## †1

**Dissection of a Modifier Network in Marfan Syndrome Reveals New Therapeutic Targets.** *J. Doyle[1], J. Pardo Habashi[1], T. Holm[1], D. Bedja[1], H. Dietz[1,2].* 1) Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 2) Howard Hughes Medical Institute.

Excessive TGFβ activity causes many of the manifestations of Marfan syndrome (MFS), including aortic aneurysm, that can be attenuated by TGFβ antagonism with neutralizing antibody or the angiotensin II (Ang II) type 1 receptor (AT1) blocker (ARB) losartan in MFS mice (Fbn1 genotype C1039G/+). TGFβ signals through both canonical (pSmad2/3) and noncanonical (MAPK: Jnk, Erk, p38) cascades. To determine their relative contribution, we introduced haploinsufficiency for Smad4 (a mediator of Smad-dependent signaling) on the C1039G/+ background (S4+/-:C1039G/+). Unexpectedly, S4+/-:C1039G/+ mice showed synthetic lethality due to aortic rupture compared to S4+/+:C1039G/+ littermates. Both S4+/+:C1039G/+ and S4+/-:C1039G/+ mice show marked and equivalent activation of the Smad2/3 and Erk1/2 cascades, which could not account for synthetic lethality. S4+/-:C1039G/+ mice uniquely show activation of Jnk, and a Jnk antagonist fully rescues synthetic lethality. Both an Erk1/2 and Jnk1 antagonist protected against aneurysm progression in S4+/+:C1039G/+ mice despite the fact that only Erk shows activation in this strain. Activation of MAPKs is dependent on both AT1 and TGFβ. These data show that Erk and Jnk signaling is additive and interchangeable in the pathogenesis of ascending aortic aneurysm. AngII signals through both the AT1 and AT2 receptors. Use of ACE inhibitors (ACEi, e.g. enalapril) inhibits both receptors, whereas selective AT1 blockade is achieved with ARBs. AT2 signaling can induce apoptosis, prompting the proposal that ACEi are preferable. To address this controversy we knocked out the AT2 gene in C1039G/+ mice (AT2KO:C1039G/+) and showed larger aortas and worsened vessel architecture compared to AT2+:C1039G/+ animals. A trial of enalapril in the MFS mice did not attenuate aortic growth, while losartan led to regression of aortic size. Much of the protection by losartan was lost in AT2KO:C1039G/+ mice, suggesting that shunting of signaling through AT2 upon AT1 blockade contributes to aortic rescue. Finally, we show that AT2 signaling mediates protection through inhibition of activation of Erk, which is increased in AT2KO:C1039G/+ mice. In contrast to losartan, enalapril fails to reduce Erk1/2 activation in C1039G/+ mice. Taken together, these data define MAPK and AT2 signaling as prognostic and therapeutic modifiers of vascular disease in MFS and provide rationale and incentive for a clinical trial of Erk and Jnk antagonists and AT2 agonists in MFS.

## 2

**Next-Generation Mendelian Genetics by Exome Sequencing.** *J. Shendure[1], S.B. Ng[1], E.H. Turner[1], P.D. Robertson[1], A.W. Bigham[2], C. Lee[1], E.E. Eichler[1], M. Bamshad[2], D.A. Nickerson[1].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Department of Pediatrics, University of Washington, Seattle, WA.

Rare monogenic diseases have been of incredible value to biomedical research, as the identification of the genes underlying phenotypes of interest has yielded fundamental, medically relevant insights into human biology. However, many rare disorders have yet to be solved. In many cases, this is because only a small number of cases/families are available, limiting the power of traditional gene mapping strategies. As most disease-causing variants affect coding sequences, comprehensive resequencing of all human genes has the potential to serve as a genome-wide scan for the underlying cause of a rare monogenic disease. While the routine sequencing of full human genomes continues to be cost prohibitive, the cost of sequencing all protein-coding regions (i.e., the "exome", ~1% of a human genome) may soon be on par with that of dense genotyping arrays. We evaluated whether second-generation methods for targeted sequencing of the human exome could serve as a genome-wide scan for monogenic diseases that is independent of linkage data or other a priori assumptions. To that end, we carried out targeted capture and massively parallel sequencing of the exomes of twelve humans. These include eight HapMap individuals representing three populations, and four unrelated individuals with a rare dominantly inherited disorder, Freeman-Sheldon syndrome (FSS). We successfully interrogated ~96% of each human exome and demonstrate the sensitive and specific identification of rare and common variants in over 300 megabases (Mb) of coding sequence, capturing both single nucleotide substitutions and small insertion-deletions. Importantly, using FSS as a proof-of-concept, we show that candidate genes for monogenic disorders can be identified by exome sequencing of a small number of unrelated, affected individuals. Notably, however, this approach is critically dependent on the availability of adequate filters for common variants. In our analysis of FSS, the availability of the 8 HapMap exomes provided an essential supplement to dbSNP. Low-cost, high throughput technologies for exome resequencing have the potential to rapidly accelerate the discovery of candidate gene(s) and mutations that underlie rare monogenic diseases that have been resistant to conventional approaches. This strategy may also be applicable to diseases with more complex genetics through larger sample sizes and appropriate weighting of nonsynonymous variants by predicted functional impact.

