 and was designed such that MboII digested only the mutant amplicon. Areas under the curve for the mutant and wild-type peaks on the instrument were assessed quantitatively. A validation experiment using dilutions of cloned mutant and wild-type DNA showed a correlation of r2=0.9993 with a lower limit of 1% mutant allele sensitivity. Using this assay, the mutant allele was detected in all 71 of the positive samples, 15 of the above 17 samples that were suggestive and 15 samples of the 70 samples that were negative by Sanger. This increased the number of patients with at least one positive sample from 23 to 27. The level of mutation in these samples ranged from 1% to ~50%. We screened 75 samples from non-Proteus individuals; all were negative for the mutation. While the PCR/MboII assay increased sensitivity compared to sequencing, variation increased at low mutant allele levels (again, <1%). To address this, we PCR-amplified the mutation from 12 samples using indexing primers and used next-gen sequencing to generate ultra-deep coverage. These samples were pooled, and sequenced on a single GAIIx lane (76 bp paired-end). 28,568,780 total reads were generated, and 335,378-1,649,523 reads were assigned to each sample. Allele frequencies were calculated at the variant position, and agreed well with the MboII results. Additional optimizations will be performed and have the potential to further improve sensitivity. (The authors confirm that the gene and mutation will be disclosed at the ICHG meeting should the abstract be selected for presentation).

262

Use and performance of the Metabochip Genotyping Array in African Americans: the PAGE Study. *S. Buyske[1,13], Y. Wu[14], J.L. Ambite[2], T. Assimes[3], E. Boerwinkle[4], C. Carty[5], I. Cheng[6], B. Cochran[7], D. Duggan[8], L. Dumitrescu[9], M. Fesinmeyer[5], C.A. Haiman[10], J. Haessler[5], L. Hindorff[11], H.M. Kang[12], C. Kooperberg[5], Y. Lin[5], L. Le Marchand[6], T. Matise[13], S. Mitchell[9], K. Mohlke[14], U. Peters[5], F. Schumacher[10], B.F. Voight[15], D. Crawford[9], K. North[14].* 1) Statistics & Biostatistics, Rutgers Univ, Piscataway, NJ; 2) Information Sciences Institute, University of Southern California, Los Angeles, CA; 3) Department of Medicine, Stanford University School of Medicine, Stanford, CA; 4) Human Genetics Center, University of Texas Health Science Center, Houston, TX; 5) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 6) University of Hawaii Cancer Center, Honolulu, HI; 7) Sponsored Programs, Baylor College of Medicine, Houston, TX; 8) Translational Genomics Research Institute, Phoenix, AZ; 9) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 10) Keck School of Medicine, University of Southern California, Los Angeles, CA; 11) Office of Population Genomics, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 12) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 13) Department of Genetics, Rutgers University, Piscataway, NJ; 14) Department of Genetics, University of North Carolina, Chapel Hill, NC; 15) Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA.

The Metabochip is a cost-efficient custom array designed for replication and fine mapping studies of metabolic, cardiovascular, and anthropometric traits. A large fraction of this array targets low frequency and rare variation at loci with established (mostly in European descent populations) trait associations, facilitating trans-ethnic study of alleles across the frequency spectrum at these loci. To date, there has been no systematic empirical evaluation of the performance of this array in a non-European population. We describe the performance of the Metabochip in the African American participants of the PAGE (Population Architecture using Genomics and Epidemiology) consortium and explore the utility of the chip for fine mapping of GWA signals in this sample using the association between high density lipoprotein cholesterol (HDL-C) and the CETP locus as a case study. A total of 5,897 African American individuals were genotyped and passed QC by three studies participating in PAGE: Atherosclerosis Risk In Communities (ARIC), Multiethnic Cohort (MEC), and Women's Health Initiative (WHI). HapMap YRI samples were also genotyped. Only 0.2% of samples failed QC. About 7.3% of the Metabochip SNPs were technical failures; 2.7% could not be uniquely mapped to the genome; 0.8% were discordant between GenCall and GenoSNP calls; 0.6% were discordant for HapMap YRI genotypes; 6.8% were monomorphic. In sum, 161,098 (81.9%) SNPs on the Metabochip passed QC. Minor allele frequency (MAF) varied dramatically between the Metabochip and GWAS arrays. E.g., 21.6% of the polymorphic Metabochip SNPs had MAF less than 2.5%, compared to 5.8% for Affymetrix 6.0 and 6.8% for Illumina 1M. At the CETP locus on chromosome 16, we observed the strongest association between HDL-C and rs17231520 (per allele effect estimate 7.81 mg/dL, SE 0.62, p=7.2E-36, MAF 7.2%). This variant, described previously in a candidate gene study as associated with HDL-C in African Americans, resides in an SP1/SP3 binding site and has been shown to influence gene transcriptional activity and CETP protein function. Rs17231520 is not on the Affymetrix 6.0 or Illumina 1M nor do they contain SNPs in high LD (r^2 0.65) with it, suggesting that this signal would not be detected by a typical GWAS in African Americans. In summary, our results demonstrate the suitability and cost effectiveness of Metabochip genotyping to generalize European ancestry GWA signals to African Americans.

263

Integrated genotyping of SNPs from multiple independent technologies: a framework and applications. *J. Flannick[1,2,3], J. Korn[1,2,3], P. Fontanillas[3], G. Grant[4], D.A. Altshuler[1,2,3].* 1) Molecular Biology, Massachusetts General Hospital, Boston, MA; 2) Genetics, Harvard Medical School, Boston, MA; 3) Medical and Population Genetics, Broad Institute, Cambridge, MA; 4) Genetic Analysis Platform, Broad Institute, Cambridge, MA.

Every genetic study aims to obtain accurate genotypes for every studied sample at every studied variant. Until whole genome sequencing becomes significantly less expensive, studies will use a variety of less comprehensive assays such as traditional microarray genotyping and low coverage sequencing. Because these assays have different strengths and weaknesses, many studies will use both --- as well as linkage disequilibrium information --- to inform genotype calls. Therefore, a principled framework is needed to produce calibrated genotype calls when these technologies agree and to arbitrate when they disagree.

We developed a framework to calculate the genotypes of all studied samples at all studied SNPs simultaneously, conditional on joint evidence from microarray and sequence data. We implemented this framework in SUMMIT, a tool that produces integrated calls from an arbitrary number of potentially overlapping sequence or microarrary data input files. We evaluated SUMMIT on several empirical datasets: low and deep coverage sequence data, a custom designed Illumina microarray, and the new Illumina OMNI array. We assessed the impact of the joint calling framework in three settings of broad importance to the next generation of genetic association studies: data integration, validation, and replication.

We show that, most important, joint calls are more accurate than calls made from sequence or microarray data independently. SNPs and samples assayed by both technologies have up to 50% fewer discordances with gold standard data; in addition, due to the joint calling approach, a significant improvement extends to samples and SNPs assayed on only one technology. We also show that SUMMIT can arbitrate disagreements between the different technologies, a useful application for studies that use microarrays to validate SNPs discovered from sequence data. Our results have implications for studies that genotype SNPs in large replication cohorts, as they suggest that genotypes called from microarrays can be improved even if only a subset of the samples have sequence data available. In conclusion, our work provides an important theoretical framework for any genetic study that uses multiple genotyping technologies, as well as a practical tool of use to any study with multiple input data sources.

264

Genotyping and sequencing on conductive nanowires: DNA testing "while you wait". *J. Burn[1,3], S.P. Whitehouse[2,3], E. Warburton[3], J. O'Halloran[3].* 1) Institute of Genetic Medicine, Newcastle Univ, Newcastle on Tyne, United Kingdom; 2) Department of Chemistry Sheffield University,UK; 3) QuantuM Dx ltd Centre for Life, Newcastle,UK NE1 4LP.

Technological advances; G banded karyotyping, Sanger Sequencing, PCR, FISH,parallel pyrosequencing, have defined the progress of genetics and genomics from the academic periphery to the centre of healthcare. We present a new technology we believe will help DNA analysis become ubiquitous. We have succeeded in developing a linker system to attach DNA to conductive nanowires. Modified bases, carrying extra charge and with chain termination properties, disturb the impedance of the nanowire as they bind, effectively converting sequence into computer code without intervening steps. Combining homegrown and in-licensed methodology, we are able to achieve DNA extraction in around four minutes and PCR amplification through a microfluidics system suitable for use in a hand held device. Our prototype is at an advanced stage of design and offers the prospect of an affordable and robust handset able to offer a genotype in a "near patient" setting for a few dollars. Small cassettes carrying the biological sample and appropriate chemistry will allow "while you wait" testing in the pharmacy or clinic for significant pharmacogenetic variants such as those associated with warfarin sensitivity. We have funding to explore the application of our disruptive genomic technology in remote settings and deprived communities. A cheap test for HIV is achievable which will have major significance in developing countries. Similar approaches to other sexually transmitted diseases offers the prospect of confidential self testing which will improve access to those at highest risk in developed countries. Extended sequencing is limited by the Debye distance which is, in effect, the electrical sphere of influence of the modified nucleotides such that only short lengths of sequence can be achieved before detection by the nanowire is lost. We will circumvent this by lining nanochannels with nanowires allowing lengths of DNA, potentially up to 10kb, to be sequenced intact. This opens the way to a relatively cheap desktop device for transcriptome, exome and whole human genome sequencing. While our nanowire techniques are not yet mature, we anticipate market entry well before our next ICHG. The issues raised merit close attention in view of the potential for a major academic, clinical, ethical and economic impact.

265

Rapid diagnosis of Glycogen Storage Diseases by next generation sequencing. *S. Abbs[1], F. Smith[1], T. Cullup[1], C. Deshpande[2], H. Mundy[2], K. Bhattacharya[3], M. Champion[2], SC. Yau[1].* 1) DNA Laboratory, Genetics, GSTS Pathology, Guy's Hospital, London, United Kingdom; 2) Guy's & St Thomas NHS Foundation Trust, London, United Kingdom; 3) Children's Hospital at Westmead, Sydney, Australia.

Massively parallel sequencing of 18 genes known to cause glycogen storage diseases (GSD) has been established to offer rapid and cost effective diagnosis for these heterogeneous metabolic disorders. An in-solution Agilent Sure Select array was designed to capture a total of ~1Mb of sequence from all 284 exons and surrounding introns from the 18 genes. After index tagging the resulting libraries were subjected to next generation sequencing on an Illumina GAIIx, with 12 samples per lane.

The initial array design used 5x tiling of 120bp probes, and resulted in more than 10 fold differences in average levels of sequence coverage for different regions. To level out and ensure a minimum 30x coverage for all exons we grouped the probes based on initial coverage, and redesigned the array with increasing probe copy numbers for groups giving lower coverage. This reduced the comparable difference in coverage to less than 4 fold, and provided valuable information for future array design. The methodology was validated to detect single nucleotide changes, deletions from 1-38bp, and exon copy number variants.

Following optimisation and validation, an initial 16 GSD patients were screened. In 14 patients (87.5%) clear pathogenic variants were detected which enabled confirmation of diagnosis. In one additional patient a single heterozygous *PHKB* mutation was detected. Heterozygous nucleotide variants were detected in between 22-54% of reads.

One of the 16 patients had presented with a clinical and biochemical diagnosis suggesting GSD type III, VI, or IX. Over a period of 12 years the patient was subjected to repeated enzyme analysis, a liver biopsy, and conventional sequencing of three GSD IX genes, all of which failed to confirm a diagnosis. A homozygous pathogenic variant was identified in the *PYGL* gene, finally confirming a diagnosis of GSD type VI. This test yielded a diagnosis in a matter of weeks, and at approximately 1/10th the combined cost that had been spent on all the inconclusive tests performed over 12 years. This case highlights the advantages of next generation technology, reducing the reliance on enzymology and invasive tests, which may give inconsistent results or may not be available. This one-stop test provides a rapid and cost-effective approach for diagnosis of all GSD's, and the methodology can be applied to any group of heterogeneous disorders.

266

Exome sequencing to determine genome-wide DNA copy number variation (EXO-CNV). *AM. Sulonen[1,2], H. Edgren[1], P. Ellonen[1], H. Almusa[1], O. Kallioniemi[1], J. Saarela[1].* 1) Institute for molecular medicine Finland FIMM, Helsinki, Finland; 2) Unit of Public Health Genomics, National Institute for Health and Welfare, Helsinki, Finland.

For the diagnostic assessment of cancer and genetic diseases, it is necessary to identify both DNA sequence alterations as well as copy number changes. Exome sequencing following hybrid-capture of the protein-coding sequences has become a key technology for the analysis of DNA sequence variation in the clinical diagnostic assessment of e.g. rare genetic diseases or human cancers. We show here, how data generated from next-generation exome sequencing can also be used to estimate copy number variations (CNV), thereby providing a deep comprehensive molecular profiling of the samples.

CNV identification from exome sequencing data (EXO-CNV) is based on the ratio of normalized reads in test vs. reference samples. In addition, we determine allele balances of single nucleotide variants, paired-end alignment anomalies and anomalously aligning read ends. Our approach can be applied to both individually enriched samples as well as to indexed samples that are captured simultaneously (comparative genomic capture). A copy number ratio was generated for each captured exon by RPKM-based normalization and calculating the ratio of these normalized signals in the test and reference samples. For individually captured samples, reference signals were counted as mean signals from multiple exome-capture experiments.

To enhance CNV break point identification, we determined the positions where multiple read ends did not align and were truncated from the same nucleotide (read end anomalies, REAs). In addition, allelic imbalances of SNPs and read pairs with unusual insert size or alignment orientation were determined.

Our method easily identified large chromosomal deletions and duplications, including trisomy of chromosome 21 from a single lane of Illumina GAIIx analysis (60 M reads). Using allele balance and REA information, a complex 580/860 kb hemi-homozygous aberration was diagnosed, where two deletions of different sizes overlapped. EXO-CNV is powerful in detecting small CNVs, theoretically down to a single-exon size, provided that there are sufficient numbers of reads.

In summary, we have developed EXO-CNV as a new approach to determine copy number variation from routinely produced exome-sequencing data, thereby providing a new comprehensive opportunity to identify pathogenetic changes in disease.

267

Development and biological validation of a high-throughput long non-coding RNA gene expression platform. *J. Hellemans[1], P. Mestdagh[2], S. Lefever[2], F. Pattyn[2], F. Speleman[2], J. Vandesompele[1,2].* 1) Biogazelle, Technologiepark 3, 9052 Zwijnaarde, Belgium; 2) Center for Medical Genetics Ghent, UZ Gent, De Pintelaan 185, 9000 Gent, Belgium.

Long non-coding RNAs (lncRNA) are an underexplored class of non-coding RNAs shown to be implicated in human health and disease, and expected to constitute a new class of biomarkers. In addition, they open up a new approach to understanding the function and organization of the genome. The lack of a high-throughput platform to reliably quantify lncRNAs has hampered their study so far. To address this, we developed a new platform for sensitive, specific and high throughput qPCR analysis of lncRNAs. In the pilot phase, we designed qPCR assays for 1200 human lncRNAs using state-of-the-art in silico quality controls, followed by an extensive empirical validation according to MIQE guidelines. Assays that did not meet the following quality control criteria were redesigned and validated: 1. specific amplification on genomic DNA according to melt curve, amplicon size and sequence analysis 2. amplification efficiency in the 90%-110% range based on standard curves that span 6 orders of magnitude Three levels of validation on biological samples were conducted. Firstly, the micro array quality control study (MAQC) samples and performance assessment were used to verify whether the platform enables accurate and precise expression analysis. The results objectively demonstrated the power of the screening platform to reliably detect small expression differences in lncRNAs. In a second study, 8 different normal human tissues were profiled for lncRNA expression levels. The global mean normalization method of qbase^{PLUS} was used for optimal removal of technical variation between samples. A set of lncRNAs with clear tissue specific expression patterns were detected based on a broad and agnostic screening. A literature survey allowed us to confirm the tissue specific expression of HULC and GOMAFU in liver and brain, respectively. A third biological proof of concept study involved a lncRNA expression profiling in a TP53 gene perturbation model using the anti-cancer nutlin-3 compound in neuroblastoma cells. This study revealed numerous lncRNAs with TP53 regulated expression levels. Interestingly, the TP53 dependant downregulation of CDKN2BAS, may hint for a novel TP53 pathway element. In conclusion, lncRNAs clearly constitute an important new class of regulatory elements. We have developed a high throughput, low-volume qPCR platform for the quantitative detection of lncRNAs. This tool offers a unique way to investigate the expression patterns of lncRNAs in health and disease.

268

Use of *PiggyBac*-mediated transient transgenic RNAi expression for rapid characterization of mammalian gene function. *B.C. Bjork[1,2], D.R. Beier[1].* 1) Genetics Division, Medicine Dept., Brigham & Women's Hospital/ Harvard Medical School, Boston, MA; 2) Dept. Of Biochemistry Midwestern University Downers Grove, IL.

The production of targeted mutations in mice is the gold standard for analysis of gene loss-of-function in mammals. However, even with the emergence of large-scale knockout mouse resources, generation of mutants using ES cells requires substantial time and resources. In particular, this approach is inefficient for high throughput applications. For instance, linkage and association studies for mutations or strain-specific traits may yield genomic intervals containing a large number of positional candidate genes. Similarly, microarray analyses typically result in lists of differentially expressed genes, with no indication of which ones may be key regulators. Even in highly resolved human GWAS studies, multiple candidate loci may exist that require an organismal (vs. a cellular) assay. An efficient method to rapidly screen genes *in vivo* would facilitate the functional analysis of mammalian genes. RNA interference (RNAi) is a powerful strategy for studying the phenotypic consequences of reduced gene expression. To develop a method for the rapid characterization of the developmental consequences of gene dysregulation, we tested the use of RNAi for “transient transgenic” knockdown of mRNA in mouse embryos. These methods included lentiviral infection as well as transposition using the *Sleeping Beauty* and *PiggyBac (PB)* transposable element systems. Of the three methods tested, the *PB* transposon system produced high numbers of transgenic embryos with the expected phenotype, demonstrating its utility as a screening assay. This approach is attractive as an alternative to gene targeting in ES cells, as it is simple and yields phenotypic information in a matter of weeks. Furthermore, we found the efficiency of *PB* for transgenesis was routinely over 65%, and we have adapted this approach for other applications including multi-gene knock-down, expression of affinity-tagged proteins, and analysis of presumptive transcriptional regulatory sequences. Perhaps the most exciting aspect of this technology is the potential to develop a method modeled on an experimental strategy commonly used in zebrafish. Specifically, if a gene has a phenotype when knocked down using morpholino injection, the effect of mutants of the same gene, or the complementary effects of different genes, can be tested by co-injection of an expression construct. Our preliminary studies and published work demonstrate this is feasible in a mammalian system, and additional optimization is in progress.

269

Functional analysis of chimeric genes in schizophrenia. *C. Rippey[1], M. Cahill[3], A. Nord[1], T. Walsh[2], M. Lee[2], P. Penzes[3], J. McClellan[4], M-C. King[1,2].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Department of Medical Genetics, University of Washington, Seattle, WA; 3) Department of Physiology, Northwestern University, Chicago, IL; 4) Department of Psychiatry, University of Washington, Seattle, WA.

Our lab and others have shown that individually rare, gene-disrupting copy number variants (CNVs) are more prevalent in individuals with schizophrenia than in unaffected controls. However, very few individual events are known to be pathogenic. A subset of CNVs connects components of two separate genes into novel, often deleterious chimeric genes. We hypothesize that some rare CNVs create brain-expressed chimeric genes that contribute to schizophrenia pathogenesis, and that we can gain insight into the pathways involved in this process by studying the effects of these chimeric genes on neuronal cells. To test these hypotheses, we screened DNA from 124 individuals with schizophrenia using the Nimblegen HD2 aCGH platform. We scanned genome-wide for CNVs greater than 30kb in length, present in at least one case and no controls (n=~240, run on the same platform), and not coinciding with events reported in public databases. From these events, we selected only those predicted to delete or duplicate the 5' end of one gene and the 3' end of another, as these CNVs should produce chimeras. We detected four such events in cases. We confirmed the breakpoints of each of these by PCR and sequencing, and tested whether a stable chimeric transcript is present in the patients' lymphoblasts. We also confirmed that in all cases, at least one of the two genes in the chimera was expressed in brain, by RT-PCR on human brain RNA. In addition, when antibodies were available, we performed immunostaining of cultured cortical neurons to ascertain the localization of the genes composing the chimeras. To study the effect of these chimeric genes in neuronal cells, we generated epitope-tagged expression vectors containing both the full-length genes involved in each fusion, and the novel chimeric genes. We then transfect these into cultured cortical neurons to evaluate their effect on protein stability and localization, cell survival, and neuronal morphology and function. In conclusion, we have detected rare CNVs that lead to stable expression of chimeric mRNA in individuals with schizophrenia. We propose that these events are likely to contribute to schizophrenia by causing aberrant expression of fusion proteins in the brains of these patients. We have begun to characterize the functional effect of these fusion proteins in neuronal cells. This represents a novel genetic mechanism for major mental illness, as well as implicating new genes and pathways in schizophrenia.