## †3

**A Global Mechanism for Non-coding RNA Dependent Chromatin Formation in Mammals.** *A. Khalil[1, 2], M. Guttman[1], M. Huarte[1], L. Goff[1], M. Kellis[1], E. Lander[1, 2], J. Rinn[1, 2].* 1) The Broad Institute of Harvard and MIT, Cambridge, MA; 2) Harvard Medical School, Boston, MA.

One of the most fundamental and unsolved problems in biology is: how does the same genome present in every cell of an organism encode a multitude of different cellular states? While epigenetic regulation by chromatin modifying complexes plays a key role in this process, it is not currently known how these complexes are targeted to specific regions of the genome. Recently, we and others have found two large intergenic non-coding RNAs (lincRNAs) known as *HOTAIR* and *XIST* to physically associate with the chromatin modifying complex PRC2 and guide it to specific genomic loci. Thus, we hypothesized that association of lincRNAs with chromatin modifying complexes may be a general theme in the establishment and maintenance of epigenetic states. To assay lincRNA interactions with chromatin modifying complexes on a genome-wide scale we developed RIP-Chip technology: RNA co-immunoprecipitation followed by hybridization to tilling arrays. Remarkably, we found that 20% of expressed lincRNAs are bound by PRC2 in several human cell types. By examining several other chromatin modifying complexes, we found that 56% of lincRNAs are associated with these complexes. We then confirmed the localization of these lincRNAs to chromatin by RNA *in situ* hybridization. Through loss-of-function experiments, we found PRC2-associated lincRNAs to be required for proper expression of specific regions of the genome which are known to be regulated by PRC2. We have also profiled lincRNAs across multiple cancer types and discovered numerous lincRNAs that are both misregulated in cancer and bound to chromatin modifying complexes. Moreover, p53 directly and temporally induces several chromatin-associated lincRNAs in response to DNA damage. Remarkably, these lincRNAs serve to regulate many key genes in the p53 pathway. Together, these results point to a general mechanism of lincRNA mediated regulation in key cancer processes via the guidance of chromatin modifying complexes. Furthermore, RIP-seq technology has revealed a plethora of small RNAs that map to peaks of various chromatin modifications. Our results are significant since: 1) This is the first genome-wide characterization of ncRNAs interaction with chromatin modifying complexes, 2) They suggest that both small and large ncRNAs can establish epigenetic landscapes by guiding chromatin modifying complexes to specific regions of the genome, and disruption of these interactions may represent a novel paradigm in cancer etiology.

† Trainee Award Finalist

## 4

**Meta-analysis of >100,000 individuals identifies 63 new loci associated with serum lipid concentrations.** *T.M. Teslovich[1], K. Musunuru[2,3], A.V. Smith[4], S. Ripatti[5,6], C. van Duijn[7], J. Rotter[8], D. Chasman[9], E. Boerwinkle[10], L.A. Cupples[11], R. Krauss[12], V. Gudnason[4], D. Rader[13], M. Sandhu[14,15], J. Kooner[16], I. Borecki[17], P. Munroe[18], L. Pentonen[2,5,19], M. Boehnke[1], G. Abecasis[1], S. Kathiresan[2,3], Global Lipids Genetics Consortium.* 1) Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI; 2) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 3) Center for Human Genetic Research and Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA; 4) Icelandic Heart Association, Kopavogur, Iceland; 5) Institute for Molecular Medicine Finland FIMM, Helsinki, Finland; 6) Public Health Genomics Unit, National Institute for Health and Welfare, Helsinki, Finland; 7) Department of Epidemiology, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands; 8) Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, CA; 9) Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA; 10) Human Genetics Center, University of Texas Health Science Center, Houston, TX; 11) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 12) Children's Hospital of Oakland Research Institute, Oakland, CA; 13) Cardiovascular Institute and Institute for Translational Medicine and Therapeutics, University of Pennsylvania School of Medicine, Philadelphia, PA; 14) Department of Public Health and Primary Care, Strangeways Research Laboratory, University of Cambridge, Cambridge, UK; 15) MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK; 16) National Heart and Lung Institute, Imperial College London, London, UK; 17) Division of Statistical Genomics, Center for Human Genome Sciences, Washington University School of Medicine, St. Louis, MO; 18) 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, UK; 19) Wellcome Trust Sanger Institute, Cambridge, UK.

Serum concentrations of low-density lipoprotein cholesterol, high density lipoprotein cholesterol, and triglycerides are among the most important risk factors for atherosclerotic cardiovascular disease and are targets for therapeutic intervention. Recent genome-wide association studies (GWASs) involving ~20,000 individuals have identified DNA sequence variants contributing to inter-individual variation in blood lipid traits. It has recently been suggested that conducting genetic studies with ever larger cohorts will be relatively uninformative for the biology of complex human disease. To test the hypothesis that additional loci of biologic relevance can be identified with progressively larger-scale GWASs, we performed meta-analyses of 40 GWASs with a combined total of 100,197 samples of European ancestry. We validate each of the 33 loci identified by previous genome-wide association studies at genome-wide significance ($p < 5x10^{-8}$), and report 63 additional loci that show genome-wide significant association with these traits for the first time. The newly reported associations include SNPs near *LDLRAP1*, a gene responsible for autosomal recessive hypercholesterolemia; *NPC1L1*, a target of the cholesterol-lowering drug ezetimibe; *SCARB1*, a liver-specific HDL receptor; *STARD3*, a cholesterol transport gene; and *LRP1* and *LRP4*, members of the LDL receptor-related protein family. In addition, there are scores of new genes previously not implicated to play a role in lipoprotein metabolism. Many of these novel loci have moderate effect sizes but lower minor allele frequencies, so that previous, smaller meta-analyses were underpowered to identify their effects. Notably, many of the loci identified in this study are significantly associated with lipid levels in samples of Indian Asian and East Asian origin; indeed, we observed surprisingly little heterogeneity in effect size estimates between European and non-European samples, suggesting that variants implicated by this study are relevant to populations around the globe. Taken together, these findings demonstrate the utility of progressively larger GWASs to identify new determinants of serum lipoprotein concentrations.

## 5

**Long Range Effects of CNVs Mediated by *cis*-Acting Promoter Competition.** *K. Lower[1], J. Hughes[1], M. De Gobbi[1], S. Henderson[2], V. Viprakasit[3], C. Fisher[1], A. Goriely[1], H. Ayyub[1], J. Sloane-Stanley[1], D. Vernimmen[1], C. Langford[4], D. Garrick[1], R. Gibbons[1], D. Higgs[1].* 1) MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford, UK; 2) National Haemoglobinopathy Reference Laboratory, Oxford Radcliffe Hospitals NHS Trust, Oxford, UK; 3) Department of Paediatrics, Faculty of Medicine, Siriaj Hospital, Mahidol University, Bangkok, Thailand; 4) Microarray Facility, The Wellcome Trust Sanger Institute, Cambridge, UK.

Copy number variations (deletions, insertions and duplications over 1kb in size; CNVs) have recently been found to account for a substantial proportion of normal inherited variation in gene expression. In addition, CNVs have frequently been found to underlie both simple and complex human genetic diseases. Interestingly, it has recently been shown that CNVs may alter the expression of neighboring genes, with up to 50% of genes affected by CNVs found to lie 10s to 100s kb outside the region of aneuploidy. The mechanisms by which this occurs are largely unknown, resulting in recent calls for detailed functional analysis of such effect. Here we show that common CNVs of the $\alpha$ globin genes, which occur in >50% of individuals in some ethnic groups, radically alters expression (by eight-fold) of an unrelated gene (*NME4*) lying 300kb away. *NME4* displays erythroid-specific transcription, which we show is under the direct regulation of the $\alpha$ globin upstream regulatory elements (MCS-Rs). Removal of the $\alpha$ globin genes results in an up-regulation of the expression of *NME4*, which occurs *in cis* to the deletion. Modified 4C analysis confirms that this *in cis* up-regulation is mediated by a physical interaction between the MCS-R and *NME4*. Therefore, this long-range effect on gene expression is due to promoter competition between the CNV genes ($\alpha$ globin) and the distal gene (*NME4*) for a shared upstream regulatory element. This represents an as yet unidentified mechanism for how CNVs may exert long-range effects on gene expression. These findings describe a new mechanism by which CNVs may commonly perturb gene regulation across a large segment of the genome, thus contributing to normal, heritable variation in gene expression and human genetic disease.

## 6

**Tospeak, a recently evolved primate gene, is disrupted in a family with severe vocal impairment.** *R.A Clarke[1,2], B.J Shen[2], A.Q Wei[2], Z. Zhao[3,4], A. Guo[4], Z.M Fang[1].* 1) St George Hosp, Sch Medicine, Univ of NSW, Kogarah, NSW 2217, Australia; 2) Orthopaedic Research Institute, University of NSW, St George Hospital, Kogarah, NSW 2217, Australia; 3) Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, Tennessee, USA; 4) Department of Psychiatry and Human Genetics and Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, Virginia 23298, USA.

A child's capacity to speak emerges as the laryngeal vocal apparatus (LVA) matures. The breath-taking utility of this developmental process evolved in primates through modifications to the anatomy, flexibility and skeletal morphology of the LVA and a gene that regulates this evolutionary development has not been identified. We have studied a unique 5 generation pedigree, KF2-01, in which a severe vocal impairment disorder is transmitted as an autosomal-dominant trait with the persistence of an infantile laryngeal configuration, dysmorphology of laryngeal cartilages, immobility of the vocal cords and fusion/ossification of specific joints in the carpals, tarsals and spine. Our previous work localised the translocation breakpoint responsible to region 8q22.2 on chromosome 8. Here we show that a previously undescribed gene, tospeak, is directly disrupted by the breakpoint. Tospeak is a putative non-coding gene that spans an expansive (542 kb) region downstream of growth and differentiation factor 6 (GDF6/BMP13), a member of the bone morphogenetic protein (BMP) gene family. Loss-of-function mutations in GDF6 are associated with malformation of laryngeal cartilages in mice and fusion of specific joints comparable to the vocally impaired members of the family. We show that the breakpoint disrupts tospeak transcription across a GDF6 enhancer known to regulate GDF6 expression in these specific joints during development. We show that the 5' end of tospeak overlaps another recently evolved gene, smalltalk, and that the tospeak promoter, which continued to evolve in human, enhances transcription in vitro ~4 fold compared with chimpanzee. We propose that the parallel evolution of tospeak transcription and LVA maturation in primates may be linked through tospeak transcription mediated modulation of GDF6's role in joint development and flexibility.