270

A Comprehensive Study Identifies Multiple Genetic Variants for Schizophrenia. *E. Van den Oord[1], K. Aberg[1], Y. Liu[2], J. Bukszár[1], J. McClay[1], A. Khachane[1], O. Andreassen[3], D. Blackwood[4], A. Corvin[5], S. Djurovic[6], H. Gurling[7], R. Ophoff[8], C. Pato[9], M. Pato[9], B. Riley[10], B. Webb[10], K. Kendler[10], M. O'Donovan[11], N. Craddock[11], G. Kirov[11], M. Owen[11], D. Rujescu[12], D. St Clair[13], T. Werge[14], C. Hultman[15], L. Delisi[16], P. Sullivan[2].* 1) Center for Biomarker Research and Personalized Medicine, Virginia Commonwealth Univ, Richmond, VA; 2) Department of Genetics, University of North Carolina at Chapel Hill, NC, USA; 3) Division of Mental Health and Addiction, Institute of Clinical Medicine, University of Oslo and Oslo University Hospital; 4) University Edinburgh, Division of Psychiatry, Royal Edinburgh Hospital, UK; 5) St James Hospital, Trinity Center for Health Sciences, Department Psychiatry, Ireland; 6) University Oslo, Institute of Psychiatry, Oslo, Norway; 7) University College of London, Windeyer Institute for the Medical Sciences, London, England; 8) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, UCLA, Los Angeles, CA, USA; 9) University of Southern California, Center for Genomic Psychiatry, Keck School of Medicine, Los Angeles, CA, USA; 10) Virginia Inst Psychiatry & Behavioral Genetics, Richmond, Virginia Commonwealth University, VA, USA; 11) Center for Neuropsychiatric Genetics and genome research, Department Psychological Medicine & Neurology, Cardiff, Wales; 12) Department Psychiatry, Division of Molecular and Clinical Neurobiology, University Munich, Munich, Germany; 13) Institute of Medical Sciences, University of Aberdeen, Aberdeen, UK; 14) Copenhagen University Hospital, Research Institute for Biological Psychiatry, Roskilde, Denmark; 15) Karolinska Institute, Department Medical Epidemiology & Biostatistics, Stockholm, Sweden; 16) VA Boston Healthcare System, Brockton, MA, USA.

Schizophrenia (SCZ) is one of the most devastating psychiatric conditions. As few genes have shown robust associations, we present one of the most comprehensive SZC genetics studies to date aimed at identifying susceptibility loci. We first performed a meta-analysis of 18 SCZ genome-wide association studies (GWAS) that involved, after stringent QC, 1,085,772 SNPs and 11,185 cases plus 10,768 controls. Second, we integrated the meta-analysis results with six large databases (linkage meta-analysis, brain transcriptome meta-analysis, SZC candidate gene database, OMIM, expression QTL database) that contained SCZ-relevant information to select 9,300 SNPs for genotyping in independent, family-based samples. After QC, 8,132 SNPs were available for 6,298 individuals from 1,811 families. Overall, our replication results showed considerable enrichment of small p-values in subjects of European, Asian, and African ancestry (fold increase of median p-value equaled 1.19, 1.12, and 1.07, respectively) as well as in the combined analyses across these groups (fold increase median p-value 1.15). We were able to replicate TCF4 and NOTCH4 that are considered to be among the most promising SCZ candidate genes. The top novel SNPs from our replication study had p-values in the 10-4-10-6 range and could all be linked to the immune system. Multiple, but not all, of these SNPs were located in the MHC region that showed a 3.7 fold overall enrichment of p-values smaller than 0.01. Our results confirmed the polygenetic nature of SCZ and provided a critical next step in understanding SCZ pathophysiology and identifying the urgently needed new drug targets.

271

Gene discovery for schizophrenia by exome sequencing of multiply-affected consanguineous Palestinian families. *H. Shahin[1], C. Rippey[2], T. Walsh[2], M. Lee[2], U. Sharaha[1], I. Ikhmayyes[3], I. Banoura[3], J. McClellan[4], M. Kanaan[1], MC. King[2].* 1) Dept Life Sci, Bethlehem Univ, Bethlehem, Palestinian Territory, Occupied; 2) Department of Medical Genetics, University of Washington, Seattle, WA; 3) Bethlehem Mental Hospital, Bethlehem, Palestine; 4) Department of Psychiatry and Behavioral Sciences, UW School of Medicine, University of Washington, Seattle, WA.

Schizophrenia is a debilitating disease that affects nearly 1% of the population. Family, twin and adoption studies all support a strong genetic component for schizophrenia. However, the genetic causes of schizophrenia remain elusive. In the Middle East and in the Palestinian population, in particular, traditional marriage patterns have led to high rates of consanguinity. As a result, recessively inherited phenotypes, whether purely Mendelian or complex, are common. Psychiatrists, geneticists, and epidemiologists have long been interested in the possible association between consanguinity and neurodevelopmental illnesses. Our rationale is that schizophrenia may be caused by rare severe mutations in genes involved in neurodevelopment, and that individuals with schizophrenia in consanguineous, severely affected families may harbor such mutations as recessive alleles. We have ascertained 5 families severely affected with schizophrenia from the Bethlehem MentalHospital, the only mental hospital in Palestine. Medical chart reviews, and when possible interviews, were conducted for all family members who were patients of the hospital. Further psychiatric history was obtained during home visits by a psychiatrist. All affected individuals meet DSM-IV criteria for schizophrenia or schizoaffective disorder. As an initial screening process, we genotyped all affected individuals for the nine large recurrent schizophrenia-associated hotspot CNVs and determined that none were homozygous or heterozygous carriers. We then performed exome sequencing on 18 affected individuals to high median depth of coverage (125x). We generated private and rare SNP and indel variants for each individual by filtering against common variants in publically available databases and 24 exomes that we have sequenced from Palestinian and other Middle Eastern individuals with no history of mental health disorders. Variants that were predicted to truncate proteins or lead to damaging mutations were Sanger sequence validated and genotyped in additional family members to exclude variants that did not segregate with schizophrenia within the families. For consanguineous families we prioritized further evaluation of variants that were predicted to be within large blocks of homozygosity. We have uncovered a number of compelling candidate mutations some in genes which have previously been implicated in idiopathic mental retardation.

272

De novo Copy Number Variants Confer Risk for Early Onset Bipolar Disorder And Schizophrenia. *J. Sebat[1], S. McCarthy[10], V. Vacic[9], K.E. Burdick[2], S. Cichon[6], A. Corvin[11], S. Gary[10], E.S. Gershon[15], M. Karayiorgou[12], J.R. Kelsoe[1,14], O. Krastoshevsky[13], V. Krause[13], E. Leibenluft[3], D.L. Levy[13], A. Malhotra[2], F. McMahon[8], J. Michaelson[1], J. Potash[4], M. Reitschel[7], T. Schulz[5], D. Malhotra[1].* 1) Psychiatry, UC San Diego, La Jolla, CA; 2) Zucker Hillside hospital, NorthShore Long Island Jewish health system, Glen Oaks, NY; 3) Section on Bipolar Spectrum Disorders, Mood and Anxiety Disorders Program, NIMH Building 15K -MSC 2670,Bethesda, MD; 4) The Johns Hopkins Hospital, John Hopkins University, 600 North Wolfe Street, Baltimore, MD; 5) Department of Psychiatry and Psychotherapy, University of Gottingen, Gottingen, Germany; 6) Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany; 7) Central Institute of Mental Health, Square J5, 68159 Mannheim, Germany; 8) Genetic Basis of Mood and Anxiety Disorders, National Institute of Mental Health, NIH, Convent Drive MSC 3719, Bethesda, MD; 9) Department of Computer Science, Columbia University, New York; 10) Stanley Center for Cognitive Genomics, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York; 11) Neuropsychiatric Genetics Research Group, Institute of Molecular Medicine and Department of Psychiatry, Trinity College Dublin, Ireland; 12) Department of Psychiatry, Columbia University, New York- 10027, USA; 13) McLean Hospital, Belmont, Massachusetts 02478; 14) Veterans Affairs San Diego Healthcare System, San Diego, CA; 15) Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL.

Family-based studies of copy number variation (CNV) have demonstrated that rare spontaneous mutations play a role in neuropsychiatric disorders including autism (ASD) and schizophrenia (SZ). However, the relevance of such mutations to bipolar disorder (BD) has not been established. We tested the hypothesis that de novo CNVs are enriched in BD, and we sought to confirm the strong association of de novo CNVs in SZ. Using a microarray consisting of 2.1 million probes, we screened for de novo CNVs ⁄ 10 kb in size in blood derived DNAs from 788 subject-mother-father trios. Diagnoses of subjects included BD (N=185), SZ (N=177) and healthy controls (N= 426). CNVs were subsequently validated by a custom tiling-resolution array comparative genomic hybridization (CGH) platform. A total of 23 de novo CNVs were detected. Mutations were detected in 0.9% of controls and at significantly increased rates in BD (4.3%, P=0.009) and SZ (4.5%, P=0.007). Age-at-onset was earlier in BD patients with de novo CNVs (P = 0.037), but a similar effect was not observed in SZ patients. De novo CNVs had a median size of 112 kb and contained a median of 2 genes. Our findings provide evidence that de novo structural mutations are associated with BD, particularly in cases with an early disease onset. Genes identified in this study may help to elucidate the neurobiological basis of mood disorders and other psychiatric disorders.

273

Small-scale exome-sequencing study followed by large-scale follow-up to detect genetic variants that increase the risk of idiopathic generalized epilepsy. *E. Heinzen[1], E. Ruzzo[1], C. Depondt[2], G. Cavalleri[3], R. Radtke[4], K. Shianna[1], D. Ge[1], C. Catarino[5], G. O'Conner[3], S. Sisodiya[5], N. Delanty[3], D. Goldstein[1], EPIGEN Consortium.* 1) Center for Human Genome Variation, Duke Univ, Durham, NC; 2) Department of Neurology, Hopital Erasme, Universite' Libre de Bruxelles, Brussels, Belgium; 3) Department of Clinical Neurological Sciences, Division of Neurology, Beaumont Hospital, Dublin, Ireland; 4) Department of Medicine (Neurology), Duke University Medical School, Durham, NC; 5) Department of Clinical and Experimental Epilepsy, UCL Institute of Neurology, Queen Square, London, UK.

Next-generation sequencing is a powerful approach for cataloging genetic variants in the human genome, however designing affordable studies with sufficient power to detect causal variants remains difficult in complex diseases with high locus heterogeneity and incomplete penetrance. Idiopathic generalized epilepsy (IGE) is a prevalent and highly heritable epilepsy disorder, yet few variants have been identified that clearly confer disease risk. In this study, we performed a small-scale exome-sequencing study of IGE patients, followed by large-scale follow-up of thousands of candidate variants. This study design enables the identification of a set of candidate variants from sequencing, with the statistical proof of the association of one or more with disease occurring with follow-up genotyping in a much larger cohort of IGE patients and controls. With this objective, we first sequenced the exomes of 118 IGE patients and 243 controls of European ancestry. Following alignment of the sequenced fragments, variant calling, and a series of quality control steps, we identified a set of 3,900 candidate single nucleotide variants predicted to affect the encoded protein product of 3,038 genes that were either enriched or exclusively present in IGE cases. Each of these variants was genotyped in a cohort of 1,000 IGE patients and 2,500 controls using a custom-designed genotyping chip. Variants continuing to show enrichment in cases compared to controls, or exclusively present in cases were evaluated in further cohorts. Genes with higher than expected variant burden compared to controls were selected for targeted gene sequencing in larger cohorts. Consistent with a model of locus heterogeneity in IGE susceptibility, we found no single variant that accounted for a large fraction of the studied cases. However, we did identify a number of variants that were observed in multiple epilepsy cases and absent in controls. Findings of particular interest included a nonsynonymous variant found only in epilepsy patients in PRICKLE1, a gene recently implicated in a specific form of IGE, and multiple rare, coding variants in NDE1, one of many genes included in the 16p13.11 deletion region associated with epilepsy disorders and recently implicated in a microcephaly condition. We continue to explore the utility of this approach in identifying genetic variants that increase the risk of IGE, and its applicability for genetic discovery in other complex diseases.

274

Homozygosity mapping and new generation sequencing identify three childhood onset symptomatic epilepsy candidate genes. *M. Kousi[1], V. Anttila[2, 3], S. Calafato[2], A.J. Coffey[2], E. Jakkula[3], M. Topcu[4], S. Gokben[5], D. Yuksel[6], F. Alehan[7], O. Kopra[1], A. Palotie[2, 3, 8, 9], A.E. Lehesjoki[1].* 1) Folkhälsan Institute of Genetics and Neuroscience Center, University of Helsinki, Finland; 2) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambs, UK; 3) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland; 4) Department of Pediatrics, Hacettepe University Faculty of Medicine, Section of Child Neurology, Ankara, Turkey; 5) Department of Pediatrics, Ege University Medical Faculty, Izmir, Turkey; 6) Dr. Sami Ulus Children's Hospital, Ankara, Turkey; 7) Division of Child Neurology, Baskent University of Medicine, Turkey; 8) Program in Medical and Population Genetics and Genetic Analysis Platform, The Broad Institute of MIT and Harvard, Cambridge, MA, USA; 9) Department of Medical Genetics, University of Helsinki and University Central Hospital, Helsinki, Finland.

The progressive myoclonic epilepsies (PMEs) are a clinically and etiologically heterogeneous group of symptomatic epilepsies characterized by myoclonic and tonic-clonic seizures, psychomotor delay, ataxia and dementia. We aimed to identify novel genes for childhood-onset PMEs in a cohort of 21 Turkish families most of which were consanguineous. Two families had two affected siblings, while the remaining 19 families had a single affected member. The patients were genotyped using Illumina's 610-quad genome-wide SNP array, and the genotypes were analyzed for runs of homozygosity using the PLINK analysis toolset. In the two families, overlapping runs of homozygosity between the siblings identified candidate loci on chromosomes 1, 3, 8, 16 and 17. The total of 484 genes residing in the candidate regions were screened for mutations by targeted next-generation sequencing. Two novel homozygous missense changes were identified to co-segregate with the disease status; one in the TXNDC6 gene within the first family, and one in the USP19 gene within the second family. The changes were not identified in 90 population controls. No additional families with changes in either TXNDC6 or USP19 were identified upon screening of a second cohort of 107 PME-affected children, rendering the clinical significance of these genes unknown. Analysis of the genotype data of the singleton samples revealed a 5 Mb run of homozygosity on chromosome 7, shared by patients in 9 families. Sequencing of the positional candidate genes revealed five distinct disease-associated mutations in a gene encoding a protein possibly related to potassium channel function, in six families. Co-localization studies showed a cytoplasmic distribution for this protein. Western blot analysis, in combination with immunohistochemistry data suggested an almost ubiquitous neuronal expression of the protein throughout the brain. Finally, evaluation of the protein's expression pattern at different time points of the developing mouse brain showed early developmental expression, suggesting a developmental role for the function of this protein. These data establish a disease association between defects in the protein we identified and neurodegenerative symptomatic epilepsy syndromes with onset in childhood. Deeper understanding of the genetic variation underlying the PMEs is crucial in order to improve the diagnostic classification and treatment of patients, as well as for understanding the underlying pathophysiology.

275

Sushi-repeat protein Srpx2 implicated in epileptic disorders of the speech cortex:in utero RNA silencing causes altered development of the rat brain cortex. *J. Cillario[1], N. Bruneau[1], M. Salmi[1], E. Buhler[1], C. Zimmer[2], A. Massacrier[1], F. Watrin[1], R. Cloarec[1], C. Cardoso[1], P. Durbec[2], F. Muscatelli-Bossy[1], A. Represa[1], P. Szepetowski[1].* 1) INMED, INSERM U901, Marseille, France; 2) IBDML, CNRS UMR6216, Marseille, France.

Early developmental alterations of the human brain cortex are increasingly recognized as a major cause for a wide variety of common pathologies such as autism, dyslexia, epilepsy, or language disorders. Mutations in secreted sushi-repeat containing protein SRPX2 cause disorders of the speech cortex that manifest as rolandic seizures, either with speech impairment (verbal dyspraxia; p.N327S) or in the context of abnormal gyration of the speech areas (bilateral perisylvian polymicrogyria; p.Y72S). We showed previously that the genes encoding SRPX2 and its receptor are both regulated by FOXP2 which in turn is mutated in verbal dyspraxia. Whether and how SRPX2 participates in the development of the brain cortex was questioned here. In utero silencing of rat Srpx2 in the developing cortex led to cell-autonomous impaired radial migration of projection neurons that was rescued by concomitant expression of wild-type rat Srpx2 or of human SRPX2 proteins. In contrast, the two mutant SRPX2 pathogenic proteins both failed to rescue the migratory phenotype and acted via loss-of-function (p.Y72S) and dominant-negative (p.N327S) mechanisms, respectively. In this latter case, the mechanism was well explained by the partial gain-of-glycosylation caused by p.N327S. Impaired positioning of projection neurons caused by Srpx2 silencing in utero was associated with post-natal increased susceptibility to epileptic seizures. Srpx2-silenced neurons that failed to migrate properly showed morphological anomalies and displayed absent or decreased velocity ex vivo (time lapse videomicroscopy). How secreted Srpx2 interferes with cell migration is currently being addressed at the molecular level. Altogether our data provide evidence that rat Srpx2 plays an important role in brain cortex development and strongly suggest that the pathophysiological mechanisms of rolandic epilepsy with verbal dyspraxia involve early developmental defects.

276

Recurrent 450 kb deletions in 2q21.1, including brain specific *ARHGEF4*, in patients with ADHD, epilepsy, and neurobehavioral abnormalities. *P. Stankiewicz[1], A.V. Dharmadhikari[1], P. Szafranski[1], J.P. Phillips III[2], V. Hannig[2], M. Williams[2], P.I. Bader[3], S.S. Vinson[4], A.A. Wilfong[5], W.J. Craigen[1], A. Patel[1], W. Bi[1], J.R. Lupski[1,6,7], S.W. Cheung[1], S.-H.L Kang[1].* 1) Dept Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Division of Medical Genetics and Genomic Medicine, Vanderbilt University School of Medicine, Nashville, TN; 3) Northeast Indiana Genetic Counseling Center, Fort Wayne, IN; 4) Developmental Pediatrics, Texas Children's Hospital, Baylor College of Medicine, Houston, TX; 5) Pediatric Neurology, Texas Children's Hospital, Baylor College of Medicine, Houston, TX; 6) Dept of Pediatrics, Baylor College of Medicine; 7) Texas Children's Hospital, Houston, TX.

Attention-deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder in children and adults characterized by inattentive, hyperactive, and impulsive behaviors that may lead to significant psychosocial impairments. Genome-wide association studies identified several loci, and CNVs found in patients with ADHD include both recurrent (16p13.11 duplication) and nonrecurrent (11p14.1 deletions harboring *BDNF*) changes. Genome-wide linkage analyses have revealed two significant susceptibility loci for ADHD at 2q21.1 and 13q12.11. Using clinical array CGH, we have identified a 450 kb recurrent deletion in 2q21.1 in five unrelated families with ADHD, epilepsy, and other neurobehavioral abnormalities from 10,360 samples referred for Chromosomal Microarray Analysis. Patient 1 is an 11-year-old boy with tall stature, obesity, gynecomastia, social phobias, head banging, and spelling difficulty. He also has an ~ 1.1 Mb duplication in 13q33.1. Patient 2 is a 5-year-old girl with dysmorphic features, hyperactivity, and behavioral issues. Patient 3 is a 6-year-old boy with mild developmental delay (DD), ADHD, behavioral problems, seizures, abnormal EEG, mild dysmorphic features, and hirsutism. His family history is positive for bipolar disease and schizophrenia on the father's side and addiction on the mother's side. Patient 4 is a 17-year-old male with ADHD, encephalopathy, and borderline cognitive disability. Patient 5 is a 2-year-old boy with profound DD, intractable epilepsy, and microcephaly. He also has an ~390 kb duplication in 17p13.3. Both CNVs were inherited from an apparently healthy mother. Parental studies are pending for patients 1-4. In addition, DECIPHER (http://decipher.sanger-.ac.uk) patient 2311 has the same loss and presented with aggressive behavior. There was only one case of larger-sized overlapping deletion identified in over 7700 control samples surveyed. The deletion is flanked by large, complex low-copy repeats with directly oriented subunits of 109 kb in size that have 97.7% sequence identity, and harbors five genes: *GBR148, FAM123C, ARHGEF4, FAM168B*, and *PLEKHB2*. The expression of *ARHGEF4* (Rho guanine nucleotide exchange factor 4; also called *ASEF*) is restricted to the brain and may regulate the actin cytoskeletal network, cell morphology and migration, and neuronal function. We suggest that deletion 2q21.1 is pathogenic for ADHD and, because of its small size, might have been missed in other previous genome-wide screening studies.