## 7

**THE NEXT-GENERATION VIEW OF THE CANCER GENOME.** *G. Getz, K. Cibulskis, M. Lawrence, D. Chiang, S.L. Carter, A. Sivachenko, E. Banks, M.A. DePristo, T. Fennel, G. Saksena, L. Ambrogio, S. Fisher, C. Nusbaum, C. Sougnez, W. Winckler, T. Golub, M. Meyerson, E. Lander, S. Gabriel.* Broad Institute of MIT and Harvard, Cambridge, MA.

Massively parallel sequencing has dramatically increased our ability to comprehensively and accurately characterize the genomic alterations in tumor genomes. During the past year we have witnessed the initial steps of this revolution, from sequencing of a single tumor genome to currently sequencing and analyzing ~20 tumor/normal pairs from various tumor types, as well as sequencing of whole exomes in a large number of samples.

Distilling these vast amounts of data to the small number of somatic events required the development of highly accurate analysis tools which are both sensitive and specific. These were then stringed together to form a systematic and robust analysis pipeline going from raw sequencing output to significant events in cancer.

We have developed new methods to detect a large variety of somatic alterations in data generated by sequencing whole-genomes, -exomes (targeted regions) and -transcriptomes (RNA-seq). These methods enabled us to: (i) call somatic mutations; (ii) detect small indels; (iii) identify chromosomal rearrangements; (iv) detect copy-number alterations; (v) estimate the purity and ploidy of a tumor DNA sample; (vi) estimate expression levels of transcripts and variants; and (vii) detect fusion-gene events.

We will describe the results of applying these methods to a wide variety of data sets, including ~20 whole cancer genomes, several transcriptomes and dozens of whole- and partial-exomes.

The emerging picture is that tumor genomes vary in terms of their mutations rates and, even more pronounced, in terms of the number of rearrangements and copy-number alterations they harbor -- some genomes have very few alterations and others are highly rearranged. The ploidy index also varies among tumors, ranging from close to diploid to ~2.8-fold more DNA/cell than a normal cell. In ovarian cancers, for example, we have detected rearrangements involving tumor-suppressor genes (both translocations that truncate the gene and a rearrangements that produces a fusion-gene) suggesting that rearrangements is another mechanism of tumor-suppressor gene deactivation.

## 8

**Sequencing Cancer Genomes.** *S. Nelson[1], B. Merriman[1], Z. Chen[1], N. Homer[1], B.D. O'Connor[1], H. Lee[1], M. Clark[1], C. Flanders[1], K. Das[1], J. Klejnot[1], A. Eskin[1], K. McKernan[2].* 1) Dept Human Genetics, UCLA, Los Angeles, CA; 2) Life Technologies, Beverly, MA.

Cancer originates from genetic mutations, and thus our understanding of cancer is poised for a major advance as it becomes practical to sequence cancer genomes and catalog all mutations. We present our first efforts at complete resequencing of cancer genomes to find all mutational variants, in support of our own research on the molecular pathology of brain cancers. We sequence three genomes: a widely studied glioblastoma cell line, U87-MG, and a tumor-normal pair from a glioblastoma patient. Sequencing is performed using the AB SOLiD 3 platform to generate over 75 Gigabases of mappable reads per genome, in the form of paired-end 50+50 base reads. A variety of analytical tools are used to detect single base variants, small and intermediate indels, and genomic rearrangement breakpoints. Validation is done on a large scale by comparison to million SNP chip data, to targeted resequencing of all exons from 5235 genes from the Sanger COSMIC cancer gene set using the Illumina GAII platform, and to whole-transcriptome sequencing. For detecting single base variants, we achieve 97% power and a false positive rate below 1 in 1000, and favorable rates for indels and breakpoints as well. Comparing the U87-MG genome to the human reference yields over 300,000 variants not in dbSNP, and over one hundred instances of genes for which all copies present have an amino-acid changing or null mutation, including a PTEN mutation described in the literature. We also observe and validate several inter-chromosomal rearrangements not detected in published SKY data for U87-MG. Of special interest in the patient tumor is an amplified region on chromosome 12q14-15 that has been previously characterized in other cancers in the literature as consisting of tens of copies of both ring and rod structures of unknown complexity. In this tumor, the sequence data show portions of this region amplified over 1000 fold, far beyond the ~10 fold estimated from our array CGH data. Further, the breakpoints in this region are extremely complex, with tens of breakpoints and hundreds of rearrangement forms within the population of tumor cells. In summary, our initial experience sequencing cancer genomes shows that existing sequencing technologies can deliver whole genome sequence that provides, at a practical level, a comprehensive and accurate view of the mutational state of cancers and reveals the true complexity of genomic mutations.

## 9

**The complete genome sequence of a human melanoma.** *R.K. Cheetham[1], E.D. Pleasance[2].* 1) Illumina Inc, Chesterford Research Park, Little Chesterford, Saffron Walden, Essex, United Kingdom; 2) Welcome Trust Sanger Institute, Hinxton, Cambridge, UK.

Accurate and rapid sequencing of individual human cancer genomes paves the way to discover the full catalogue of somatic mutations that underlie the molecular basis of tumorigenesis. We sequenced the genome of the melanoma cell line COLO-829 to 40x coverage and the corresponding control genomic DNA to 30x using Illumina's Genome Analyzer. We characterised the profile of candidate somatic mutations (gain of novel variant in the tumour sample compared to the normal control) and identified more than 30,000 single base substitutions. Two-thirds of the substitutions are C to T changes; 98 percent of these occur in CC or TC dimers, consistent with the expected mutational spectrum of UV-induced events in melanoma. This dataset provides a previously unparalleled source of information to characterise the influence of sequence context on somatic mutations in melanoma. The more than 300 protein coding changes include 44 out of the 50 somatic variants detected previously in this sample by exon sequencing, confirming a high sensitivity of detection in the whole genome data. We are also able to identify somatic short indels, structural variants, and copy number changes. This study illustrates the precision of somatic mutation detection in whole genome sequencing and the feasibility for large-scale studies to identify novel genetic elements that are involved in tumorigenesis and may be used as diagnostic markers.

## 10

**Next generation sequencing of the genome of relapsed acute lymphoblastic leukemia.** *C.G. Mullighan[1], L.A.A. Phillips[1], C. Lee[2], K. McKernan[2], H. Peckham[2], Q. Doan[3].* 1) Pathology, St Jude Children's Research Hospital, Memphis, TN; 2) Life TEchnologies, Beverly, MA; 3) Life Technologies, Foster City, CA.

Acute lymphoblastic leukemia (ALL) is a cancer of leukocyte progenitors, and the commonest cancer in childhood. Relapse of ALL is a leading cause of pediatric cancer death, and occurs in most adult ALL patients. Our prior work profiling genetic alterations in relapsed ALL using SNP microarrays and candidate gene resequencing identified substantial evolution in the nature of genomic abnormalities from diagnosis to relapse, including the acquisition of novel mutations in transcription factors, tumor suppressors and kinases at relapse. These findings indicate that comprehensive identification of all sequence variants in relapsed ALL will be important to obtain a complete understanding of the genetic changes contributing to relapse, and to identify novel targets for therapy. To identify all sequence variants, we are performing next-generation, massively parallel transcriptomic resequencing of purified leukemic samples obtained at diagnosis and relapse of five patients with ALL using the Applied Biosystems SOLiD Sequencing system. Variant data is being compared to data generated by whole genome resequencing of non-leukemic DNA from each patient in order to distinguish tumor-acquired from inherited variants. Simultaneous sequencing of both diagnosis and relapse samples is enabling the the identification of mutations present at diagnosis, and those that are acquired at relapse. Novel variants will be tested in a panel of over 60 matched diagnosis-relapse ALL sample pairs to determine the frequency of mutation in each gene. Data analysis is ongoing, and will be presented in full at the meeting.

## 11

**The DNA Methylome of Pediatric Acute Lymphoblastic Leukemia.** *J. Davidsson[1], H. Lilljebjörn[1], A. Andersson[1], S. Veerla[1], J. Heldrup[2], M. Behrendtz[3], T. Fioretos[1], B. Johansson[1].* 1) Dept. of Clinical Genetics, Lund University Hospital, Lund, Sweden; 2) Dept. of Pediatrics, Lund University Hospital, Lund, Sweden; 3) Dept. of Pediatrics, Linköping University Hospital, Linköping, Sweden.

Acute lymphoblastic leukemia (ALL) is the most common childhood malignancy, with high hyperdiploidy (51-67 chromosomes) and the t(12;21)(p13;q22) (*ETV6/RUNX1* fusion) representing the most frequent abnormalities. Although these arise in utero, there is long latency before overt ALL, showing that additional changes are needed. Gene dysregulation through hypermethylation may be such an event; however, this has not previously been investigated in a detailed fashion. We performed genome-wide methylation profiling using bacterial artificial chromosome arrays and promoter-specific analyses of high hyperdiploid and *ETV6/RUNX1*-positive ALLs. In addition, global gene expression analyses were performed to identify associated expression patterns. Unsupervised cluster and principal component analyses of the chromosome-wide methylome profiles could successfully subgroup the two genetic ALL types. Analysis of all currently known promoter-specific CpG islands demonstrated that several B-cell- and neoplasia-associated genes were hypermethylated and underexpressed, indicating that aberrant methylation plays a significant leukemogenic role. Interestingly, methylation hotspots were associated with chromosome bands predicted to harbor imprinted genes and the tri-/tetrasomic chromosomes in the high hyperdiploid ALLs were less methylated than their disomic counterparts. Decreased methylation of gained chromosomes is a previously unknown phenomenon that may have ramifications not only for the pathogenesis of high hyperdiploid ALL but also for other disorders with acquired or constitutional numerical chromosome anomalies.