277

Genome-Wide Association Study of TB in the South African Colored Population: Comprehensive Gene and pathway-based Association Study. *R.E.C. Emile Chimusa[1], M.M Marlo Möller[2], E.G.H Eileen G. Hoal[2], N.M Nicola Mulder[1].* 1) Clinical Laboratory Sciences, University of Cape Town, Cape Town, Western Cape, South Africa; 2) Molecular Biology and Human Genetics, MRC Centre for Molecular and Cellular Biology, DST/NRF Centre of Excellence for Biomedical TB Research, Faculty of Health Sciences, Stellenbosch University, Tygerberg, South Africa.

The Coloured population of Western Cape in South Africa is uniquely admixed of ancestry from various populations include African, European and Malaysian or Indonesian descent, however this mixed population has the highest incidence of tuberculosis (TB) in sub-Saharan of Africa. Both the environmental and migration factors that may be involved in determining susceptibility and resistance to active tuberculosis are difficult to be controlled, however the understanding of the genetic basis of the TB susceptibility and of its multiple genetic factors, is critical for informing the development of novel interventions. Because of the complex nature of the immune system and the polygenic nature of TB, incorporating both the association signal from Genome-wide Association Study (GWAS) and the available human protein-protein interaction (PPI) information for testing the combined effects of SNPs and searching for significantly enriched sub-networks for such a complex disease provides increased evidence to elucidate the genetic susceptibility. Here we conducted the first SNP-based, comprehensive gene and pathway-based Genome-wide association study for TB risk on 692 cases and 91 controls of the South African Coloured population. Through this new paradigm for GWAS, we found evidence of significant association of the (CYP2C9) cytochrome P450, family 2, sub-family C (10q23.33, p= 9.0018e-09) and (DGKH) diacylglycerol kinase eta (13q14.11, 2.7852e-08) genes with susceptibility to TB. Our comprehensive gene and pathway-based association analysis revealed the evidence of 52 highly significant TB-related candidate genes and 10 TB-related canonic pathways. Our present results demonstrate that the combination of SNP-based, Gene-based and pathway-based association in complex diseases such as TB provides increased evidence for identifying novel genes in which each single SNP conferred a small disease risk beyond those could be detected by single SNP analysis from standard GWAS. In addition, our results demonstrated evidence of convergence of the genetic signals to novel sub-networks of the human interactome that are enriched with interesting TB related biological pathways and functional groups associated with TB.

278

Use of a large pedigree from the genetic isolate of Norfolk Island to localize a novel X chromosomal migraine susceptibility locus. *L.R. Griffiths[1], B. Maher[1], H. Cox[1], L.M. Haupt[1], C. Bellis[2], J. Blangero[2], J. Curran[2], R.A. Lea[1].* 1) Genomics Res Ctr, Sch Med Sci, Griffith Health Institute, Griffith University, Southport, Australia; 2) Southwest Foundation of Biomedical Research, Texas, US.

Pedigree linkage studies in our laboratory have previously implicated an X chromosomal migraine gene component. These studies have shown significant linkage of independent migraine pedigrees to chromosome Xq27 and Xq28. The present study has continued to focus on this chromosome by performing a complete X chromosomal scan of a very large migraine pedigree from the genetically isolated population of Norfolk Island. A novel pedigree-based association approach incorporating logistic regression was employed to analyse X chromosome-wide SNP data (15,154 SNPs) in this pedigree. Data was ascertained for 288 related individuals comprising a large core pedigree of the Norfolk Island population. This pedigree is comprised of 76 individuals affected with International Headache Society (IHS) diagnosed migraine, including 25 who had migraine without aura and 51 diagnosed with the migraine with aura sub-type. All SNPs were ranked based on P-values adjusted for the effects of relatedness, gender and age. SNP prioritization showed that the top 25 SNPs contained 13 SNPs localized to previously identified loci on Xq27 and Xq28, adding further support to the involvement of these genomic regions in migraine. In addition, 10 of the 25 top ranked SNPs mapped to a new 377Kb locus at Xq12 with the strongest SNP association being P = 9.67x10-5. Haplotype analysis of this new region highlighted the presence of 2 major haplotype blocks each containing a haplotype that was significantly over-represented in the migraineurs compared to controls (P<0.0005). In conclusion, we have provided strong evidence supporting the presence of a novel migraine susceptibility locus on Xq12 in a large pedigree from the isolated population of Norfolk Island. Investigations are now underway examining these Xq12 haplotypes in unrelated case-control populations to determine whether this locus is unique to Norfolk or influences migraine susceptibility more generally.

279

Novel loci identified by the largest genome-wide association study of lupus to date. *J. Bentham[1], D.L. Morris[1], M.E. Alarcón-Riquelme[2], V. Anand[1], A.M. Delgado-Vega[3], P.R. Fortin[4], J. Martín[5], C.L. Pinder[1], J.D. Rioux[6], J.E. Wither[4], D.S. Cunninghame Graham[1], T.J. Vyse[1], CaNIOS-GenES, BIOLUPUS.* 1) Medical and Molecular Genetics, King's College, London, SE1 9RT, United Kingdom; 2) GENYO, Centro de Genómica e Investigación Oncológica Pfizer, Universidad de Granada, Junta de Andalucía, Avenida de la Ilustración 114, 18007 Granada, Spain; 3) Department of Genetics and Pathology, University of Uppsala, Dag Hammarsjölds väg 20, 75185 Uppsala, Sweden; 4) Toronto Western Hospital, Main Pavilion, 399 Bathurst St, Toronto, Ontario, M5T 2S8, Canada; 5) Instituto de Parasitología y Biomedicina Lopez-Neyra, CSIC, Calle del Conocimiento, 18100 Armilla, Spain; 6) Montreal Heart Institute, Université de Montréal, 5000 rue Belanger, Montreal, Quebec, H1T 1C8, Canada.

Systemic lupus erythematosus (SLE) is an inflammatory autoimmune disease with complex etiology. We have undertaken a GWAS using the Omni1-Quad chip, comprising 4,254 unrelated lupus sufferers and 8,578 controls; this is substantially larger and denser than any study in lupus hitherto. We have also imputed a much denser set of SNPs using the 1000 Genomes data as a reference, providing further evidence of association at novel loci. Our study includes cases and controls from across Europe, allowing us to investigate further the differences in lupus susceptibility loci between northern and southern Europeans.

We confirm the clear signal at the MHC ($p = 1.2 \times 10^{-193}$) and at other known susceptibility loci such as *STAT4* ($p = 8.4 \times 10^{-56}$) and *ITGAM* ($p = 1.5 \times 10^{-45}$). The power of our study has allowed us to identify novel genotyped loci with lower odds ratios and minor allele frequencies than found in previous work. The loci associated with lupus in our study include *NAB1* ($p = 5.8 \times 10^{-16}$, OR = 1.19-1.33) and *GATA4* ($p = 2.8 \times 10^{-8}$, OR = 1.10-1.23), on chromosomes 2 and 8 respectively. As well as these novel autosomal genes, we have identified a new signal on the X chromosome: *NAA10* ($p = 5.2 \times 10^{-11}$, OR = 1.17-1.34). A number of the genes found in our study have been shown to be associated with other autoimmune diseases (e.g., *LRRC18* in rheumatoid arthritis), allowing the interrelationships between these diseases and lupus to be more precisely defined.

Other results include an association with lupus in Europeans for *WDFY4* ($p = 1.9 \times 10^{-9}$, OR = 1.12-1.26), which has been previously associated with SLE in Asians, and the identification of likely false positives from other studies; for example, *IL10* shows no signal in our study. Our study design has allowed us to examine the geographical differentiation (north versus south) in lupus susceptibility loci within Europe. We find that *ARAP2* has a stronger signal in northern Europeans ($p = 2.2 \times 10^{-12}$, OR = 2.50-5.19) than in Europeans as a whole ($p = 2.0 \times 10^{-5}$, OR = 1.37-2.37). The lowest p-value for southern Europeans is at *STAT4* ($p = 2.8 \times 10^{-27}$, OR = 1.68-2.13) rather than the MHC ($p = 3.5 \times 10^{-15}$, OR = 1.89-2.92).

The results of our GWAS provide a dense map of genetic susceptibility to lupus across the genome, and provide exciting new directions for research into the etiology of the disease.

280

Independent rare and common variants underly association signals for celiac disease. *G. Trynka[1], KA. Hunt[2], J. Romanos[1], L. Franke[1], . Coeliac Disease Genetics Cons.[1,2], . WTCCC[5], S. Onengut-Gumuscu[3], RH. Duerr[4], P. Deloukas[5], V. Plagnol[6,7], DA. Van Heel[2], C. Wijmenga[1].* 1) Department of Genetics, University Medical Center Groningen, Netherlands; 2) Blizard Institute, Barts and The London School of Medicine and Dentistry, London, UK; 3) Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA; 4) University of Pittsburgh School of Medicine, Pittsburgh, USA; 5) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK; 6) Genetics Institute, University College London , London, UK; 7) JDRF/Wellcome Trust Diabetes and Inflammation Laboratory, University of Cambridge, Cambridge, UK.

Celiac disease is an autoimmune disease with 1% prevalence in the general population and a complex genetic background. With our latest genome wide association scan (GWAS) on 15,283 case-control samples we established association to 26 non-HLA loci. To refine the association signals and to identify variants conferring stronger effects to disease susceptibility, we performed high-density genotyping using Immunochip, a custom Illumina iSelect platform (196,548 variants) covering the 183 reported immune-mediated disease loci. This platform includes ~30% rare variants with frequencies <5% derived from re-sequencing projects and variants identified in the 1000Genomes-pilot CEU. Each of the celiac disease loci is covered by 467 SNPs on average. We genotyped 12,041 celiac disease cases and 12,228 controls from 7 different populations on Immunochip. We identified 13 new celiac risk loci at genome wide significance, bringing the total number of known loci (including HLA) to 40. We noted an excess of intermediate range p-values at the remaining 148 autoimmune loci, confirming the large genetic overlap between immune-mediated diseases and indicating hundreds of genes underly celiac disease susceptibility. Multiple independent association signals were found at over a third of these loci, attributable to a combination of common (IL12A locus, maf>0.1), low frequency (CTLA4 locus, 0.05>maf>0.005), and rare (SOCS1 locus, maf<0.005) genetic variants. Many signals could be localized to much finer-scale regions than the corresponding HapMap3 CEU recombination blocks, and in some cases to regulatory regions of immune system genes. Investigation of independent populations revealed population-specific high-risk variants (ORs>2). We define a complex genetic architecture of risk regions, and localize risk signals, providing a next step towards elucidating causal disease mechanisms. With these approach we narrowed down the list of 134 genes within 36 currently associated regions to 59 plausible, causal genes.

281

Targeted sequencing of coding regions implicated via GWAS discovers novel rare causal mutations influencing Inflammatory Bowel Disease. *M.A. Rivas[1,2,3], M. Beaudoin[4], A. Gardet[5], C. Stevens[2], Y. Sharma[6], F. Kuruvilla[2], D. Ellinghaus[11], M. Dubinsky[7], S.B. Brant[9], R.H. Duerr[10], D. Altshuler[2], S. Gabriel[2], G. Lettre[4], A. Franke[11], M. D'Amato[8,12], D.P.B. McGovern[13], J.H. Cho[6], J.D. Rioux[4], R.J. Xavier[2,3,5], M.J. Daly[2,3], NIDDK and International IBD Genetics Consortium.* 1) Clinical Medicine, Oxford University, Oxford, United Kingdom; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Analytic and Translational Genetics Unit (ATGU), Massachusetts General Hospital, Boston, MA, USA; 4) Université de Montréal and Research Centre, Montreal Heart Institute, Montreal, Quebec, Canada; 5) Gastrointenstinal Unit, Center for the Study of the Inflammatory Bowel Disease and Center for Computational and Integrative Biology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; 6) Yale School of Medicine, New Haven, Connecticut, USA; 7) The Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA; 8) Karolinska Institutet, Department of Clinical Science Intervention and Technology, Stockholm, Sweden; 9) Meyerhoff Inflammatory Bowel Disease Center, Department of Medicine, School of Medicine, and Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA; 10) Division of Gastroenterology, Hepatology and Nutrition, Department of Medicine, School of Medicine, and Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA; 11) Institute of Clinical Molecular Biology, Schittenhelmstr. 12, D-24103 Kiel, Germany; 12) Karolinska Institutet, Department of Biosciences and Nutrition, Stockholm, Sweden; 13) Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA.

In CD, individual genome-wide association scans (GWAS) have robustly identified over 71 susceptibility loci. In UC, GWAS studies have identified a total of 47 susceptibility loci and in total 99 distinct associations have been documented for IBD. While these new findings have already provided novel insight into disease pathways, the common SNPs identified are generally of modest effect. Moreover, most of the associated variants do not have any known function and many implicate regions with multiple genes, limiting biological extrapolation.

In this study we set out to identify rare variants in coding exons within GWAS regions for Crohn's disease. Should independently associated rare coding variation with higher statistical and biological effects be discovered in a gene within a region implicated by GWAS, the gene and variants become directly implicated. Targeted resequencing in the coding exons of 56 genes in 350 cases and 350 controls in regions associated to Crohn's disease and follow-up genotyping of 70 variants in over 40,000 samples identifies over 10 independent rare variants associated to Inflammatory Bowel Disease. We identify five additional independent risk factors in NOD2, two additional protective variants in IL23R that are independent of, but demonstrate similar effects as previously defined coding variants, highly significant association is seen for a novel splice variant in CARD9 with a strongly protective effect, and additional rare variants in IL18RAP, splice site in CUL2, PTPN22, C1orf106, and MUC19.

The results of this experiment are highly relevant to ongoing debates in human genetics. While we found little support for the hypothesis that common variant associations are simply an indirect LD-driven byproduct of higher-penetrance rare alleles, additional independent acting rare alleles in genes implicated by common variant association are documented. The value of these results is likely much more in the realm of functional biology than in nudging the tally of variance explained marginally forward. In addition to the functional confirmation of NOD2 alleles, the identification of a novel CARD9 isoform that is strongly protective against disease development provides a concrete handle with which to study disease biology and potentially a model that could be mimicked therapeutically.

282

Genetic pleiotropy in the inflammation-related quantitative trait loci. *Y. Okada[1,2], A. Takahashi[1], M. Kubo[3], Y. Nakamura[4], K. Yamamoto[2], N. Kamatani[1].* 1) Lab for Statistical Analysis, CGM, RIKEN, Tokyo, Japan; 2) Department of Allergy and Rheumatology, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan; 3) Lab for Genotyping Development, CGM, RIKEN, Yokohama, Japan; 4) Lab of Molecular Medicine, HGC, IMSUT, Tokyo, Japan.

Pleiotropy is a phenomenon that a single gene influences multiple phenotypes, and recent development of genome-wide association studies (GWAS) revealed that a number of genetic loci had pleiotropic effects. Since the locus with pleiotropy could have broad spectrums of impacts, the investigation for them would contribute to our understanding of genotype-phenotype etiologies.

Here, we report pleiotropy identified through GWAS for inflammation-related quantitative trait loci (QTL). We enrolled around 30,000 unrelated Japanese individuals, and performed QTL-GWAS for white blood cell subtype counts (neutrophils, lymphocytes, monocytes, basophils, and eosinophils) and serum C-reactive protein (CRP) levels. Associations of the genetic loci identified through the GWAS with other hematological and biochemical phenotypes were further evaluated.

Through the GWAS, we identified 15 significantly associated loci that satisfied $P < 5\times10^{-8}$. Of these, 10 loci were novel (the *CDK6* locus for neutrophils, the *ITGA4*, *MLZE*, *STXBP6* loci and the MHC region for monocytes, the *SLC45A3-NUCKS1*, *GATA2*, *NAALAD2*, *ERG* loci for basophils, and the *IL6* locus for CRP). We observed a variety of patterns of pleiotropic associations within these phenotypes, or with other hematological and biochemical phenotypes. For example, the *HBS1L-MYB* locus demonstrated significant associations with all of the evaluated hematological phenotypes (white blood cell subtypes, red blood cell-related phenotypes, and platelets), suggesting its substantial role in general hematopoiesis. On the other hand, the *GATA2* locus was solely but strongly associated with basophils and eosinophils, suggesting its contribution on allergic reactions.

One issue raised in the analysis on pleiotropy is inherent correlations among phenotypes. Regarding the pleiotropic associations observed in the *IL6* locus on multiple hematological and biochemical phenotypes, we adjusted the observed associations considering the correlations among the phenotypes. Significant pleiotropic associations were still observed after the adjustments, which demonstrated the pivotal and independent roles of the *IL6* locus on the regulations of inflammation.

Our study would enhance our knowledge for pleiotropy and the etiology of inflammation.

283

Novel loci identified for osteoarthritis. *K. Panoutsopoulou on behalf of the arcOGEN Consortium and the replication studies.* Human Genetics, Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

To enable a well-powered genome-wide association study (GWAS) for osteoarthritis (OA), we have formed the arcOGEN Consortium, a UK-wide collaboration, and have performed the largest GWAS to date in 7,410 knee and/or hip OA cases and 11,009 controls. OA cases were ascertained based on radiographic evidence of disease (Kellgren-Lawrence grade⁄ 2) or clinical evidence of disease to a level requiring total joint replacement (TJR) and were genotyped on the Illumina Human610 platform. The study used two different types of controls: 11,009 population-based, unrelated UK controls which came from 5 distinct sources (the 1958 Birth Cohort and the UK Blood Donor Service from the WTCCC2 study, the 1958 Birth Cohort from the T1DGC study, the ALSPAC study and the PoBI study) genotyped on a variety of Illumina platforms; and 1,828 unrelated, disease-free controls from the TwinsUK cohort (females only) genotyped on the Illumina Human610 platform. Analyses were also performed stratifying the cases by severity, i.e., total joint replacement (TJR) and by joint and gender. We took forward 129 signals with $p<10^{-5}$ for in silico replication in 5 other GWAS (deCODE, EGCUT, GARP, Rotterdam RS1 and RS2 cohorts, and TwinsUK comprising a total of 5,064 cases and 40,619 controls) and based on these results we prioritized 26 SNPs with $p<5x10^{-6}$ for de novo replication in 2,409 additional arcOGEN cases and 2,319 population-based controls. Following large-scale replication we have identified 5 loci to be associated with OA at the genome-wide significance threshold ($p<5x10^{-8}$). The most highly associated SNP, rs6976 in the TJR analysis (OR for allele T: 1.12 [1.08-1.16] $p=7.2x10^{-11}$) is located in chr3p21.1, a gene-rich region. Three other genome-wide significant signals (rs4836732, rs9350591 and rs10492367) lie in regions/genes with no apparent role in OA, while rs835487 (association of G allele with total hip replacement: OR=1.13 [1.09-1.18] $p=1.6x10^{-8}$) is located within chondroitin 4 sulfotransferase 11, involved in chondrocyte development during cartilage morphogenesis. We also report association of rs8044769 at the *FTO* locus with OA in females, $p=6.85x10^{-8}$. It appears that the *FTO* SNP exerts its effect on OA through obesity since the signal is attenuated after adjustment for BMI. We have identified five novel OA loci conferring modest risk for OA. Functional studies will be needed to identify the underlying causative genes and understand their involvement in the aetiology of OA.

284

Genome-wide association study finds 9 susceptibility loci for Dupuytren's disease and suggests a major role for WNT-signaling. *G.H.C.G. Dolmans[1,2], P.M.N. Werker[1], H.C. Hennies[3,4,5], D. Furniss[6,7], E.A. Festen[2,8], L. Franke[2], K. Becker[3,4], P. van der Vlies[2], B.H. Wolffenbuttel[9], S. Tinschert[10], M.R. Toliat[3], M. Nothnagel[11], A. Franke[12], N. Klopp[13], H.-E. Wichmann[13,14,15], P. Nürnberg[3,4,5], H. Giele[6,7], R.A. Ophoff[16,17], C. Wijmenga[2].* 1) Plastic Surgery, University Medical Center Groningen, Groningen, Groningen, Netherlands; 2) Department of Genetics, University Medical Center Groningen and University of Groningen, the Netherlands; 3) Cologne Center for Genomics, Division of Dermatogenetics, University of Cologne, Germany; 4) Cluster of Excellence on Cellular Stress Responses in Aging-associated Diseases, University of Cologne, Germany; 5) Center for Molecular Medicine, University of Cologne, Germany; 6) Nuffield Department of Surgical Sciences, University of Oxford, John Radcliffe Hospital, Oxford, UK; 7) Department of Plastic and Reconstructive Surgery, West Wing, John Radcliffe Hospital, Oxford, UK; 8) Department of Gastroenterology, University Medical Center Groningen and University of Groningen, the Netherlands; 9) Department of Endocrinology, University Medical Center Groningen and University of Groningen, the Netherlands; 10) Institute of Clinical Genetics, Faculty of Medicine Carl Gustav Carus, Dresden University of Technology, Dresden, Germany; 11) Institute of Medical Informatics and Statistics, Christian-Albrechts University, Kiel, Germany; 12) Institute of Clinical Molecular Biology, Christian-Albrechts University, Kiel, Germany; 13) Institute of Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany; 14) Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität and Klinikum Grosshadern, Munich, Germany; 15) Klinikum Grosshadern, Munich, Germany; 16) Department of Medical Genetics and Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, the Netherlands; 17) Center for Neurobehavioral Genetics, University of California, Los Angeles, USA.

Purpose Dupuytren's disease (DD) is a benign fibromatosis of the hands and fingers, often leading to the patients being unable to straighten affected fingers fully. We hypothesized that the susceptibility to DD is influenced by multiple genetic and environmental factors. We performed a genome-wide association study to identify DD susceptibility genes and further understanding of its pathogenesis. Methods The genome-wide association study included 960 Dutch DD cases and 3,117 controls. The top-35 associated single nucleotide polymorphisms (SNPs) were replicated in three independent cohorts, comprising 1,365 DD cases and 8,445 controls from Germany, the UK and the Netherlands. Results Initially we observed genome-wide significance for eight SNPs at three loci. After replication and joint analysis of 2,325 DD cases and 11,562 controls, we had 11 SNPs from nine different loci that showed genome-wide significance. Six DD loci contain genes known to be involved in the WNT signaling pathway. Conclusions This study implicates nine different loci involved in the genetic susceptibility to DD. The presence of WNT signaling pathway genes in six of the nine loci suggests that it is likely to be a key player in the fibromatosis process observed in DD.

285

Rare coding mutations and risk for early-onset myocardial infarction: an exome sequencing study of >2,000 cases and controls. *R. Do on behalf of the NHLBI's Exome Sequencing Project - Early-onset Myocardial Infarction.* Center for Human Genetics, Massachusetts General Hospital, Cambridge, MA.

Myocardial infarction (MI), the leading cause of death in the U.S., is a heritable phenotype and the role for inheritance is greatest when MI occurs early in life. Whereas genome-wide association studies (GWAS) have identified at least 30 common variants associated with MI, the modest proportion of overall heritability explained suggests that variants low in frequency (0.5% to 5% frequency) or rare (<0.5% frequency) may contribute to risk for early-onset MI (EOMI). To test the hypothesis that rare coding mutations contribute to EOMI risk, we are sequencing the protein-coding regions - the exome - of ~1,100 cases with EOMI (men ≤ 50; women ≤ 60) and ~1,100 controls free of MI. Using next-generation sequencing, we have targeted 32.7 megabases at 188,260 exons from 18,560 genes. In the first 970 exomes sequenced, we have generated ~6 billion bases of sequence per individual. Each targeted base was read 146 times on average, and for each individual ~87% of all bases were covered with at least 20x depth. We performed burden-of-rare-variant tests, single SNP association tests, and imputed exomic variants into completed MI GWAS datasets. In burden-of-variant tests, we find an excess of rare mutations (all non-synonymous with MAF < 1% (T1) or 5% (T5)) in several genes including CHRM5 (P=0.0001 for T1), DKK2 (P=0.0003 for T5), and LRIG2 (P=0.0002 for T5). In single SNP association tests, we re-discovered a known nonsense mutation in PCSK9 that confers protection against MI (0 in cases; 6 in controls) in 466 cases compared with 504 controls. In imputation using EOMI exomes as the reference panel, we re-discovered the association of a known low-frequency missense SNP in LPA (I4399M, 2% allele frequency, P < 5x10-8). We are replicating findings from the discovery study using three approaches: (1) Sanger sequencing in independent samples (500 cases and 500 controls) of specific genes with signal based on a burden of rare mutations; (2) genotyping of 212 low-frequency SNPs in >10,000 independent MI cases and controls; and (3) imputation of exomic variants into >35,000 MI cases and controls with GWAS data. These replication results should provide insight into the role of rare variants in conferring MI risk and the role of exome sequencing to understand the inherited basis for complex traits.

286

Identification of familial cancer susceptibility genes using whole exome sequencing. *I.G. Campbell, J. Ellul, J. Li, M. Doyle, E.R. Thompson.* Research Division, Peter MacCallum Cancer Centre, East Melbourne, Victoria, Australia.

Recent advances in technology have opened up the possibility of using next generation sequencing to efficiently uncover predisposing mutations in individuals with inherited cancer in an unbiased manner. We are conducting whole exome sequence analysis of germline DNA from multiple affected relatives from breast, sarcoma and colorectal cancer families, respectively, with the aim of identifying rare protein truncating and non-synonymous variants that are likely to include novel cancer predisposing mutations. Data from >100 exomes show that on average each individual only carries 30-50 protein truncating mutations and 300-400 rare non-synonymous missense variants. By considering only those variants shared by multiple affected relatives the number of candidate predisposing mutations can be dramatically reduced to just 3-5 truncating mutations and 10-20 non-synonymous variants per family. Among the first 10 breast cancer families studied in detail, two harbour mutations in known breast cancer genes that were missed by clinical genetic testing either because the index case was a phenocopy who did not carry the mutation (BRCA2) or because the gene is not routinely tested in the context of beast cancer without additional clinical manifestations (PTEN). Among the remaining families candidate genes are currently being assessed for segregation among family members and for prevalence among an additional 800 unexplained breast cancer families. In particular, we found truncating mutations in two genes that are responsible for recognized DNA repair defect syndromes but have not previously been associated with an increased risk of breast cancer. In summary, whole exome sequencing of multiple individuals from within each cancer family is proving to be an efficient strategy for rapidly identifying novel familial predisposing mutations.

287

Genome-Wide Gene-Environment Study Identifies Glutamate Receptor Gene *GRIN2A* as a Parkinson's Disease Modifier Gene via Interaction with Coffee. *H. Payami[1], T.H. Hamza[1], H. Chen[2], E.M. Hill-Burns[1], S.L. Rhodes[3], J. Montimurro[1], D.M. Kay[1], A. Tenesa[6], V.I. Kusel[1], P. Sheehan[1], M. Eaaswarkhanth[1], D. Yearout[1,4], A. Samii[4], J.W. Roberts[7], P. Agarwal[8], Y. Bordelon[9], Y. Park[10], L. Wang[5], J. Gao[2], J.M. Vance[5], K.S. Kendler[11], S. Bacanu[11], W.K. Scott[5], B. Ritz[3,9,12], J. Nutt[13], C.P. Zabetian[4], S.A. Factor[14].* 1) Wadsworth Ctr, NY Dept Hlth, Albany, NY; 2) Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina; 3) Department of Epidemiology, UCLA School of Public Health, University of California at Los Angeles, Los Angeles; 4) VA Puget Sound Health Care System and Department of Neurology, University of Washington, Seattle, WA; 5) Dr. John T. Macdonald Foundation Department of Human Genetics and John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Florida; 6) Institute of Genetics and Molecular Medicine & The Roslin Institute, University of Edinburgh, Scotland; 7) Virginia Mason Medical Center, Seattle, WA; 8) Booth Gardner Parkinson's Care Center, Evergreen Hospital Medical Center, Kirkland, WA; 9) Department of Neurology, UCLA School of Medicine, University of California at Los Angeles, Los Angeles, California; 10) Nutritional Epidemiology Branch, Divisions of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland; 11) Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond VA; 12) Department of Environmental Health Sciences, Center for Occupational and Environmental Health, UCLA School of Public Health, University of California at Los Angeles, Los Angeles; 13) Department of Neurology, Oregon Health & Sciences University, Portland, Oregon; 14) Department of Neurology, Emory University School of Medicine, Atlanta, GA.

Our aim was to identify genes that influence the inverse association of coffee with the risk of developing Parkinson's disease (PD). We used genome-wide genotype and caffeinated-coffee-consumption data on 1458 persons with PD and 931 without PD from the NeuroGenetics Research Consortium (NGRC), and performed a genome-wide association and interaction study (GWAIS), testing each SNP's main-effect plus its interaction with coffee using Kraft's 2df joint analysis, adjusting for sex, age and two principal-components. We then stratified subjects as heavy- or light-coffee-drinkers and performed genome-wide association study (GWAS) in each group. We replicated the most significant SNP. Finally, we imputed the NGRC dataset, increasing genomic coverage to examine the region of interest in detail. The primary analyses (GWAIS, GWAS, Replication) were performed using genotyped data. In GWAIS, the most significant signal came from rs4998386 in *GRIN2A*, which at P_{2df}=1E-6, surpassed all known PD susceptibility genes in significance, including *SNCA* and *MAPT*. Imputation revealed a block of SNPs in *GRIN2A* that achieved P_{2df}<5E-8. *GRIN2A* is an important gene for the central nervous system. Accelerated evolution of *GRIN2A* in primates contributed to the dramatic increase in the size and complexity of the human brain which defines human evolution. *GRIN2A* encodes a subunit of the NMDA glutamate receptor which is central to excitatory neurotransmission and the control of movement and behavior. In stratified GWAS, *GRIN2A* signal was present in heavy-coffee-drinkers (rs4998386 OR=0.43; P=6E-7; imputed SNPs OR=0.41, P=3E-8) but not in light-coffee-drinkers. The a-priori Replication hypothesis that “Among heavy-coffee-drinkers, rs4998386_T carriers have lower PD risk than rs4998386_CC carriers” was confirmed: Replication OR=0.59, P=1E-3; Pooled OR=0.51, P=6.5E-8. Compared to light-coffee-drinkers with rs4998386_CC genotype, heavy-coffee-drinkers with rs4998386_CC genotype had 18% lower risk (P=3E-3), whereas heavy-coffee-drinkers with rs4998386_TC genotype had 59% lower risk (P=6E-13). This study is proof of concept that inclusion of environmental factors can help identify genes that are missed in GWAS. Both adenosine-antagonists (caffeine-like) and glutamate-antagonists (*GRIN2A*-related) are being tested in clinical trials for treatment of PD. *GRIN2A* may therefore be a useful pharmacogenetic marker for prevention and treatment trials and therapy. Study was funded NIH R01 NS36960.

288

Targeted next-generation DNA sequencing of the mitochondrial proteome. *A.G. Compton[1], S.E. Calvo[2,3,4], S.G. Hershman[2,3,4], S.C. Lim[1,5], C. Garone[6,7], E.J. Tucker[1,5], A. Laskowski[1], D.S. Lieber[2,3,4], S. Liu[2], J. Christodoulou[8,9], J.M. Fletcher[10], S. DiMauro[6], D.R. Thorburn[1,5,11], V.K. Mootha[2,3,4].* 1) Murdoch Childrens Research Institute, Royal Children's Hospital, Melbourne, Victoria, Australia; 2) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 3) Department of Systems Biology, Harvard Medical School, Boston, MA; 4) Broad Institute of Harvard and MIT, Cambridge, MA; 5) Department of Paediatrics, University of Melbourne, Melbourne, Victoria, Australia; 6) Department of Neurology, Columbia University College of Physicians and Surgeons, New York, NY; 7) Department of Human Genetics, Universities of Turin and Bologna, Italy; 8) Genetic Metabolic Disorders Research Unit, Children's Hospital at Westmead, NSW, Australia; 9) Disciplines of Paediatrics & Child Health and Genetic Medicine, University of Sydney, NSW Australia; 10) Genetics and Molecular Pathology, SA Pathology, Women's and Children's Hospital, North Adelaide, South Australia, Australia; 11) Victorian Clinical Genetics Services, Royal Children's Hospital, Melbourne, Victoria, Australia.

Inherited defects in mitochondrial oxidative phosphorylation (OXPHOS) are the most common inborn error of metabolism, affecting at least 1 in 5000 live births and predominantly affect organs with high energy consumption. At a biochemical level they are characterized by a defect in the mitochondrial respiratory chain, which generates approximately 90% of the cellular ATP needed to fuel the body. OXPHOS diseases are notoriously difficult to diagnose, as they show extreme clinical heterogeneity. They can present early in infancy or in adulthood, can be severe or mild in presentation, and typically impact multiple organ systems. They are also genetically heterogeneous with both mitochondrial DNA (mtDNA) and nuclear DNA (nDNA) encoding components of OXPHOS. OXPHOS diseases can be caused by mutations in mtDNA and more than 90 known nDNA genes encoding OXPHOS subunits and biogenesis factors. However, the majority of patients with OXPHOS disease do not have a molecular diagnosis. We developed a targeted DNA capture and next-generation sequencing method to detect variants within the mtDNA and exons of 1034 nuclear genes encoding the mitochondrial proteome (MitoExome). We applied this MitoExome sequencing to a group of 40 unrelated patients with clinical and biochemical evidence of severe OXPHOS disease. Sequence variants were filtered by allele frequency, predicted effect on protein function, and compatibility with recessive inheritance. Prioritized variants were followed-up experimentally for evidence of pathogenicity by testing segregation with disease in the family and effect on cellular abundance of mRNA transcripts, protein products or OXPHOS subunits. Molecular diagnoses were provided for 9 patients with mutations in 7 known OXPHOS disease genes (*ACAD9*, *TSFM*, *GFM1*, *POLG*, *BCS1L*, *COX6B1* and *AARS2*). Variants in 16 candidate genes are being investigated as possible new disease genes. MitoExome sequencing shows promise both as a research tool for discovering novel disease loci and as a noninvasive diagnostic approach for patients with clinical evidence of mitochondrial disease.

289

Megabase deletion of the William-Beuren Syndrome critical region reveals remarkable robustness in chromatin interaction landscape. *N. Gheldof[1], M. Leleu[2], J. Rougemont[2], A. Reymond[1].* 1) Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; 2) Bioinformatics and Biostatistics Core Facility, School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), and Swiss Institute of Bioinformatics, Lausanne, Switzerland.

The William-Beuren Syndrome (WBS) is a neurodevelopmental disorder caused by a 1.6 Mb deletion on human chromosome 7. This rearrangement can alter the expression levels of normal-copy flanking genes. The mechanism behind this flanking effect is currently unknown. We hypothesized that long-range chromatin interactions involved in regulation of these genes can be affected by deletion of the critical region. We investigated any changes in looping conformation of six non-hemizygous flanking genes that were differentially expressed in cells from a WBS patient compared to cells from a healthy individual by using an adaptation of the Chromosome Conformation Capture on chip technique (3C-seq). In normal cells, we observed significant interactions between the six genes, as well as interactions with the WBS critical region. We also identified new interacting partner genes. For example, WBSCR17, a gene mapping to the flank of the critical region, and involved in “tameness” in dogs was found to interact with the six genes. We propose that the WBS region forms a multi-looped chromatin globule with these genes, which are co-regulated. In WBS cells, the interaction profile of the six genes is similar, but we detected significant changes in looping intensity. Not only changes in shorter-range interactions within the chromatin globule were observed, but also changes in very long-range interactions, extending along the entire length of the chromosome. Overall, comparison of the interaction profiles between normal and WBS cells suggest that the structural rearrangement did not result in broad chromatin reorganization, where many new interactions are formed, but rather in distinct quantitative changes in interaction frequency. These changes occur locally within the globule of colocalizing genes, thereby possibly jointly affecting their gene regulation. Also chromosome-wide changes in interaction frequency are detected, suggesting that expression might be affected more globally. The 3C-seq method provides a mean to find clusters of genomic regions or genes that are co-regulated, and might thus be useful in discovering potential novel contributors to the phenotypic variation of disorders involving genomic rearrangements.

290

Variation in chromatin accessibility is a key determinant of heritable variation in gene expression. *J. Degner[1,2], A. Pai[1], R. Pique-Regi[1], J.B. Veyrieras[1,3], D. Gaffney[1,4], J. Pickrell[1], S. De Leon[4], K. Michelini[4], N. Lewellen[4], G. Crawford[5,6], M. Stephens[1,7], Y. Gilad[1], J. Pritchard[1,4].* 1) Human Gen, Univ Chicago, Chicago, IL; 2) Committee on Genetics, Genomics and Systems Biology, University of Chicago; 3) BioMiningLabs, Lyon, France; 4) Howard Hughes Medical Institute, University of Chicago; 5) Institute for Genome Sciences and Policy, Duke University; 6) Departments of Pediatrics, Division of Medical Genetics, Duke University; 7) Department of Statistics, University of Chicago.

JD, AP, and RPR contributed equally.

While it is well established that variation in gene expression levels can be influenced by single nucleotide polymorphisms (SNPs), little is known about the regulatory mechanisms by which this most often occurs. To address this gap, we used DNaseI sequencing to measure genome-wide chromatin accessibility in 70 Yoruba lymphoblastoid cell lines (LCLs), for which genome-wide genotypes and expression levels from RNA-sequencing are also available. We obtained a total of 2.8 billion uniquely mapped DNase-seq reads, which allowed us to produce genome-wide maps of chromatin accessibility for each individual. We identified 4,947 locations at which DNase-seq read depth correlates significantly with variation at a nearby SNP or indel (FDR=10%). We call such variants “chromatin accessibility Quantitative Trait Loci” (or “caQTLs”). Most caQTLs lie within or very near to the target DNaseI hypersensitive sites, and they are strongly enriched within inferred transcription factor binding sites. We find that a substantial fraction (15%) of caQTLs are also significantly associated with expression of nearby genes (eQTLs), suggesting that changes in chromatin accessibility or transcription factor binding frequently lead to expression changes. Conversely, 12% of eQTL SNPs are also significant caQTLs. After accounting for incomplete power, we estimate that the true fraction of eQTLs that are also caQTLs is 41%, while the fraction of caQTLs that are eQTLs is 36%. In summary, caQTLs are abundant in the human genome, and are likely to be significant contributors to phenotypic variation and disease.

291

Cell- and individual-specific open chromatin in primary human blood cells. *P. Deloukas[1], K. Voss[2], S. Kanoni[1], C.A. Albers[2], A. Rendon[2], K. Stirrups[1], D. Paul[1].* 1) Human Genetics, Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) Department of Haematology, University of Cambridge, Cambridge, United Kingdom.

Maps of open chromatin provide a powerful means of identifying active regulatory elements across the genome. We applied the formaldehyde-assisted isolation of regulatory elements (FAIRE) method to define open chromatin profiles in primary human monocytes, erythroblasts and megakaryocytes (three individuals per cell type). We systematically assessed the variation of open chromatin signatures across both cell types and individuals. Distinct patterns of open chromatin reflecting the lineage commitment of the three differentiated blood cells could be observed. On average, 10% of the identified open chromatin sites per cell type showed more than four standard deviations variation in normalized peak height across individuals. We have established a generic platform based on FAIRE to guide the interpretation of GWA signals by intersection of cell-selective open chromatin profiles. We first looked at GWA signals from haematological traits and found substantial overlap of open chromatin with GWA signals in our data. As proof of principle we elucidated the molecular mechanism by which the non-coding GWA index SNP rs342293 at chromosome 7q22.3, associated with mean platelet volume and function, modulates platelet phenotype through transcriptional regulation of PIK3CG. We are currently focusing on the functional annotation of genetic loci linked to cardiovascular traits. For coronary artery disease we are integrating fine-mapping data from a large meta-analysis in 22 established CAD loci.

292

A complete haploid human reference genome using a hydatidiform mole resource. *W.C. Warren[1], R. Agarwala[2], T. Graves[2], C. Alkan[3], F. Antonacci[3], M. Dennis[3], B. Fulton[1], S. Shiryaev[2], D.M. Church[2], P. Minx[1], U. Surti[4], M. Nefedov[5], P.J. de Jong[5], E. Mardis[1], E. Eichler[3], R.K. Wilson[1].* 1) Washington University, St Louis, MO; 2) National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD; 3) University of Washington, Genome Sciences, Seattle, WA; 4) Department of Obstetrics, Gynecology, and Reproductive Sciences, University of Pittsburgh School of Medicine, Pittsburgh, PA; 5) Children's Hospital Oakland Research Institute, Oakland, CA.