## 12

**Detection of tumor-specific somatic mutations by transcriptome sequencing of a cytogenetically normal acute myeloid leukemia.** *A. Benet-Pages[1], S. Eck[1], P.A. Greif[2,3], H. Popp[2], A. Dufour[2], T. Meitinger[1,4], T.M. Strom[1,4], S.K. Bohlander[2,3].* 1) Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg-Munich, Germany; 2) Department of Medicine III, Universität München, Munich, Germany; 3) Clinical Cooperative Group "Leukemia", Helmholtz Zentrum München, German Research Center for Environmental Health, Munich, Germany; 4) Institute of Human Genetics, Technische Univeristät München, Munich, Germany.

Approximately half of acute myeloid leukemia (AML) patients have at least one chromosomal aberration, whereas the other half classifies as cytogenetically normal (CN-AML). Most of the genetic events that initiate the disease are still undiscovered. To identify tumor-specific somatic coding mutations, we sequenced the transcriptome of a CN-AML and a matched remission sample by second-generation sequencing technology (Illumina GAII). SNPs were called with the MAQ software. Additional filters were applied to exclude known and possible sequencing artefacts. We generated 20.4 and 15.6 million 32 bp paired-end reads of the CN-AML and remission sample, respectively, which mapped to exons of UCSC genes. 8.9% of reads for the AML and 5.0% reads of the remission sample mapped to intergenic regions. Of the 11178 transcripts with a higher expression than 60 reads per gene (corresponding to approximately 1 transcript per cell), we sequenced 5911 with an average coverage of greater than seven. By comparing the 63159 SNPs discovered in the CN-AML sample with the respective results in the remission sample, we identified 3 non-synonymous mutations not present in either the remission sample or in dbSNP which were confirmed by capillary sequencing. Among them is a nonsense mutation affecting the RUNX1 gene, which forms a well known fusion gene in AML (RUNX1/RUNX1T1), and a missense mutation in the tumor-associated gene TLE which interacts with RUNX1. A second missense mutation was identified in the SHKBP1 gene which acts downstream of the AML-associated FLT3 regulatory pathway. Two additional mutations were false positives. These results demonstrate that our technique of transcriptome sequencing is an efficient method to discover novel mutations in AML.

## 13

**Whole Exome and Whole Transcriptome sequence of four human tumor cell lines.** *M. Dean[1], L. Li[2], K. McGee[1], M. Nickerson[1], B. Gold[1], P. Bouffard[2], Z. Markovic[2], A. Goncalves[2], Q. Zhao[2], B. Boese[2], T. Harkins[3].* 1) Cancer & Inflammation Program, NCI-FCRDC, Frederick, MD; 2) 454 Life Sciences, A Roche Company; 1 Commercial Street, Branford, CT; 3) Roche Applied Science 9115 Hague Rd, Indianapolis, IN.

Technology to sequence all of the exons of the human genome (the exome) as well as the transcripts (the transcriptome) provides an unprecedented opportunity for insight into the cancer genome and the development of cancer biomarkers. To validate this approach we have deeply sequenced the exome and transcriptome of 4 well characterized human tumor cell lines from the NCI-60 panel; PC3 (prostate), MCF7 (breast), SN12C and 786-0 (kidney). These cell lines have extensive characterization on gene expression, chromosome structure, and on the cytotoxicity of over 100,000 compounds. Alignment to the consensus genome permitted identification of genomic sequence variants and 2583-5307 variants were identified including 92-297 novel non-synonymous variants identified in each cell line. The variants also include many putative frameshift and non-sense mutations. Array genotype data on the same samples were accessed (ftp-snp500cancer.nci.nih.gov/nci60) and used to confirm the SNP calls and revealed only a small discordance: 1% of sites were inconsistent with the array data. Known mutations in tumor suppressor genes, in these cell lines, such as *TP53*, *PTEN*, *MLH1*, *CDKN2A* and *VHL* were confirmed, and a comparison of 35 SNPs in the coding region of cancer-related genes were highly concordant between the Exome and Transcriptome analyses. We also identified multiple regions of homozygosity at an unprecedented resolution. For example, in the clear cell renal cancer cell line SN12C, the known mutation in *VHL* (311delG) was identified and LOH across the entire length of chromosome 3 was confirmed. Novel non-synonymous variants identified include the cancer-related genes *HIF1A*, *ERBB4*, *IL5RA*, *ATM*, *FANCM*, *FANCI*, *AR*, and *BLM*. Pathway analysis of these novel non-synonymous variants in GeneGo revealed insight into potential functions of the newly identified variants and candidates for targeted therapy. For example, in the MCF7 breast cancer cell line, 7 genes interacting with MYC, 8 interacting with P53, 7 with ubiquitin/caspases, and 5 genes with estrogen receptor contain previously uncharacterized variants. This multi-pronged approach, if applied to individual malignancies, has the potential to accelerate the identification of susceptibility loci and develop targets for drug development and rational therapies.