A hydatidiform mole is simply defined as a pregnancy with no embryo and clinically presents in approximately 1 in 1500 pregnant women in North America. Hydatidiform moles are classified as partial or complete (CHM). CHM have a diploid genome that is a result of the replication of the paternal (sperm) genome, typically XX, creating a haploid equivalent. It is still a mystery as to how and why this process occurs at fairly moderate frequency. With the aim of understanding the biology of CHMs and leveraging the absence of allelic variation, we have begun to construct the first haploid reference genome sequence. We first generated a rich resource of reagents from one characterized complete hydatidiform mole resource (CHM1TERT) including physical maps, genotypes (iSCAN), a large-insert BAC library (CHORI-17) and a corresponding cell line. Next, we isolated genomic DNA from a cell line CHM1 and sequenced to 100-fold coverage with Illumina 100 base length paired sequences. From these sequences we generated an assisted assembly of consensus CHM1 bases with a novel multi-phase aligner and then using a heuristic governed space search with word size 30 overlaps we attempt to fill scaffold gap space. We have completely end-sequenced the BAC library (BES) (170,000 clones), and finished BAC clones totaling 15 Mb for human genome closure efforts, with an additional 35 Mb presently being finished. On chromosome 1, this single haplotype genome has been a critical resource for improvement of the extant reference sequence adding an additional 476 kbp not present in the human reference sequence (GRCh37), discovering missing human genes and revealing large complex patterns of novel structural polymorphism. The consensus sequence derived from CHM1 sequence alignments (assisted assembly) spans 92.2% of the known human reference. Using a read depth approach, we discovered 99 Mb segmental duplications (autosomal, >=20 Kb, n=1207 intervals) and 22.7 Mb deletions (autosomal, >=20 Kb, n=164 intervals). These patterns of copy-number variation are also present in other human genomes but the absence of allelic variation is being used to resolve the structure of these regions for the first time at the sequence level. The CHM1 reference genome we are generating has the potential to not only improve our understanding of this clinical condition but allowing previously intractable regions near acrocentric, centromeric and segmental duplications to be more fully characterized.

293

Identification of three novel genes for Kleefstra syndrome associated phenotypes establishes a chromatin modification module of the *EHMT1* gene network. *T. Kleefstra[1], J. Kramer[1], L.E.L.M. Vissers[1], M.H. Willemsen[1], K. Neveling[1], T. Koemans[1], W.M. Nillesen[1], H. Zhou[1], T. Prescott[2], R.D. Clark[3], W. Wissink-Lindhout[1], H.G. Brunner[1], A.P.M. de Brouwer[1], H.G. Stunnenberg[4], J.A. Veltman[1], A. Schenck[1], H. van Bokhoven[1].* 1) Dept. of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands; 2) Dept. of Medical Genetics, Oslo University Hospital Oslo, Norway; 3) Division of Medical Genetics, Department of Pediatrics, Loma Linda University School of Medicine, Loma Linda, California 92354 USA; 4) Dept. of Molecular Biology, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen, Netherlands.

It has recently become evident that part of the enormous genetic heterogeneity in intellectual disability (ID) can be understood in terms of “modules” in which several genes act together in common molecular complexes or biological pathways. Mutations in genes that perturb epigenetic gene regulation contribute significantly to the prevalence of ID. We aim to identify novel ID genes acting in epigenetic pathways and use these to reconstruct a `chromatin modification module' based on interconnective genotypes and phenotypes. We focus on patients with a phenotype reminiscent of Kleefstra syndrome (KS) [OMIM 610253], which is caused by haploinsufficiency of the Eu-chromatin Histone Methyl Transferase 1 (*EHMT1*) gene in about 25% of the cases. The EHMT1 protein functions as an epigenetic modulator through mono- and dimethylation of lysine 9 histone 3 in euchromatic regions. We have collected a unique cohort of patients with a phenotype compatible with KS without a mutation in *EHMT1*, suggesting the involvement of other genes in the module. Six of these patients were selected for next generation sequencing to identify mutations in novel ID genes. We identified three *de novo* mutations in different genes, all encoding modulators of chromatin structure. To further establish the functional relationships between *EHMT1* and the new genes, we modelled these genes in Drosophila melanogaster. Disruption of at least *EHMT1* resulted in a developmental phenotype, which included impaired neurodevelopment and defects in learning and memory. In addition we will further test out hypothesis and characterize whether the mutant new genes cause comparable phenotypes to the *EHMT1* mutants. Initial genetic interaction studies between *EHMT1* and the novel genes already revealed reciprocal modulation of the mutant phenotypes, which provides further evidence for functional interaction between some of the KS genes. The identification of these novel genes and the elucidation of their functional relationships sheds light on the epigenetic mechanisms involving the *EHMT1* chromatin modification module in neuronal development.

294

Functional evaluation of candidate mutations identified in whole-exome sequencing of patient genomes. *Y. Hitomi, E.K. Ruzzo, K. Pelak, K.V. Shianna, D.B. Goldstein.* Center for Human Genome Variation, School of Medicine, Duke University, Durham, NC.

Whole-genome and whole-exome sequencing are powerful new tools for identifying disease-causing mutations. A central challenge of analyzing human complete sequence data however is distinguishing pathogenic and benign variants, especially given that in many cases variants are being identified in very poorly characterized genes. Here we present analyses of two such cases. In this first, whole-exome sequencing was applied to two unrelated families with children affected with a previously undescribed microcephaly condition. The sequence data identified a nonsynonymous variant in the *asparagine synthetase (ASNS)* gene and we were able to show that this variant appears to result in a loss of protein function. Although *ASNS* mRNA in several mammalian cell line transfectants showed similar expression level, the mutant form of the ASNS protein product was drastically diminished. This difference in protein expression may induce abnormal brain growth caused by amino acid metabolism insufficiency. Next, we also detected an indel which causes frame-shift in *tectonin beta-propeller repeat containing 2 (TECPR2)* gene by whole-exome sequencing of the three unrelated family trios with cerebellar ataxia and developmental delay. In this case, we were able to show that as expected the frame-shift does indeed reduce protein expression. Additional data attempting to elucidate how variations in these genes cause disease is also discussed. These analyses illustrate that straightforward functional work in cellular systems can significantly facilitate the interpretation of whole-genome and whole-exome sequence data.

295

Identification of non-coding mutations in X-linked intellectual disability. *J. Gecz[1,2], L.L. Huang[3], S. Willis-Owen[1,5], M. Field[4], A. Hackett[4], M. Shaw[1], C. Shoubridge[1,2], A. Gardner[1], M. Corbett[1].* 1) Dept Genetics and Molecular Pathology, SA Pathology at Women's & Children's Hosp, Adelaide, South Australia, Australia; 2) School of Paediatrics and Reproductive Health, The University of Adelaide, Adelaide, Australia; 3) Institute of Reproductive & Stem Cell Engineering, Central South University, Changsha, China; 4) GOLD Service, Hunter Genetics, Newcastle, Australia; 5) Imperial College, London, UK.

Intellectual disability (ID) is a highly heterogeneous trait with estimated hundreds if not thousands of genes involved genome-wide. X-chromosome linked intellectual disability (XLID), which represents about 10-15% of ID, has been most extensively studied. So far in excess of 100 XLID genes have been identified. However, these genes resolve no more than 50-70% of the XLID families. In order to address the unresolved families, we have started to investigate the involvement of non-coding and regulatory sequences. We have done this by targeting entire non-repetitive genomic intervals of 2 families with no relevant coding variation. In one unresolved family, known as MRX3 (J Med Genet, 28, 372-7; 1991) we identified regulatory mutation affecting YY1 binding site for the HCFC1 gene. EMSA and luciferase reporter assays demonstrate its functional effect, which as tested on patients cells (LCLs; n=3) lead to HCFC1 mRNA up-regulation. YY1 itself has recently been found mutated in a sporadic case with ID (Nat Genet, 42(12):1109-12, 2010). In collaboration with EURO MRX and IGOLD Consortia we have identified further missense changes in HCFC1, which might compromise its function. In the other, large (>100 individuals) XLID family with non-syndromic intellectual disability (with ~1Mbp minimum linkage interval; LOD>3.0; unpublished data) we have identified at least two small, non-coding deletions. In order to assign functional significance to any of these, we have complemented our inquiry with genome-wide array expression profiling (on LCLs) and found one of the candidate XLID genes (TAF1) significantly down-regulated. On of these, an 8bp deletion was identified within the first intron of the ZMYM3 gene, which has been proposed to be involved in XLID (Hum Mol Genet, 5(7):887-97, 1996). The functional effect of this deletion on ZMYM3 mRNA expression or processing, or its effect on another, neighbouring gene (e.g. TAF1) has not yet been determined. Currently ~86% of all disease causing mutations are in protein coding regions, however, this may represent ascertainment bias, which can now be addressed at least partly by the application of massively parallel sequencing technology in conjunction with the wealth of information provided by projects like ENCODE.

296

Small lethal non-coding copy-number variations in the gene desert region on 16q24.1. *P. Szafranski[1], A. de Klein[2], J. Pinner[3], K.E. Kolodziejska[1], Z. Ou[1], K.N. Mohan[1], M. Chopra[3], G. Peters[4], S. Arbuckle[5], S.F. Guiang[6], V. Hustead[6], R. Hirsch[7], D. Witte[7], C. Langston[8], P. Sen[9], P. Stankiewicz[1].* 1) Dept of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Dept of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands; 3) Dept of Molecular and Clinical Genetics, Royal Prince Alfred Hospital, Sydney, NSW, Australia; 4) Cytogenetics Dept, The Children's Hospital at Westmead, Westmead, NSW, Australia; 5) Histopathology Dept, The Children's Hospital at Westmead, Westmead, NSW, Australia; 6) University of Minnesota Medical Center, Minneapolis, MN; 7) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 8) Dept of Pathology, Texas Children's Hospital, Baylor College of Medicine, Houston; 9) Dept of Pediatrics - Nutrition, Baylor College of Medicine, Houston, TX.

Alveolar Capillary Dysplasia with Misalignment of Pulmonary Veins (ACD/MPV, MIM 265380) is a rare neonatal lethal developmental lung disorder defined by malposition of pulmonary veins adjacent to small pulmonary arteries, medial thickening of small pulmonary arteries, deficient lobular development, a paucity of alveolar wall capillaries, and sometimes lymphangiectasis. Over the last 10 years, we have accumulated the largest collection of ACD/MPV samples (mainly FFPE lung tissues) in the world. In 2009, we reported *de novo* genomic deletions and heterozygous point mutations in *FOXF1* in unrelated patients with sporadic ACD/MPV. We also described deletions located upstream to *FOXF1*, with the gene coding sequence remaining intact, in two patients with ACD/MPV. We have now identified novel differentially-sized overlapping *de novo* deletions upstream to *FOXF1* in five unrelated newborns with histopathologically verified ACD/MPV, who all died in the first month of life with severe respiratory distress and pulmonary hypertension. The deletions enabled us to narrow and define a putative critical regulatory region ~ 85 kb in size located ~ 250 kb upstream to *FOXF1*. Interestingly, among deletions for which parental origin could be determined, six upstream and five others harboring *FOXF1*, each arose on the maternal chromosome. These results strongly corroborate previous observations that *FOXF1* may be paternally imprinted. Bioinformatic analyses of the upstream deletion region revealed an ~ 1.5 kb highly conserved segment with high regulatory potential, harboring a consensus GLI2 transcription factor binding site and a 172 bp long CpG island located 254 bp away from it. Gli2 was previously described to up-regulate the activity of the Foxf1 promoter in mice. Using ChIP-on-chip experiments, we have identified strong GLI2 interactions between these two sites (p=0.007, n=2). Moreover, using bisulfite sequencing and methylation array analyses, we have found that the CpG island is differentially methylated, suggesting that *FOXF1* may be regulated by methylation-dependent binding of GLI2. Further experimments, including reporter assays, chromosome conformation capture (4C), and mouse studies are pending to investigate this hypothesis. Finally, to better estimate the recurrence risk, we have sequenced the deletion breakpoints in five cases and did not find any evidence of low-level somatic mosaicism in parental blood samples using PCR for patient-specific junction fragments.

297

Iron homeostasis modulates the activity of microRNA pathway through PCBP2. *Y. Li[1], L. Lin[1], Z. Li[2], X. Ye[3], B. Aryal[2], Z. Paroo[2], Q. Liu[3], C. He[2], P. Jin[1].* 1) Dept Human Gen, Emory Univ, Atlanta, GA; 2) Dept of Chem and Inst for Biophysi Dyna, The Univer of Chicago, Chicago, IL; 3) Dept of Biochem, Univer of Texas Southwestern Medical Center, TX.

RNA interference (RNAi) is a well-conserved mechanism that uses small noncoding RNAs to silence gene expression post-transcriptionally. Gene regulation by RNA interference (RNAi) has been recognized as one of the major regulatory pathways in eukaryotic cells. Although the major components in RNAi/miRNA pathway have been identified, little is known about the regulation of the RNAi pathway itself. To dissect cellular components modulating RNAi using chemical biology approach, we have previously developed a cell-based assay to monitor the activity of the RNAi pathway. Using this system and 8,000 diversified compounds, we have performed pilot screen, and identified several potent RNAi enhancers. Among them are several structurally distinct small molecules with metal/Fe chelating activity. We found that RNAi-enhancing activity by metal/Fe chelators is iron-dependent, and the level of intracellular iron could modulate the activity of endogenous RNAi/miRNA pathway. Iron (Fe) has been shown involved in many cellular processes. Tight regulation of iron uptake, distribution and export must be achieved to maintain iron homeostasis, which is mediated through transport, storage and regulatory proteins. To determine the protein(s) involved in modulating the activity of RNAi/miRNA pathway, we performed a mini-screen using siRNAs against the genes involved in iron homeostasis, and found that RNAi-enhancing activity by metal chelators requires PCBP2, a cytosolic iron chaperone. Furthermore, we found that PCBP2 is associated with Dicer and can promote the processing of miRNA precursors. Level of cytosolic iron could modulate the association between PCBP2 and Dicer, as well as the multimerization of PCBP2 and its ability to bind to miRNA precursors, which can alter the processing of miRNA precursors. Our results reveal an unexpected link between the RNAi/miRNA pathway and iron metabolism.

298

Genome-wide poly(A) mapping reveals widespread changes in use of polyadenylation sites in Oculopharyngeal muscular dystrophy. *E. de Klerk[1], A. Venema[1], S.Y. Anvar[1], J.J. Goeman[2], J.T. den Dunnen[1], G.B. van Ommen[1], S.M. van der Maarel[1], V. Raz[1], P.A.C. -t Hoen[1].* 1) Leiden University Medical Center, Center for Human and Clinical Genetics, Leiden, The Netherlands; 2) Leiden University Medical Center, Department of Medical Statistics and Bioinformatics, Leiden, The Netherlands.

Regulatory elements in the 3'-untranslated region (3'-UTR) affect stability, translation and nuclear transport of mRNA. Loss or gain of regulatory sequences occurs through alternative polyadenylation (APA) events, which determine the length of the 3'-UTR and hence the level of gene expression. The poly(A) binding protein nuclear 1 (PABPN1) is involved in pre-mRNA polyadenylation, where it stimulates poly(A) polymerase and controls the length of the poly(A) tail. A poly-alanine expansion in the N-terminus of PABPN1 causes Oculopharyngeal muscular dystrophy (OPMD), a late onset muscle disorder characterized by muscle weakness starting in the eyelid and pharyngeal muscles. Although PABPN1 is ubiquitously expressed, the pathological features are muscle specific. A previous microarray expression profiling study from OPMD and control mice provided evidence for a significant number of genes showing differential regulation of transcript isoforms with short and long 3'-UTRs. Here we investigated the alternative usage of polyadenylation sites in OPMD mice at a much higher resolution, employing next generation sequencing. For this, we developed an assay based on the HeliScope single molecule sequencer platform to accurately map poly(A) sites in a strand specific and quantitative way, to reveal differences in overall gene expression and relative use of alternative poly(A) sites. We identified 6,506 transcripts with more than one poly(A) site, corresponding to almost 50% of the detected transcripts. Significant changes in the positions of the most frequently used APA sites were observed for 2,012 transcripts in affected muscles of OPMD mice, demonstrating widespread alterations in 3'UTR length. APA sites switches are almost exclusively resulting in the use of a more proximal poly(A) site, leading to a shortening of the 3'UTR. Changes in APA sites usage overlap significantly with changes in expression levels. Almost 80% of the transcripts showing alternative poly(A) sites usage are differentially expressed, most of them being upregulated. This could be explained by the loss of negative regulatory elements in transcripts with shorter 3'UTR. Therefore we suggest that switches in poly(A) site usage contribute to disturbed gene expression patterns and muscle weakness in OPMD. Our results indicate a possible new role for PABPN1 in poly(A) site selection.

299

Transcriptional activation of lysosomal exocytosis promotes cellular clearance. *D.L. Medina[1], A. Fraldi[1], V. Bouche[1], F. Annunziata[1], G. Mansueto[1], C. Spampanato[1], C. Puri[1], A. Pignata[1], J.A. Martina[1], M. Sardiello[3,4], R. Polishchuk[1], R. Puertollano[2], A. Ballabio[1,3,4,5].* 1) Telethon Institute of Genetics and Medicine (TIGEM), Via P. Castellino 111, 80131 Naples, Italy; 2) Laboratory of Cell Biology, NHLBI, National Institutes of Health, Bethesda, Maryland 20892, USA; 3) Dept. of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 4) Jan and Dan Duncan Neurological Research Institute, Texas Children Hospital, Houston, TX 77030, USA; 5) Medical Genetics, Department of Pediatrics, Federico II University, Via Pansini 5, 80131 Naples, Italy.

Lysosomes are cellular organelles primarily involved in degradation and recycling processes. These organelles are also involved in a secretory pathway known as lysosomal exocytosis. During lysosomal exocytosis, a Ca^{2+}-regulated process, lysosomes are docked to the cell surface and fuse with the plasma membrane (PM), emptying their content outside the cell. This process has an important role in secretion and PM repair. While the main steps of lysosomal exocytosis have been elucidated, little is known about its regulation and how this process is coordinated with lysosomal biogenesis. We recently discovered that lysosomal biogenesis and lysosomal degradative function are transcriptionally regulated by the bHLH-leucine zipper transcription factor EB (TFEB). TFEB activation was able to reduce the accumulation of the pathogenic protein in a cellular model of Huntington disease and ameliorated the phenotype of cells from a murine model of Parkinson disease. Here we show that the transcription factor EB (TFEB) regulates lysosomal exocytosis. TFEB increases the pool of lysosomes in the proximity of the PM and promotes their fusion by raising intracellular Ca^{2+} levels through the activation of the lysosomal Ca^{2+} channel MCOLN1. Induction of lysosomal exocytosis by TFEB overexpression rescued pathologic storage and restored normal cellular morphology both in vitro and in vivo in lysosomal storage diseases (LSDs). Our data indicate that lysosomal exocytosis may directly modulate cellular clearance and suggest a novel therapeutic strategy for disorders associated to intracellular storage.

300

Development and pharmacological rescue of a murine model of primary open angle glaucoma. *G. Zode[1,2], M. Kuehn[3], C. Searby[1,2], K. Mohan[5], S. Grozdanic[5], D. Nishimura[1,2], K. Bugge[1,2], M. Anderson[3,4], A. Clark[6], E. Stone[1,3], V. Sheffield[1,2,3].* 1) Howard Hughes Medical Institute; 2) Departments of Pediatrics,University of Iowa, Iowa, IA 52242; 3) Ophthalmology and Visual Sciences,University of Iowa, Iowa, IA 52242; 4) Molecular Physiology and Biophysics,University of Iowa, Iowa, IA 52242; 5) Department of Veterinary Clinical Sciences, College of Veterinary Medicine, Iowa State University, Ames, IA 50011; 6) Dept. Cell Biology & Anatomy and the North Texas Eye Research Institute, University of North Texas Health Science Center at Fort Worth, TX 76107.

Primary open angle glaucoma (POAG) is a common chronic optic neuropathy, characterized by progressive loss of retinal ganglion cell (RGC) axons with the resultant irreversible loss of vision. Myocilin (MYOC) mutations are the most common genetic cause of POAG. Mechanisms underlying MYOC-associated glaucoma are not fully understood, and a mouse genetic model that closely mimics human POAG has not been developed. Here, we report the development of a transgenic mouse model (Tg-MYOC^{Y437H}), which expresses human MYOC containing the Y437H mutation within relevant eye tissues, and which displays glaucoma phenotypes closely resembling those seen in POAG patients. Tg-MYOC^{Y437H} mice displayed ocular manifestations of glaucoma including elevated IOP (7mm Hg increase vs WT; $p<0.0001$) at 3 months of age, progressive structural and functional loss of RGCs (40% loss by 12 months; $p<0.001$), and optic nerve degeneration (45% reduction in axon count by 12 months; $p<0.001$). We demonstrate that mutant MYOC is not secreted into the aqueous humor and accumulates in the endoplasmic reticulum (ER) of the trabecular meshwork (TM), thereby inducing ER stress in the TM of Tg-MYOC^{Y437H} mice. Chronic and persistent ER stress is associated with TM cell death and elevation of IOP in Tg-MYOC^{Y437H} mice. Reduction of ER stress by topical treatment with the chemical chaperone, phenyl butyric acid (PBA) prevents glaucoma in Tg-MYOC^{Y437H} mice by promoting the secretion of mutant MYOC to the aqueous humor, and by decreasing intracellular accumulation of MYOC in the ER, thus preventing TM cell death. These results demonstrate that ER stress is linked to the pathogenesis of POAG and is a target for treatment in human patients.

301

Modeling the Complex Human Disorder of Tooth Agenesis in Mouse by Combinatorial Reduction in Gene Dosage of *Prdm16*, *Pax9*, and *Msx1*. *I. Saadi[1], B.C. Bjork[1], J.J. Lund[1], X.P. Wang[1], T. Siggers[1], D.J. O'Connell[1], P. Purcell[1], A. Turbe-Doan[1], J.M. Dobeck[2], R. Meira[3], J.R. Avila[4], A. Modesto[5], Z. Skobe[2], M.L. Bulyk[1], A.R. Vieira[5,6], J.C. Murray[4], D.R. Beier[1], R.L. Maas[1].* 1) Department of Medicine, Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2) The Forsyth Institute, Cambridge, MA; 3) Department of Pediatric Dentistry, Brazilian Lutheran University, Canoas, RS, Brazil; 4) Department of Pediatrics, University of Iowa, Iowa City, IA; 5) Department of Pediatric Dentistry, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA; 6) Department of Oral Biology and Center for Craniofacial and Dental Genetics, School of Dental Medicine; Department of Human Genetics, Graduate School of Public Health, and Clinical and Translational Science Institute, University of Pittsburgh, Pittsburgh,.

Oligogenic inheritance of complex diseases such as nonsyndromic tooth agenesis is proposed to result from “subtle” mutations in combinations of genes that function in the same developmental pathway, or that encode functional partners. We posited that analysis of compound heterozygous mice for genes known to play a role in tooth development could reveal a subset of these combinations that are gene-dosage sensitive. As a result, we report the identification of a three-way interaction between *Prdm16*$^{csp1/+}$, *Pax9*$^{+/-}$ and *Msx1*$^{+/-}$ mice that affects incisor development. Rodent incisors have the ability to continuously regenerate due to presence of an endogenous epithelial stem cell compartment called the cervical loop. The cervical loop requires FGF3 and FGF10 signals from the mesenchyme for its induction and maintenance. We identify a dental mesenchymal transcription factor gene, *Prdm16*, which interacts genetically with *Pax9* and *Msx1* to coordinately regulate expression of the FGF and BMP4 signals that induce and maintain this stem cell niche for incisors. Homozygous reduction in *Prdm16*, or heterozygous reductions in *Prdm16* in combination with haploinsufficiency for either *Pax9* or *Msx1* (*Prdm16*$^{csp1/+}$; *Pax9*$^{+/-}$ or *Prdm16*$^{csp1/+}$; *Msx1*$^{+/-}$), all lead to impaired cervical loop formation in lower incisors. In the *Prdm16*$^{csp1/+}$; *Pax9*$^{+/-}$ or *Prdm16*$^{csp1/+}$; *Msx1*$^{+/-}$ compound heterozygotes, this is accompanied by progressive loss of ameloblasts, enamel hypoplasia and eventual breakage and lower incisor loss in adulthood. *Prdm16*$^{csp1/+}$; *Pax9*$^{+/-}$; *Msx1*$^{+/-}$ triple heterozygotes show an even more severe dental phenotype, with embryonic bud stage arrest in lower incisor development, and an upper incisor phenotype that resembles that seen in *Prdm16*$^{csp1/+}$; *Pax9*$^{+/-}$ or *Prdm16*$^{csp1/+}$; *Msx1*$^{+/-}$ lower incisors. In addition, PRDM16 interacts physically with PAX9 and MSX1 *in vitro* and *in vivo*, and their respective DNA binding sites cluster in conserved non-coding genomic regions surrounding the *Bmp4* gene. Collectively, these data establish a role for *Prdm16* during mammalian tooth development and suggest a combinatorial model by which *Prdm16*, *Pax9* and *Msx1* regulate *Bmp4* expression in mouse incisor mesenchyme. Lastly, in humans, specific allelic combinations of *PRDM16* and *PAX9*, and of *PRDM16* and *MSX1*, are over-represented in patients with tooth agenesis, suggesting a potential translational relevance of such genetic modeling in mice.

302

Exome sequencing combined with functional analysis of zebrafish mutants elucidates a novel pathogenesis in high bone mass osteogenesis imperfecta. *K. Keupp[1,2], A. Nair[3], O. Semler[4], Y. Li[1,2], H. Thiele[5], P. Frommolt[5], J. Becker[2], N. Arkasu[6], E. Schoenau[4], P. Nürnberg[5], M. Hammerschmidt[7], T. Carney[3], B. Wollnik[1,2].* 1) Center for Molecular Medicine (CMMC), Cologne, Germany; 2) Institute of Human Genetics, University of Cologne, Germany; 3) Institute of Molecular and Cell Biology, Proteos, Singapore; 4) Children's Hospital, University of Cologne, Germany; 5) Center for Genomics, University of Cologne, Germany; 6) Gene Mapping Laboratory, Department of Medical Genetics, Hacettepe University Medical Faculty, Ankara, Turkey; 7) Institute of Developmental Biology, University of Cologne, Germany.

Osteogenesis imperfecta (OI) is an inherited collagenopathy mainly characterized by dramatically increased bone fragility associated with numerous fractures after minimal trauma. Affected individuals often show a decreased bone mineral density. In our study we investigated a Turkish consanguineous family with autosomal recessive inheritance of OI and affected children suffering from numerous fractures in early childhood. Interestingly, an increased bone mineralisation density with z-scores up to 4.3 was present in our OI patients and the fracture rate dramatically decreased upon bisphosphonate therapy. We performed a whole-exome sequencing strategy in one of the affected individuals and used an innovative filtering strategy to identify the causative gene. Coding sequences were captured using the Agilent SureSelect protocol followed by deep sequencing using an Illumina Genome Analyzer. We used all identified variants throughout the genome to determine homozygous stretches via linkage analysis in this index patient. Subsequently, the homozygous stretches were used as a stringent filter to successfully identify the causative homozygous missense mutation in a gene involved in bone morphogenesis, which has been not yet associated with OI. The mutation was not found in 300 ethnically matched control individuals. Moreover, the mutation is located within the signal peptide of the protein and we provide evidence for an impaired secretion and alteration in post translational modification of the mutant. In order to determine the underlying molecular pathogenesis, we mapped three zebrafish mutants with hypomorphic mutations in a functional domain of the zebrafish ortholog of our novel OI gene. Mutant fish presented with a delayed osteogenesis and defects in bone formation, several fractures in fin rays, and osteopenia in vertebrae, which develops in later larval stages to a significant high bone mass phenotype in these mutants. Detailed analysis of osteoblast function, collagen network structure, and collagen processing determined the underlying pathogenic mechanism responsible for the increase mineralisation and bone fragility. Taken together, we present a novel cause for autosomal recessively inherited osteogenesis imperfecta with a high bone mass phenotype and provide fascinating insights into the underlying molecular mechanism.

303

Recessive Mutations In The Guanidine Nucleotide Exchange Factor DOCK6 Lead to Abnormal Actin Cytoskeleton Organization and Adams-Oliver Syndrome. *R. Shaheen[1], E. Faqeih[2], A. Sunker[1], H. Morsy[3], T. Al-Sheddi[1], HE. Shamseldin[1], N. Adly[1], M. Hashem[1], FS. Alkuraya[1,4,5].* 1) Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia; 2) Department of Pediatrics, King Fahad Medical City, Riyadh, Saudi Arabia; 3) Human Genetics Department, Medical Research Institute, Alexandria University, Alexandria, Egypt; 4) Department of Pediatrics, King Khalid University Hospital and College of Medicine, King Saud University, Riyadh, Saudi Arabia; 5) Department of Anatomy and Cell Biology, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia.

Adams-Oliver syndrome (AOS) is defined by the combination of aplasia cutis congenita (ACC) and terminal transverse limb defects (TTLD). It is usually inherited as an autosomal dominant trait but autosomal recessive inheritance has also been documented. In a patient with autosomal recessive AOS, we combined autozygome analysis with exome sequencing to identify a homoallelic truncating mutation in DOCK6 which encodes an atypical Guanidine Exchange Factor (GEF) known to activate two members of the Rho GTPase family: Cdc42 and Rac1. Another homoallelic truncating mutation was identified upon targeted sequencing of DOCK6 in an unrelated patient with AOS. Consistent with the established role of Cdc42 and Rac1 in actin cytoskeleton organization, we demonstrate a cellular phenotype typical of defective actin cytoskeleton in patient cells. These findings, combined with an expression profile of Dock6 that is consistent with AOS phenotype as well as the very recent demonstration of dominant mutations of ARHGAP3 in AOS, establish Cdc42 and Rac1 as key molecules in the pathogenesis of AOS and suggest that other regulators of these Rho GTPase proteins may be good candidates in the quest to define the genetic spectrum of this genetically heterogeneous condition.

304

Exome sequencing identifies phospholipase B4 (PLCB4) as a gene causing auriculocondylar syndrome (ACS). *M.J. Rieder[1], G.E Green[2], C.M. Cunniff[3], B.D. Stamper[4], S.S. Park[4], J.M. Johnson[7], S.B. Emery[2], T.C. Cox[4,5], A.V. Hing[4,5], J.A. Horst[6], M.L. Cunningham[4,5].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Department of Otolaryngology- Head & Neck Surgery, University of Michigan, Ann Arbor, MI; 3) Department of Pediatrics, University of Arizona, Tucson; 4) Seattle Children's Research Institute, Seattle, WA; 5) Department of Pediatrics, Division of Craniofacial Medicine, University of Washington, Seattle, WA; 6) Department of Pediatric Dentistry, UCSF, San Francisco, CA; 7) Department of Radiology, University of Vermont, Burlington, VT.

Auriculocondylar syndrome (ACS) is an autosomal dominant craniofacial malformation syndrome characterized by highly variable micrognathia, temporomandibular ankylosis, cleft palate, and a distinctive ear malformation consisting of a cleft between the lobule and the pinna giving the appearance of a question mark. In some cases, the mandibular and intraoral phenotype is a distinguishing characteristic that suggests the syndrome may represent maxillary duplication rather than micrognathia. Mouse models with maxillary duplication are due to disruption of the endothelin-1/DLX5-DLX6 pathway, which is regulated via the phospholipase C (PLC)/inositol-triphosphate (IP3) signaling cascade. In this study, we performed exome sequencing, using Illumina second-generation sequencing and Roche/Nimblegen SeqCap EZ v2 capture on three kindreds (duplex cases with parents and two parent-child trios with affected children) and two single cases to identify mutations/genes shared amongst the probands. Variant analysis using discrete filtering failed to identify a single gene that was consistent with either a dominant or recessive model for all cases. Sequencing of parental DNAs in the families, where they were available, led to the identification of two missense changes in *PLCB4* (a phosphatidylinositol-specific C isomer that catalyzes the formation of inositol-triphosphate), one being a *de novo* mutation in a single trio (Tyr623Cys), and a second found in the duplex cases with one mildly-affected parent (Asn329Ser). Analysis of *PLCB4* protein structure shows that these residues are within a highly conserved catalytic domain (calcium and substrate binding sites). Notably, missense mutations in the catalytic domain of the zebrafish homologue plc beta 3 results in micrognathia, fusion to the upper jaw, and patterning suggesting a maxillary `identity'. Subsequently, we performed Sanger sequencing on the *PLCB4* coding region (exons 11-26 containing active sites) in two multi-generation pedigrees showing transmission of ACS distinguishing phenotypes. Each pedigree contained a novel, segregating *PLCB4* mutation at positions Arg621His and Asn650His, respectively. None of the identified *PLCB4* mutations in this study had been identified previously. These results from both small kindred and two large pedigrees identify *PLCB4* as a novel gene for ACS, with a likely mechanism through the PLC/IP3 pathway.

305

Dominant negative mutations affect the cytoplasmic functions of the co-chaperone DNAJB6 and cause limb-girdle muscular dystrophy. *C. Golzio[1], J. Sarparanta[2], PH. Jonson[2], S. Sandell[3,4], H. Luque[2], M. Screen[2], K. McDonald[5,6], JM. Stajich[5], I. Mahjneh[7], A. Vihola[2], O. Raheem[3], S. Penttilä[3], S. Lehtinen[2], S. Huovinen[3], J. Palmio[3], G. Tasca[8], E. Ricci[8], P. Hackman[2], M. Hauser[5,6], N. Katsanis[1], B. Udd[2,3,9].* 1) Center for Human Disease Modeling, Duke Medical Center, Durham, North Carolina, USA; 2) Folkhälsan Institute of Genetics and Department of Medical Genetics, Haartman Institute, University of Helsinki, Helsinki, Finland; 3) Neuromuscular Research Center, Department of Neurology, University Hospital and University of Tampere, Tampere, Finland; 4) Department of Neurology, Seinäjoki Central Hospital, Seinäjoki, Finland; 5) Department of Medicine, Duke University Medical Center, Durham, North Carolina, USA; 6) Center for Human Genetics, Duke University Medical Center, Durham, North Carolina, USA; 7) Department of Neurology, Pietarsaari Hospital, Pietarsaari, Finland, and Department of Neurology, University of Oulu, Oulu, Finland; 8) Institute of Neurology, Catholic University School of Medicine, Rome, Italy, and Don Carlo Gnocchi Onlus Foundation, Italy; 9) Department of Neurology, Vaasa Central Hospital, Vaasa, Finland.

Limb-girdle muscular dystrophy type 1D (LGMD1D) was linked to chromosome 7q36 over a decade ago, but its genetic cause has remained elusive. Here we report four missense mutations in the ubiquitously expressed co-chaperone *DNAJB6* in nine families from Finland, the U.S., and Italy. All alleles segregated with the phenotype under an autosomal dominant model and included three independent nucleotide changes that encode the same p.Phe93Leu change. Functional testing of mutant protein *in vivo* showed that all patient mutations introduce dominant-negative alleles that have a toxic effect mediated by the cytoplasmic, but not the nuclear, isoform of DNAJB6. Further, we show that DNAJB6 interacts with members of the CASA complex, including the myofibrillar-myopathy-causing protein BAG3, and that the patient mutations reduce the protective anti-aggregation effect of DNAJB6. Our data provide a genetic cause and mechanistic insights into the etiopathology of LGMD1D, indicate a functional link with other forms of muscular dystrophy, and highlight how mutations expressed ubiquitously can exert their effect in a tissue-, cellular compartment-, and isoform-specific manner.

306

Mutations in CLMP cause Congenital Short Bowel Syndrome, pointing to the major role of CLMP in intestinal development. *R.M.W. Hofstra[1], C. van der Werf[1], T.D. Wabbersen[2], N.H. Hsiao[3], J. Paredes[4], H.C. Etchevers[5], P.M. Kroisel[6], D. Tibboel[8], C. Babarit[5], R.A. Schreiber[9], E.J. Hoffenberg[10], M. Vekemans[5], S.J. Zeder[7], I. Ceccherini[11], S. Lyonnet[5], A.S. Ribeiro[4], R. Seruca[4], G.J. te Meerman[1], S.C.D. van Ijzendoorn[3], I.T. Shepherd[2], J.B.G.M. Verheij[1].* 1) Dept Med Gen, UMCG, Univ Groningen, Groningen, Netherlands; 2) Department of Biology, Emory University, Atlanta, USA; 3) The Membrane Cell Biology section, Department of Cell Biology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; 4) The Cancer Genetics Group, the Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal; 5) Département de Génétique, INSERM, U781, Hôpital Necker-Enfants Malades, Université Paris Descartes, Paris, France; 6) Institute of Human Genetics, Medical University of Graz, Graz, Austria; 7) Department of Pediatric Surgery, Medical University of Graz, Graz, Austria; 8) Department of Pediatric Surgery, Erasmus MC-Sophia Children's Hospital, PO Box 2060, 3000 CB, Rotterdam, The Netherlands; 9) The Division of Gastroenterology, BC Children's Hospital, Rm K4-200, 4480 Oak Street, Vancouver, Canada; 10) 10Department of Pediatrics, Section of Pediatric Gastroenterology, Hepatology, and Nutrition, University of Colorado, Denver, CO, USA; 11) Laboratorio di Genetica Molecolare, Istituto Giannina Gaslini - 16148 Genoa, Italy.

Background & Aims: Short Bowel Syndrome usually results from surgical resection of the small intestine for diseases such as Crohn's disease, intestinal atresias, volvulus and necrotizing enterocolitis. Patients with Congenital Short Bowel Syndrome (CSBS) are born with a substantial shortening of the small intestine with a mean length of 50 cm compared to a normal length at birth of 190-280 cm. They are also born with intestinal malrotation. Because of the many consanguineous families reported CSBS is considered as an autosomal recessive disorder. In this study we aimed at identifying and characterizing the gene underlying CSBS. Methods: We applied homozygosity mapping using 610 K SNP arrays on five CSBS patients. After the identification of the underlying gene we determined the expression pattern of the encoded protein in human embryos. Moreover, we overexpressed both wild type and mutant proteins in CHO and T84 cells and we generated a zebrafish model. Results: We identified loss-of-function mutations in Coxsackie- and adenovirus receptor like membrane protein (CLMP) underlying CSBS. CLMP, a tight-junction protein, is expressed in the intestine of human embryos throughout development. Mutation of CLMP abrogated its normal localization at the cell membrane. Knock-down experiments in zebrafish resulted in general developmental defects, including shortening of the intestine and absence of goblet cells, which are characteristic for the mid-intestine in zebrafish, which resembles the small intestine in humans. Conclusions: Loss-of-function of CLMP leads to Congenital Short Bowel Syndrome, likely by interfering with tight-junction formation, with intestinal development and with gut length determination.

307

Massively parallel sequencing identifies a previously unrecognized X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency. *G.J. Lyon[1,2], A. Rope[2], K. Wang[1,6], R. Evjenth[3], J. Xing[2], J.J. Johnston[4], J. Swensen[2], W.E. Johnson[5], B. Moore[2], C.D. Huff[2], L.M. Bird[4], J.C. Carey[2], J.M. Opitz[2], C.A. Stevens[2], C. Schank[5], H. Fain[2], R. Robison[2], B. Dalley[2], S. Chin[2], S. South[2], T.J. Pysher[2], L.B. Jorde[2], H. Hakonarson[1], J.R. Lillehaug[3], L.G. Biesecker[4], M. Yandell[2], T. Arnesen[3].* 1) Pediatrics, Center for Applied Genomics, CHOP, Philadelphia, PA; 2) University of Utah; 3) University of Bergen; 4) Genetic Disease Research Branch, National Human Genome Research Institute; 5) Brigham Young University; 6) University of Southern California.

We have identified two families with a previously undescribed lethal X-linked disorder of infancy comprising a distinct combination of an aged appearance, craniofacial anomalies, hypotonia, global developmental delays, cryptorchidism, cardiac arrhythmia, and cardiomyopathy. We used X-chromosome exon sequencing and VAAST, a recently developed probabilistic disease-gene discovery algorithm, to identify a missense variant in NAA10, a gene encoding the catalytic subunit of the major human N-terminal acetyltransferase (NAT). A parallel effort on a second unrelated family converged on the same variant. The absence of this variant in controls, the amino acid conservation of this region of the protein, the predicted disruptive change, and the co-occurrence in two unrelated families with the same rare disorder suggest that this is the pathogenic mutation. We confirmed this by demonstrating a significantly impaired biochemical activity of the mutant hNaa10p, and from this we conclude that a reduction in acetylation by hNaa10p causes this disease. This is the first evidence of a human genetic disorder resulting from direct impairment of N-terminal acetylation, one of the most common protein modifications in humans. We have also demonstrated that VAAST can readily identify and characterize the genetic basis of a previously unrecognized X-linked syndrome.

308

Whole exome sequencing reveals de novo variants that affect chromatin remodeling processes in Nicolaides-Baraitser syndrome patients.

J.K.J. Van Houdt[1], B. Nowakowska[1], S.B. de Sousa[2,3], B.D.C. van Schaik[6], E. Seuntjes[1], A. Sifrim[1,5], Y. Moreau[5], G. Peeters[1], K. Devriendt[1], R.C.M. Hennekam[4], J.R. Vermeesch[1]. 1) Center for Human Genetics, KULeuven, Leuven, Belgium; 2) Department of Clinical Genetics, Great Ormond Street Hospital for Children, London, UK; 3) Serviço de Genética Médica, Hospital Pediátrico de Coimbra, Coimbra, Portugal; 4) Department of Pediatrics, Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands; 5) Faculty of Engineering - ESAT/SCD, KULeuven, Leuven, Belgium; 6) Clinical Epidemiology, Biostatistics and Bioinformatics (KEBB), Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands.