## 14

**Aberrant microRNA expression in retinoblastoma by high-throughput next generation sequencing.** *E. Chao, J. Schug, G. Grant, A. Ganguly.* Department of Genetics, University of Pennsylvania School of Medicine, Philadelphia, PA.

Retinoblastoma (RB) is the most common intraocular malignancy of childhood. Carriers of a germline RB1 mutation are at high risk of developing cancer (93% penetrance). Aggressive screening measures are currently required to identify early-stage disease with no therapeutic options available prior to cancer development. MicroRNAs (miRNA) are short, 21-24bp, oligonucleotides whose dysregulation plays an important role in cancer. Moreover, studies have shown that these small molecules have many advantages as potent, non-toxic chemoprevention agents. We propose to identify miRNA aberrations in RB as a first step towards development of effective chemopreventtion agents in this high-risk patient population. Towards this goal, we profiled miRNA expression in normal retina, RB tumors and RB cell lines by microarray (Agilent) and high-throughput next-generation sequencing (Illumina). We generated nearly 2 million reads from each sample, which mapped to 540 known miRNA sequences. Thirty-six miRNAs were differentially expressed ($p<0.05$) in tumors compared with normal retinal tissue. A 24-miRNA signature differentiates normal tissue from cell lines, well, and poorly-differentiated tumors. These results were confirmed by quantitative PCR. Several members of the let-7 family were highly dysregulated; let-7c is nearly absent in tumors, with 9-fold down-regulation ($p=0.008$), suggesting a role for this miRNA as a tumor suppressor. To confirm this role we will transfect RB cell lines deficient in let-7c with synthetic miRNA and assay for changes in cell survival, cell proliferation and cellular invasion compared with controls. We hypothesize that these experiments will demonstrate a functional role for let-7c in tumorigenesis, and may elucidate a potential candidate for miRNA-based chemoprevention in retinoblastoma.

## 15

**The Cancer Genome Atlas study on serous ovarian carcinoma: preliminary analysis of copy number alterations.** *W. Winckler[1], G. Getz[1], G. Saksena[1], S. Carter[1], B. Weir[1], M. Lawrence[1], R. Onofrio[1], S. Gabriel[1], E. Lander[1], M. Meyerson[1,2] for The Cancer Genome Atlas Research Network.* 1) Broad Institute, Cambridge, MA; 2) Dana Farber Cancer Institute, Boston, MA.

The Cancer Genome Atlas (TCGA) is a comprehensive effort to delineate the molecular basis of cancer through the application of genome analysis technologies including DNA sequencing, copy number, gene expression, and epigenetic analysis. Launched by the U.S. National Cancer Institute and National Human Genome Research Institute, the pilot project will assess the feasibility of a full-scale effort to systematically explore the entire spectrum of genomic changes involved in human cancer. TCGA recently reported an integrative genomic analysis of glioblastoma (*Nature*, 2008), and is now investigating serous ovarian carcinoma. Here we present preliminary analysis of copy number alterations in the first 162 ovarian carcinoma specimens. Each sample was separately analyzed using Affymetrix SNP 6.0, Agilent 244K, Agilent 1M, and Illumina 1M Duo arrays. Copy number analysis of the combined data using GISTIC (*PNAS*, 2007) reveals that these serous ovarian carcinomas are highly rearranged, with over 40 regions of statistically significant recurrent focal amplification and another 40 such regions of focal deletion. Within these regions of recurrent alteration, we observe previously known genes, including loss of *PTEN* and *RB1* and gain of *CCNE1* and *MYC*. Additionally, many regions of significantly recurrent alteration contain novel plausible biological candidate genes and/or miRNAs, some known in other tumor types (e.g. loss of *NF1*) and others with less prior evidence of involvement in cancer. Recurrent copy neutral loss of heterozygosity is observed, most notably on 17q. The inclusion of >100 planned additional tumors will help to further refine the boundaries and identify the target(s) in each region. Integration of these copy number results with mutation, gene expression, and methylation data from the same samples enables a more comprehensive assessment of the genes and pathways that are dysregulated in ovarian carcinoma. As in the glioblastoma study, integrative analysis identifies pathways that are altered in a large fraction of patients. These results suggest that the completion of the TCGA ovarian carcinoma project and the expansion to additional tumor types with a similarly well-powered design will expand knowledge of the molecular basis of cancer and hopefully identify novel targets for therapeutic intervention.

## 16

**Elucidation of tumor aneuploidy, non-aberrant cell infiltration and LOH from genome wide SNP data in breast carcinomas.** *P. Van Loo[1,2,7], S.H. Nordgard[1,7], O.C. Lingjærde[3], H.E. Giercksky Russnes[1], W. Sun[4,5], P. Marynen[2], C.M. Perou[5], B. Naume[6], A.-L. Børresen-Dale[1,8], V.N. Kristensen[1,8].* 1) Department of Genetics, The Norwegian Radium Hospital, University of Oslo, Oslo, Norway; 2) Department of Human Genetics, VIB and University of Leuven, Leuven, Belgium; 3) Biomedical division, Department of Informatics, University of Oslo, Oslo, Norway; 4) Department of Biostatistics, University of North Carolina, Chapel Hill, NC; 5) Department of Genetics, University of North Carolina, Chapel Hill, NC; 6) The Cancer Clinic, The Norwegian Radium Hospital, Oslo, Norway; 7) These authors contributed equally to this work; 8) These authors share senior authorship.