Nicolaides-Baraitser syndrome (NBS) was first described in 1993, but only recently well delineated in 25 patients. Main characteristics are short stature, sparse hair, typical facial morphology, brachydactyly, interphalangeal joint swellings, and intellectual disability with marked language impairment. The syndrome occurs in persons with different ethnic backgrounds with no significant difference in occurrence in males and females, no familial cases are known and parental consanguinity has not been reported. This suggests that NBS is caused by a dominant de novo mutation in the affected individuals. We performed whole exome sequencing on 8 unrelated NBS patients by targeted exome enrichment and massive parallel sequencing. We obtained between 5.1 to 6.7 Gb of sequence data per individual. On average, 60% of the bases originated from the targeted exome, resulting in a mean exome coverage of 40x, with 80% of the targeted exons covered at least ten times. About 8500 variants per sample were nonsynonymous, frame shift or splice variant changes. Because of the probable dominant nature of this disease, we filtered out SNPs described previously in the control populations, using the NCBI dbSNP build 131 as well as 1000 Genomes Project database. The comparison showed that only 1% of all initially found variants were included for further investigation. We focused our analysis on genes for which at least 3 out of 4 individuals carried a novel, non-synonymous variant at different genomic positions. We identified 5 genes that were affected by distinct missense variants in 3 unrelated cases. Validation by Sanger sequencing confirmed that for one of the genes all detected variants were heterozygous in the affected individuals, and for the three patients for which the parents could be tested it was proven to be de novo.

309

Genetic counsellor education and professional standing in 17 European countries. *C. Cordier[1,2], U. Hosterey-Ugander[3], D. Lambert[4], MA. Voelckel[2,5], H. Skirton[6].* 1) Oncology and haematology, University hospital of Strasbourg, Strasbourg, France; 2) French Association of Genetic Counsellors, AFCG, Marseille, France; 3) Klinisk genetik, Gotebörg, Sweden; 4) Children's university hospital, Dublin, Ireland; 5) Department of genetic, Timone's hospital, Marseille, France; 6) Faculty of health, University of Plymouth, Tauton, United Kingdom.

Quality genetic healthcare services should be available throughout Europe. However, due to enhanced diagnostic and genetic testing options, the pressure on genetic counselling services has increased. Appropriately trained genetic counsellors and genetic nurses can offer clinical care for patients seeking information or testing for a wide range of genetic conditions and the European Society of Human Genetics is making efforts to set up a system of accreditation for genetic counsellors, to ensure safe practice. We undertook a descriptive, cross-sectional survey to obtain baseline data on the role, education and practice of genetic counsellors and nurses in European countries. To collect the data, we approached a number of key informants (leaders in national genetics organisations or experienced practitioners) to complete an online survey, reporting on the situation in their own country. Twenty-eight practitioners responded, providing data from 17 European countries. The findings indicate huge variation in genetic counsellor numbers, roles and education across Europe. In United Kingdom and the Netherlands there are more than 4 counsellors per million population, in five countries there is less than one counsellor per million, while in Turkey, Czech Republic and Germany there are no counsellors. There are specific educational programmes for genetic counsellors in seven countries but only France has a specific governing legal framework for genetic counsellors. We will present these and other data to demonstrate the disparity in approaches to the education and use of genetic counsellors across Europe. This study underpins the need for a coherent European approach to accreditation of counsellors.

310

Differences observed in the communication of breast cancer risk factors among European GP's and breast surgeons. *CM. Julian-Reynier[1], DGR. Evans[2], AD. Bouhnik[1], H. Harris[3], C. Van Asperen[4], A. Tibben[4], J. Schmidtke[5], I. Nippert[6].* 1) Inst Paoli-Calmettes, BP156, INSERM UMR 912, Marseille CDX 9, France; 2) Genetic Medicine MAHSC, Central Manchester University Hospitals, Manchester UK; 3) University of Manchester, Manchester, UK; 4) Center for Human and Clinical Genetics, Leiden University Hopital, Leiden, Netherlands; 5) Institute of Human Genetics, Hannover Medical School, Hannover, Germany; 6) Women' s Health Research, Westfaelische Wilhelms-Universitaet, Münster Medical School, Münster, Germany.

Cancer genetic clinics are focused on the communication of familial/genetic risks but other cancer risk factors intervene in breast cancer occurrence. Our objective was to investigate the options chosen by European Breast Surgeons (BS) and General Practitioners (GP) to present breast cancer risk factors according to the « state of the art » of a better risk communication. A cross-sectional design with postal self-administered questionnaires was used to study the attitudes of primary care providers (N=3999) and those surgeons (N=3293) who most usually treat breast cancer in each of 4 EU countries (UK, France, the Netherlands and Germany). Assessment of family risks and communication about risk factors such as alcohol, obesity, oral contraception, HRT, physical exercise, early pregnancy, breast feeding were first investigated and then the usual formats used/ preferred to present risks (numbered, verbal, framing ...) were surveyed. The overall answer rates were 30% and 37% for GPs (N=1197) and BS (N=1223), respectively. Among the different risk factors, risk family history and hormonal replacement therapy (HRT) were the most frequently explained factors either by GP's or by BS. Countries and providers differed significantly ($p<0.001$) for the kind of factors discussed in particular for those other than family history. Event frequency was the most frequently used presentation (47% GPs; 52% BS), in particular in the UK. Absolute risks were more frequently presented than relative risks. Only 11% of GPs' and 17% of BS would present risk communication including absolute and relative risks with both negative and positive framing and no verbal presentation of risk. Preferences and declared behaviours will be presented according to specialty and country after multivariate adjustment on personal characteristics. In order to optimise cancer risk communication in medicine, initial and vocational training could be reinforced by published guidelines issued from multidisciplinary task forces.

311

Estonian public attitude towards and expectations for personalized medicine. *L. Leitsalu[1], A. Allik[1], A. Metspalu[1,2,3].* 1) Estonian Genome Center of the University of Tartu, Tartu, Tartumaa, Estonia; 2) Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 3) Estonian Biocentre, Tartu, Estonia.

Since the year 2000, the Estonian Genome Center of the University of Tartu (EGCUT), in partnership with primary care providers (PCPs) nationwide, has recruited over 50,000 participants for the Estonian Biobank. The EGCUT is regulated by the Estonian Human Genes Research Act, which states that the participants have the right to receive feedback on their genetic information, and that the feedback should be accompanied by medical counseling, which would most likely occur through PCPs. Besides the necessary developments in genomics research and policy making, it is crucial to educate the PCPs as well as the public in order to facilitate the effective application of genomic information in clinical care and to avoid unsound effects. For that purpose, EGCUT is investigating the knowledge in genetics and the perspectives regarding the use of genomic information, among the PCPs as well as the general public. The first survey, carried out annually over the past 10 years, investigated the general public's opinion and awareness of the EGCUT on a cohort of 1000 individuals between the age of 15 and 74. The 2011 results showed that 67% of the respondents had heard of the EGCUT, and of those, 55% support strongly the idea of EGCUT, 28% have a wait-and-see attitude, 10% say they need more information, 5% cannot comment, and 2% are against the idea of EGCUT. A second survey was conducted among the PCPs collaborating with the EGCUT, investigating their knowledge in genetics and their perspectives regarding the use of genomic information in health care practice. In total, 64 of the 130 PCPs responded. Of the respondents, 96.4% believe that predictive genetic testing will improve health care, 73.3% believe that predictive genetic testing will be used in their practice in the next 5 years, and 75% reported having patients showing interest in their genomic health data. In order to examine the general public's perspectives on the use of genomic information in health care, another survey has been recently conducted on 1000 individuals between the age of 15 and 74. The topics investigated include the knowledge of genetics, the attitudes towards the use of genomic information in health care, and the health behavior. The upcoming results will be used for the development of educational resources in order to improve genetic literacy among the general public, and to facilitate the effective use of genomic information in clinical care.

312

***First Trimester Screening, it's not a routine test* An education module for General Practitioners to help women to make an informed choice.** *K. Dunlop, K. Barlow-Stewart.* Ctr for Genetics Education, Sydney, NSW, Australia.

NSW, Australia, Government policy (2007) states that health care professionals should inform all pregnant women about first trimester screening (FTS) and give appropriate risk information to enable an informed choice. Given the timing of the test, this requirement usually applies to general practitioners (GPs), with increased risk results usually referred to genetic counsellors (GCs) or obstetricians. This presents particular challenges for GPs given patients' different perceptions and interpretations of risk; the pros and cons of screening; the possible implications of an increased risk result; the potential for coercion in guiding decision making; and the inconsistencies in the availability of FTS and genetic counselling services. To support GPs facing these challenges, and at the request of providers of professional development, an education module, First Trimester Screening- it's not a routine test, was developed to be used in group educational settings or stand alone. Content included a slide presentation of information and three video case studies of GP consultations that addressed issues anecdotally reported to commonly occur. Development included expert consultation, case studies prepared in consultation with GPs and GCs, actors in the patient roles with a practicing GP and having a creative director finesse the dialogue and direct filming to ensure an engaging product. The module was piloted in GP education sessions (1 hr) in two different divisions of General Practice with GPs (52) and GCs (5). GPs were also asked to choose from a list of ways to observe the videos such as focusing on the step by step approach suggested and report on this process. The module was rated highly relevant and useful (96%) and reflected questions that arise in GP practice. 7/52 GPs (13%) requested that they would like to see further video cases included, in particular discussing increased risk results. Comments on how it would impact their future practice included having more thorough discussion of FTS, discuss over more than one session, adopt ideas from the way the GP phrased her comments and be more confident. Bias “for screening” was noted by 10/52 GPs (19%) and 3/5 GCs (60%), and requires review. With the increased community access to information about genetic risk, this educational approach may have a significant role to play as more than ever GPs are required to assist patients in understanding their risk and facilitating decision making around FTS.

313

Increasing the ability to predict FMR1 repeat instability: regression analysis of CGG/AGG structure in 546 transmissions from parents with 45-69 CGG repeats. *E.G. Allen[1], S.L. Sherman[1], S.L. Nolin[2], A. Glicksman[2], E. Berry-Kravis[3], F. Tassone[4], C. Yrigollen[4], A. Cronister[5], M. Jodah[5], N. Ersalesi[2], W.T. Brown[2], R. Shroff[6], S. Sah[6], G.J. Latham[6], A.G. Hadd[6].* 1) Department of Human Genetics, Emory University, 615 Michael St. , Atlanta, GA. 30322; 2) New York State Institute for Basic Research in Developmental Disabilities, 1050 Forest Hill Road, Staten Island, NY 10314; 3) Rush University Medical Center, 1725 West Harrison Street, Chicago, IL 60612; 4) Dept. of Biochemistry and Molecular Medicine, 4303 Tupper Hall, One Shields Avenue, Davis, CA 95616; 5) Genzyme Genetics, Cambridge, MA; 6) Asuragen, Inc., Austin, TX.

Screening pregnant women to determine their fragile X mutation status has become increasingly common in recent years. This has resulted in the identification of intermediate (45-54 repeats) and small premutation (55-69 repeats) alleles whose stability is poorly understood. Until now, however, the presence of the second X chromosome in females has prevented large scale analysis of sequence data. An elegant new PCR assay has allowed us to examine the effect the FMR1 CGG repeat structure has on repeat instability in females. In the current study, we tested whether AGG interruptions and the length of pure 3' CGG repeats, two traits that have been indicated as important in previous studies, are better predictors for instability than the currently used total repeat length. We focused our study on parental alleles of 45-69 CGG repeats. First, there was a significantly higher proportion of unstable transmissions among paternal (62.8%, n=77) compared with maternal (50.5%, n=469) transmissions; although, the range of repeat size change from parent to child (delta) was smaller for paternal (56) compared with maternal transmissions (161). Thus, we analyzed maternal and paternal transmissions separately, using regression models with delta as the outcome variable and total repeat length, number of AGG interspersions, and length of the 3' pure CGG repeat as predictor variables. For females, the pure CGGs at the 3' end of the repeat region was the best predictor of delta ($p<.0001, R^2=0.21$) and explained twice the variance compared with total repeat length alone. For paternal transmissions, either 3' pure repeats or number of AGGs alone were significant predictors of delta ($p<.001$, $R^2=0.13$ and 0.14, respectively). Having both predictors in the model did not improve the fit, nor did a quadratic term for 3' pure repeats. Compared with total repeat length, either 3' pure repeats or number of AGGs explained three times the variance in delta compared to total repeat length alone. The difference between maternal and paternal transmissions is further indicated by the magnitude of the beta coefficients of the predictors in the models. In summary, either the number of AGGs or the 3' pure CGG repeat length along with parental origin more accurately estimates the risk for instability compared to the current standard of repeat size alone. Thus AGG genotyping can immediately improve counseling regarding repeat instability prediction.

314

Twenty-six years of prenatal testing for Duchenne and Becker Muscular Dystrophy in the Netherlands: Has it made an impact? *A.T.J.M. Helderman-van den Enden[1,2], K.M. Madan[1], M.H. Breuning[1], A.H. van der Hout[3], E. Bakker[1], C.E.M. de Die-Smulders[2], H.B. Ginjaar[1].* 1) Human and Clinical Genetics , Leiden University Medical Center, Leiden, Netherlands; 2) Maastricht University Medical Center PO Box 5800, 6202 AZ Maastricht, the Netherlands; 3) Department of Genetics, University Medical Center Groningen, University of Groningen, The Netherlands.

Background: Carrier testing and prenatal diagnosis for Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) were introduced in 1984. We investigated the effect of 26 years (1984-2009) of genetic counselling on the incidence of these diseases in the Netherlands. Methods: The incidence of DMD and BMD from 1993 to 2002 was compared to that in the period 1961-1974. Information on all prenatal diagnoses performed from 1984 to 2009 was analysed. Findings: There was no significant difference in incidence between the two periods. The percentage of second affected boys in the DMD families decreased from 38% in the period 1961-1974 to 12% in 1993-2002. Of the 635 prenatal diagnoses performed in the 26 year period 51% were males. Of these, nearly half (46%) were affected or had an increased risk of having DMD/BMD. As a result 145 male foetuses were aborted and 174 unaffected boys were born. Nearly 80% of females, now 16 years or older, who were prenatally diagnosed but not tested for carrier status have still not been tested. Interpretation: Although there was no apparent decrease in the incidence of the disease, prenatal diagnosis was used by DMD families for preventing the birth of a second affected boy. It was possible for DMD families to have unaffected boys. Our study has revealed: 1) a high proportion of families with de novo mutations in the DMD gene cannot make use of prenatal diagnosis, partly because the older affected boys are diagnosed at around 4 years. 2) Current policy dictates that female foetuses should not be tested. As adults these untested females risk having a son with DMD and cardiomyopathy. The consensus within the genetic community on not testing girls should be reconsidered. Prenatal and neonatal screening of males and pre-conceptional and prenatal testing for carrier status in females could improve the chances of prevention of DMD in the future.

315

Tryptic Peptide Analysis of WBC to Diagnose Genetic disorders: Application to Primary Immunodeficiency Disorders and Nephropathic Cystinosis. *S. Hahn[1,2], S. Kerfoot[2], S. Jung[2], V. Vasta[2], K. Golob[2], T. Torgerson[1,2], T. Vilboux[3], W. Gahl[3].* 1) Dept Pediatrics, Univ Washington/Seattle Children's Hospital, Seattle, WA; 2) Seattle Children's Hospital Research Institute, Seattle, WA; 3) Section on Human Biochemical Genetics, Medical Genetics Branch, NHGRI, NIH, Bethesda, MD.

PURPOSE: There are many genetic diseases that would benefit from early diagnosis but that have no population screening platform. Examples include primary immunodeficiency disorders such as X-linked agammaglobulinemia (XLA), Wiskott Aldrich Syndrome (WAS), and Severe Combined Immune Deficiency (SCID), as well as nephropathic cystinosis. BTK, WASP, CD3, and cystinosin are low abundance proteins deficient in these conditions. They are localized in cell membranes, cytoplasm, or on the cell surface and are not detectable in plasma. We propose that these proteins can be observed in proteolytically digested extracts of WBCs and will be absent or significantly reduced in affected cells, making diagnosis possible. **METHODS:** Candidate peptides were screened by in silico trypsin digestion modeling followed by a BLAST search to insure that the sequences are unique within the human genome. The final “signature” peptides were selected by evaluating the MRM chromatogram for the isotopically labeled peptide and the WBC digest peptide in control (n=20) and patient samples. Five lymphocyte cell lines were used to establish the absence of signature peptides for primary immunodeficiencies. Three PBMC samples from XLA were blindly analyzed. WBCs from 10 cystinosis patients with known genotypes were also tested. The amount of each peptide in the WBC was determined by taking the ratio of the peak area for the signature peptide to that of the labeled peptide and reported as normalized to actin. **RESULTS:** Three blinded samples lacked only BTK peptides, confirming that these samples were from XLA patients. All five cell lines clearly showed the biochemical phenotype of each cell line. Cystinosin was non-detectable in 9 cystinosis patients. One patient showed a cystinosin peak but at a very low concentration compared to control. **CONCLUSION:** Our method quantified the proteolytic peptides for the target proteins, BTK, WASP, CD3, and cystinosin in various cell lines and patient samples. Targeted proteins from these conditions were either absent or significantly diminished. This approach can be potentially utilized as part of a multiplex analysis for many genetic conditions in which the protein of interest is significantly reduced. While the transition from white blood cells to dried blood spots will be challenging, we believe that with further enrichment and optimization, this method can be potentially applied to newborn screening.

316

Genetic test evaluation models: how well do they fit the translational process? *F.E. Caballero[1,2], R.N. Battista[1,2,4], A.M. Laberge[2,3,4].* 1) Department of Health Administration (DASUM), Université de Montréal, Montréal, Québec, Canada; 2) CHU Sainte-Justine Research Center, Montréal, Québec, Canada; 3) Department of Pediatrics, Université de Montréal, Montréal, Québec, Canada; 4) Apogée-Net/CanGèneTest Network.

Purpose: Compare genetic test evaluation models in terms of their place in the translational pathway. Context: Over the past two decades, the exponential growth of new discoveries and advances in the realm of genetics has led to a significant increase in the number of genetic test available. Concerns have been expressed regarding the appropriate assessment of these technologies prior to their widespread use. Several frameworks have been developed to provide a systematic, evidence-based approach for the evaluation of new genomic applications. We propose an overview and comparison of different models presented in the literature for the evaluation of genetic tests. Methods: Using PubMed and other publicly available databases and using appropriate keywords (e.g. “genetic test” AND “evaluation”), we identified genetic test evaluation models. The information used to evaluate genetic tests in each model was situated along the translational pathway developed by Khoury et al. A general test model was used as a comparison (USPSTF). Results: Genetic test evaluation models come from different sources (government agencies, researchers) and countries (North America and Europe). We identified seven evaluation models specific to genetic tests. While most suggested a formal streamlined evidence collection procedure, review and synthesis of the available data, few focused on issues beyond T3 in their approach. Only two addressed specifically the decision-making process with a way to classify or compare tests. Conclusion: The evaluation of genetic tests often stops at the evidence review, synthesis and recommendation for a specific test, but decision makers need a way to apply the results of these evaluations in the context of limited resources and competing interests. Models that provide a way to compare or classify tests are more likely to be used by decision-makers. We suggest the evaluation process should be viewed as a comprehensive process ranging from early transition from the research setting to clinical practice all the way to post implantation stages where their impact on health outcomes are observed. By shedding light on the strengths and limitations of these various evaluative methods, we hope to allow decision makers at the policy, management and clinical level to take informed resource allocation decisions in respect to the introduction and use of genetic tests.

317

A replication study examining novel common single nucleotide polymorphisms identified through a prostate cancer genome wide association study (GWAS) in a Japanese population. *J. Batra[1], F. Lose[2], S. Chambers[3, 4], R. Gardiner[5], J. Aitkin[4], J. Yaxley[6], J. Clements[1], A. Spurdle[2], Australian Prostate Cancer Bioresource.* 1) Australian Prostate Cancer Research Centre-Queensland and Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia; 2) Molecular Cancer Epidemiology, Queensland Institute of Medical Research, Brisbane, Queensland, Australia; 3) Griffith Health Institute, Griffith University, Brisbane, Queensland, Australia; 4) 4Viertel Centre for Research in Cancer Control, Cancer Council Queensland, Brisbane, Queensland, Australia; 5) University of Queensland, Centre for Clinical Research, Royal Brisbane Hospital, Brisbane, Queensland, Australia; 6) Brisbane Private Hospital, Brisbane, Queensland, Australia.

Five novel prostate cancer risk loci have been identified in a recent genome wide association study (GWAS) of Japanese individuals. We proposed that apart from population specific linkage disequilibrium (LD) patterns, limitations of Stage1 GWAS power to prioritize SNPs for publication report and/ or study design could explain the missing report on these loci in previously conducted GWAS on Caucasians. Thus, we undertook a replication study in 1,357 prostate cancer patients and 1,403 normal Australian males of European decent. The minor allele frequency (MAF) for the analysed SNPs was not markedly different between European and Japanese samples except for the rs13385191 as indicated by F_{ST} scores from the HapMap data. In the absence of the knowledge of where a causative SNP may be located, we investigated the possible differences in LD patterns between Japanese and European populations for a ±10kb region around the 5 SNPs identified by the Japanese GWAS. This extended analysis of Fst values suggest, but do not prove, that heterogeneity may drive differences in LD patterns between Japanese and European populations for rs13385191 and rs9600079, but is unlikely to do so for rs12653946, rs1983891 and rs339331.

We found rs12653946 at 5p15 to be significantly associated with risk of prostate cancer in Caucasian males with an odds ratio of 1.20; (95% Confidence Interval = 1.07, 1.34), P = 0.002, with per-allele effect size similar to that reported in Japanese men. The risk estimate did not differ markedly for cases stratified by Gleason score <7 vs.⁄ 7, or by report of family history, as reported by the Japanese study. On the basis of LD calculations, the rs12653946 SNP represents an independent locus, distinct from the previously identified *TERT-CLPTM1L* cancer nexus region. Further, our bioinformatic analysis indicates that rs12653946 falls in the intron of a testis-expressed gene strongly predicted to translate a conceptual 8.1kDa protein named *tojy*.

Our findings indicate that follow-up of apparently ethnic-specific risk associations are warranted to highlight risk-associated loci for experimental studies, and for incorporation in future risk prediction models for prostate cancer.

318

Polymorphisms in MMPs, TIMPs, and RECK interact to influence susceptibility to prostate cancer and aggressive disease among two groups of African American men. *W. Hernandez[1], S. Hooker[2], R. Kittles[3].* 1) Dept Med, Univ Chicago, Chicago, IL; 2) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Section of Hematology/ Oncology, Department of Medicine, University of Illinois at Chicago.

In the USA, African American men exhibit the highest rate of prostate cancer (PCa) and are more likely to die from the disease. The metalloproteinases (MMPs) are capable of degrading all components of the extracellular matrix and excess MMP activity is considered to be a key event in cancer invasion and metastasis. Tissue inhibitors of metalloproteinases (TIMPs) and the reversion-inducing cysteine-rich protein with Kazal motifs (RECK) regulate MMP activity and have been shown to play important roles in several types of cancer initiation and aggressiveness. Here we examined polymorphisms located in exons, 3'UTR, 5'UTR, and promoter regions of MMPs, TIMPs, and RECK and prostate cancer susceptibility and aggressive disease. A total of 88 SNPs were genotyped in 575 men self-described as African American (cases=257, controls=318) from Washington, DC, using MALDI-TOF Mass spectrometry analysis (Sequenom Inc., San Diego, CA). Statistical association analyses were conducted adjusting for age as well as West African ancestry. Based on our findings from this discovery set of samples we selected 30 of the 88 SNPs to genotype in a second group of African American men from Chicago, IL (cases=332, controls=157). Single marker analyses revealed that the presence of the minor allele (MA) for rs514921 (MMP-1) and rs1042703 (MMP-14) reduced the risk of PCa (rs514921 OR=0.63, P=0.007 and rs1042703 OR=0.66, P=0.03) in the DC group. In addition, rs602128 (MMP-3) significantly increased PCa risk by 1.5 fold (P=0.009). In the Chicago group, rs20544 (MMP-9) increased the risk of aggressive prostate cancer (OR=2.0, Pemp=0.001). We also found rs4643046 (MMP-8) to significantly decrease risk of aggressive disease (OR=0.45, Pemp=0.0003) as well as the risk of PCa. We observed the same for a haplotype of two SNPs located in exon 5 of TIMP-4 (OR=0.54, P= 0.02; OR=0.41, P=0.04). We further explored SNPxSNP associations and found several significant associations, but most importantly, a few replicated in both populations. In the two populations, SNPs rs3918241 and rs2274756 increased the risk of prostate cancer (OR=2.4, P=0.006; OR-2.64, P=0.03 respectively) while SNPs rs17035945 and rs743257 increased the risk of aggressive disease (OR=3.1, P=0.005 and OR=1.9, P=0.04 respectively). Our data suggest that SNPs in MMPs, TIMPs, and RECK interact to influence prostate cancer risk and aggressive disease in African Americans.

319

Exome sequencing reveals sequential alterations of cancer driver genes during progression of metastatic prostate cancer. *M.L. Nickerson[1], K.M. Im[1], K.J. Misner[1], A.L. Yates[1], D.W. Wells[1], H.C. Bravo[2], K. Fredrikson[3], W. Tan[1], M. Yeager[4], P. Milos[5], B. Zbar[1], G.S. Bova[6], M. Dean[1].* 1) Cancer & Inflammation Program, National Cancer Institute, National Institutes of Health, Frederick, MD; 2) Center for Bioinformatics and Computational Biology, Department of Computer Science, University of Maryland, College Park, MD; 3) Roche Diagnostics Corporation, Indianapolis, IN; 4) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD; 5) Helicos Biosciences, Cambridge, MA; 6) Department of Pathology, Johns Hopkins Hospital, Johns Hopkins University, Baltimore, MD.

Metastatic prostate cancer is a highly morbid condition, poorly controlled using existing therapies, and the molecular ontology of cancer progression to metastatic disease is incompletely defined. The high frequency of prostate cancer in older men and the poor outcome of patients with metastatic disease make it critical to identify molecular markers to distinguish indolent from aggressive disease. Exome sequencing of five metastatic tumors and non-cancerous tissue from a clinically aggressive case of prostate cancer revealed a germline, 185delAG mutation in *BRCA1* and 62 somatic, nonsynonymous variants in 62 genes. Six distinct alterations of 4 known drivers of cancer were detected, including 2-hits in *BRCA1* and *TMPRSS2*, and novel missense alterations of *PBRM1* and *TET2*. Twenty-two of the 62 somatic variants were detected in matched primary prostate adenocarcinoma, including the *PBRM1* alteration, and were considered candidate cancer initiating genes. Thirty-one somatic variants were observed in all five metastatic deposits but not the primary tumor and allowed the somatic genotype of a metastatic progenitor cell to be inferred. These variants defined a set of candidate metastasis genes, including *TMPRSS2* and *TET2*. Nine variants were present in a subset of deposits and allowed the sequential spread of cancer to be partially mapped. Nineteen genes with somatic alterations in the initial patient were examined in prostate tumors from additional patients and revealed somatic alterations in *TET2* in 10% of samples. We show altered expression of *TET2* in a prostate cancer cell line altered hydroxymethylcytosine levels, indicating CpG demethylation may play a role in prostate cancer. This study demonstrated the utility of exome sequencing for the genetic dissection of cancer. Somatic alterations of four cancer driver genes occurred at different times during progression in addition to a germline *BRCA1* deleterious allele. Two cancer driver genes produce proteins with roles in chromatin (*PBRM1*) and DNA (*TET2*) epigenetic modification and indicated involvement of the SWI/SNF complex in prostate cancer.

320

Genomic Signatures of Metastasis in Prostate Cancer. *A. Pearlman[1], C. Campbell[1], E. Brooks[1], A. Genshaft[1], S. Shajahan[1], M. Ittmann[2], G. Bova[3], J. Melamed[4], I. Holcomb[5], R. Schneider[6], Y. Shao[7], H. Ostrer[1].* 1) Human Genetics Program, NYU Langone Medical Center, New York, NY; 2) Department of Pathology, Baylor College of Medicine, Houston, TX; 3) Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD; 4) Department of Pathology, NYU School of Medicine, New York, NY; 5) Department of Pathology, Stanford University School of Medicine, Stanford, CA; 6) Translational Research, NYU Cancer Institute, and Department of Microbiology, NYU School of Medicine, New York; 7) Division of Biostatistics, NYU School of Medicine, New York, NY.

Identifying the subset of mutations that cause a primary tumor to metastasize could provide valuable prognostic and therapeutic clinical utility. To discern the mutations that cause metastases in prostate cancer, genomic DNA copy number alterations (CNAs) were studied in 294 prostate cancer primary tumors and 49 metastases from 5 independent cohorts. We developed novel methods to identify CNAs concordant in metastases and primary tumors from men treated with radical prostatectomy that later progressed to form distant metastases, while discordant with local primary tumors that did not metastasize. This signature included putative metastasis genes as well as genes whose role in metastasis was not previously identified. Eight solute carrier superfamily members were identified including a deleted SLC7A5 gene. The SLC7A5 gene encodes a regulator of cellular L-glutamine concentration which at high concentrations, constitutively fuels mTOR activity. Six Cadherin family members included CDH17 as an amplified gene. RNAi knockdown of CDH17 was previously shown to inhibit hepatocellular carcinoma metastasis. Five potassium channels were predictive of metastasis. Altered cytoplasmic potassium ion concentrations promote escape from anoikis, the process of apoptosis that occurs when anchorage-dependent cells detach from the surrounding matrix. 193 of these genes were shown to undergo somatic mutation in prostate cancer, 5 resulted in potentially deleterious protein altering mutations and 12 were identified in promoter regions. The metastatic signature was observed in cell lines from lung, breast, colon, and melanoma, suggesting that other tumors may share pathways of metastasis. In addition, we developed a prognostic model based on the metastasis signature which we assessed with other early pre-treatment clinical predictors using multivariate logistic regression and Cox proportional hazards models. Our proposed metastatic potential score was the only significant predictor of metastasis in a multivariate logistic regression model and resulted in a predictive accuracy of 80%, p = 0.01. This score was also the only significant parameter in a multivariate Cox proportional hazards model (HR = 2.88; HR95%CI = 1.15-7.2; p=0.02) to predict metastasis-free survival. The data indicate a prognostic clinical utility and the functional attributes of the predictive genes provide targets for the development of therapeutics for prostate and likely other cancers.

321

Most, but not all, common risk alleles for colorectal cancer act through predisposition to adenomas. *L.G. Carvajal-Carmona[1], A.G. Zauber[2], A.M. Jones[1], J. Wang[3], D. Morton[4], M.M. Bertagnolli[3], I. Tomlinson[1,5].* 1) Wellcome Trust Ctr Human Gen, Univ Oxford, Oxford, United Kingdom; 2) Department of Epidemiology and Statistics, Memorial Sloan-Kettering Cancer Center, New York, USA; 3) Department of Surgery, Brigham and Women's Hospital, Boston, USA; 4) Academic Department of Surgery, University of Birmingham, Edgbaston, United Kingdom; 5) Oxford Comprehensive Biomedical Research Centre.

Background & aims. Genome-wide association studies have identified 18 tagSNPs associated with the risk of developing colorectal cancer (CRC), but the underlying functional mechanisms are largely unknown. The great majority of CRCs probably arise from adenomas, and the adenoma is therefore an intermediate phenotype for CRC. Methods. We determined whether each CRC predisposition SNP influenced susceptibility to adenomas or only had effects on the growth of carcinomas. We tested all 18 SNPs in three series of adenoma cases without CRC (APC, APPROVe and CORGI studies, N=2,390) and 3 sets of controls (CORGI, National Blood Service and 1958 Birth Cohort studies, N=7,272). We compared results in the adenoma cohorts with those from CRC cohorts. Results. Twelve SNPs (rs10936599, rs16892766, rs6983267, rs10795668, rs3802842, rs4444235, rs1957636, rs4779584, rs4939827, rs9929218, rs961253, rs4925386) were associated with adenoma susceptibility, independent of a patient's family history of CRC. The remaining 6 SNPs (rs6691170, rs6687758, rs7136702, rs11169552, rs10411210, rs4813802) were not associated with adenoma risk, and were therefore likely to act solely on adenoma-carcinoma transition and/or carcinoma progression, or act in an adenoma-independent pathway. rs4444235, close to BMP4, had a significantly stronger association with adenoma than carcinoma. rs6983267, proximal to MYC, was associated with multiple adenomas, but no SNP was associated with advanced adenomas. Conclusions. Common polymorphisms influence CRC risk at different stages of colorectal carcinogenesis. If genetic risk stratification were to be used to influence population screening or chemoprevention, sub-groups of affected or unaffected individuals could be defined with different risks of developing adenoma, carcinoma or both.

322

Variation in epigenetic regulatory pathways and breast cancer susceptibility. *T. Kirchhoff[1], J. Rendleman[1], Y. Antipin[2], B. Reva[2], C. Sander[2], M.M. Gaudet[3], R.J. Klein[2], K. Offit[2].* 1) New York University School of Medicine, NYU, New York, NY, USA; 2) Memorial Sloan-Kettering Cancer Center, New York, NY, USA; 3) American Cancer Society, Atlanta, GA, USA.

Epigenetic aberrations are the hallmarks of many human cancers. Although the epigenome is being extensively studied on molecular level, little is known about the germline alterations in these pathways contributing to human malignancies. In our study we hypothesize that the common genetic variation of the genes involved in the epigenetic regulation may alter these processes and thus confer an effect on risk associated with breast cancer. To support this hypothesis, we have demonstrated that the epigenetic regulatory genes (ERGs) are significantly enriched in prior breast cancer GWAS studies of CGEMs and AJ breast cancer GWAS. In order to provide a systematic evaluation of this hypothesis, we have designed a case/control association analysis of 180 ERGs genotyped on 1100 common (maf>0.05)tagging SNPs(tSNPs) in 2500 breast cancer cases and 2500 controls, both of European ancestry. The ERGs were selected using bioinformatics strategies as belonging to several classes of proteins involved in epigenetic regulation. The genotyping was performed using Sequenom iPLEX technology and while the full scan is in progress, here we report the data of the first phase on 131 genes and 711 tSNPs in 2500 breast cancer cases and 2500 controls. Using logistic regression, we have identified a variant on chromosome 6 showing a significant association with breast cancer risk (OR = .7754, 95% CI =.68-.90, p = .00044). The associated locus on chromosome 6 is centered around the region of a methylatransferase involved in embryonic development, recently shown to be aberrantly expressed in breast cancers. Additional associations include loci on chromosome 12 (OR = 1.197, p = .019), chromosome 6 (OR = 1.178, p = .020), chromosome 21 (OR = .8578, .7745, 1.177; p = .022, .023, .047), chromosome 1 (OR = .8584, p = .032), and on chromosome 2 (OR = 1.182, p = .043). The haplotype and epistatic analyses are currently in progress using the imputed data from HapMap phase 2 and 1000 genomes to access the presence of potentially functional variants and the gene-gene interactions between the different components of epigenetic regulatory pathways. Besides the validation of identified variants in International consortia, and testing the associations for other clinico-pathological variables, the functional investigation will continue by correlation of risk alleles with expression and methylation profiles in the large collection of breast tumors.

323

Alteration of *PBRM1* and *VHL* in clear cell renal cancer. *M. Dean[1], K.M. Im[1], K. Misner[1], D. Wells[1], B. Gold[1], L.S. Schmidt[2], B. Zbar[1], N. Rothman[3], W.-H. Chow[3], F. Waldman[4], W.M. Linehan[2], L.E. Moore[3], M.L. Nickerson[1].* 1) Cancer and Inflammation Program, National Cancer Institute, National Institutes of Health, Frederick, MD; 2) Urologic Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD; 3) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD; 4) School of Medicine, University of California San Francisco, San Francisco, CA.

Sporadic clear cell renal cell carcinoma (ccRCC) is the most common form of kidney cancer and is characterized by a high frequency of somatic alteration of the *von Hippel-Lindau* (*VHL*) gene, including sequence variants that alter the protein, hypermethylation of promoter CpGs that silence expression, and loss of the *VHL* locus on the distal region of chromosome arm 3p. The *polybromo 1* (*PBRM1*) gene, also located on chromosome 3p, was recently shown to be a second frequently mutated gene in ccRCC. The complete spectrum of *PBRM1* mutations as well as association with tumor grade, stage, and other clinicopathologic parameters such as *VHL* inactivation and gene expression have yet to be clarified. We sequenced PBRM1 in 198 sporadic ccRCC tumors for which matching normal DNA (blood or non-cancerous tissue) was available. This data was analyzed relative to *VHL* inactivation and chromosome 3p LOH, patient ethnicity, exposure and clinical history, and tumor characteristics. We observed alteration of *PBRM1* in 37% of tumors and most alterations (81%) were insertions, deletions, or nonsense changes indicative of a tumor suppressor. There were no significant differences in clinical parameters among cases with and without a *PBRM1* mutation. Alteration of both *PBRM1* and *VHL* occurred in 34% of tumors and 3% were *PBRM1*mut/*VHL*wt. *VHL* alone was altered in 51% of tumors and neither gene was altered in ~12% of cases. Alterations in these two tumor suppressor genes were observed in clear cell carcinoma of the kidney in approximately 88% of cases examined. Upregulation of HIF transcription factors by *VHL* inactivation is associated with an induction of angiogenesis and a shift in cell metabolism that may be complemented by loss of function of PBRM1 protein. The presence of PBRM1 protein is known to define a subtype of SWI/SNF complex (PBAF) structurally related to yeast Rsc whose function in chromatin remodeling is not yet known.

324

A new disease: retinoblastoma driven by *MYCN* amplification with normal *RB1* tumor suppressor gene alleles. *B.L. Gallie[1], D.E. Rushlow[2], S. Yee[3], J.Y. Kennett[4], P. Boutros[5], N.L. Prigoda-Lee[2], W. Halliday[6], S. Pajovic[1], C. Spencer[1], B.L. Thériault[1], H. Dimaras[7], A. Raizis[8], C. Houdayer[9], D. Lohmann[10].* 1) Campbell Family Cancer Research Institute and Ontario Cancer Institute, Princess Margaret Hosp, Toronto, Canada; 2) Retinoblastoma Solutions, Toronto Western Hospital Research Institute, University Health Network, Toronto, ON, Canada; 3) Department of Molecular Genetics, University of Toronto, Toronto, Canada; 4) British Columbia Cancer Research Centre, Vancouver, Canada; 5) The Ontario Institute for Cancer Research, Toronto, Canada; 6) Department of Pathology, Hospital for Sick Children, Toronto, Canada; 7) Departments of Ophthalmology & Visual Science and Hematology/Oncology, Hospital for Sick Children, Toronto, Canada; 8) Department of Molecular Pathology, Canterbury Health Laboratories, Christchurch, New Zealand; 9) Service de Génétique Oncologique, Institut Curie and Université Paris Descartes, Paris, France; 10) Institut für Humangenetik, Universitätsklinikum Essen, Germany.

Background: Dogma states that all retinoblastoma tumors have lost both alleles of the *RB1* tumor suppressor gene. We can efficiently identify 95% of *RB1* mutant alleles. In 2% of tumors we failed to find any mutant *RB1* allele ($RB1^{+/+}$). **Methods:** In the $RB1^{+/+}$ tumors, we characterized expression of pRB and other proteins, specific genomic changes of $RB1^{-/-}$ retinoblastomas, genome copies and clinical features. One $RB1^{+/+}MYCN^{A}$ cell line was studied for growth rate in comparison to $RB1^{-/-}$ cell lines. **Results:** In 1% of retinoblastomas we discovered high level amplification of the *MYCN* oncogene and normal *RB1* ($RB1^{+/+}MYCN^{A}$ tumors). The genomic instability characteristic of $RB1^{-/-}$ retinoblastomas was not present. The children with the $RB1^{+/+}MYCN^{A}$ tumors presented at very young ages (6 months vs 24 months for non-hereditary retinoblastoma). Distinctive histopathological features included prominent nucleoli. *In vitro* the $RB1^{+/+}MYCN^{A}$ cell line rapidly died when MYCN levels are reduced by shRNA, in comparison to $RB1^{-/-}$ cell lines with late acquired *MYCN* amplification that showed only slowed growth. Subsequent to this dataset, we predicted that a child with unilateral retinoblastoma at age 4 months would have a 15% chance to have a $RB1^{+/+}MYCN^{A}$ tumor; the enucleated tumor indeed had 40 copies of *MYCN* and all the described features. **Conclusions:** Through our clinical analysis of *RB1* alleles in more than 1000 retinoblastoma probands, we have discovered a previously unrecognized disease: retinoblastoma that is not caused by *RB1* gene mutations. We hypothesize that children presenting with unilateral retinoblastoma filling the eye before 6 months of age, and those with extraocular/metastatic retinoblastoma before 1 year of age, are likely to have $RB1^{+/+}MYCN^{A}$ tumors. Despite their rapid growth in very young children, $RB1^{+/+}MYCN^{A}$ tumors may be more responsive to therapy than $RB1^{-/-}$ tumors, since they have less genomic instability and therefore less capacity to achieve drug resistance. Although $RB1^{+/+}MYCN^{A}$ tumors are aggressive and rapidly growing, therapeutic targeting of MYCN may cure. Children with $RB1^{+/+}MYCN^{A}$ tumors have normal population risk of heritable disease, despite their young age.