Whole genome SNP genotyping can measure copy number aberrations of tumors in an allele specific manner. The analysis of these SNP arrays has caught particular interest, because of the promises to deduce aneuploidy of cancer cells, infiltration of normal tissue and tumor heterogeneity. Technically however this has proven difficult, especially for whole cancer genomes. We have developed an algorithm to accurately dissect the allele specific copy number from SNP array data of cancer samples. Our method infers the aneuploidy of the tumor cells and the percentage of infiltrating non-aberrant cells, and calculates a genome-wide allele specific copy number profile correcting for both these properties. We applied this method to a set of 112 breast carcinomas and evaluated the results by visual inspection of the resulting copy number profiles one by one. For 55 cases (49 %), we concluded that the method's interpretation is likely correct. In 11 of these cases, this was verified by focused FISH experiments. Upon close inspection of the remaining 57 cases, we observed multiple local minima in the (percentage non-aberrant cell infiltration - ploidy) parameter space, and we manually excluded solutions deemed incorrect. This approach allowed us to find an acceptable solution in an additional 46 cases (41 %). The remaining 11 cases (10 %) had to be excluded from our further analysis, as no acceptable solution was found, possibly due to noise in the profiles, too much non-aberrant cell involvement or too high tumor heterogeneity. We find that the 101 breast carcinomas were on average infiltrated with 52 % non-aberrant cells and that 51 % of them are aneuploid with an average copy number of 3 or higher. These features could not be deduced from array-CGH data and would influence the interpretation of copy number gains and losses. Furthermore, subdividing our cases into the 5 previously identified breast cancer subtypes, we find characteristic recurring patterns of loss of heterozygocity (LOH) in each of the five subgroups. Finally, subgroup profiles of gains and losses can be made apparent that (due to the non-tumor cell infiltration) would not be visible by array CGH. In conclusion, our novel algorithm to infer genome-wide allele specific copy number profiles from SNP array data highlights considerable tumor aneuploidy and non-aberrant cell infiltration, and brings forward characteristic genome-wide patterns of gains, losses and LOH in breast carcinomas.

## 17

**Developmental changes in human neocortical transcriptome revealed by RNA-Seq.** *C.E. Mason[1,2], M. Johnson[3], K. Bilguvar[1,4], J. Marioni[5], N. Carriero[6], P. Zumbo[7], S. Mane[7], M. Gunel[1,3], N. Sestan[3,8], J. Noonan[2], M.W. State[1,2].* 1) Program on Neurogenetics, Yale University, New Haven, CT, 06511, USA; 2) Department of Genetics and Yale Child Study Center, Yale University School of Medicine, New Haven, CT, 06511, USA; 3) Department of Neurobiology, Yale University School of Medicine, New Haven, CT, 06511, USA; 4) Department of Neurosurgery, Yale University School of Medicine, New Haven, CT, 06511, USA; 5) Department of Human Genetics, University of Chicago, Chicago, Illinois, 60637, USA; 6) Yale High Performance Computing Cluster, Yale University, New Haven, CT, 06511, USA; 7) Keck Biotechnology Laboratory, New Haven, Connecticut, 06520, USA; 8) Kavli Institute for Neuroscience, Yale University School of Medicine, New Haven, CT, 06511, USA.

Human brain development is a highly complex process that depends on the precise temporal execution of genetic events orchestrated by the expression of key transcription factors. Slight missteps in this developmental process can lead to severe neurological disorders; this is especially true of the cerebral cortex, the seat of our higher cognitive abilities. Understanding this process has been helped by previous large-scale expression studies, but most are from model organisms. A small number of recent exceptions have revealed genome-wide gene expression and splicing patterns in the developing human brain, but how these expression patterns compare to the adult has never been investigated. To explore in detail the normal genetic mechanisms of development in the brain, we used RNA-seq to directly compare the transcriptome of the adult human temporal lobe to the transcriptome of the same area at late mid-gestation. Our goal was to leverage RNA-Seq to identify specific transcripts that are enriched in fetal or adult temporal lobe, as well as enabling more insights into the other noncoding transcriptional dynamics of human cortical development. After confirming the reproducibility, sensitivity and specificity of our RNA-seq data, we found that the majority (71%) of annotated genes are active in the neocortex. Using conservative statistical thresholds, we identified approximately 6,200 transcripts whose expression patterns were significantly enriched at either the fetal or adult stage. We also found high levels of transcription outside previously identified gene units, which was more pronounced in fetal versus adult brain. We did not observe differences in either the number of genes expressed, the amount of splicing, or the number of transcripts exhibiting alternative splicing between the fetus and adult, suggesting that early brain development is not distinguished by global increases in transcriptional activity or isoform variation. Rather, specific families of genes (transcription factors, cadherins, RNA-binding motifs) showed altered patterns of expression in either the fetal or adult neocortex. In particular, we identified a specific upregulation of the zinc finger (ZNF) and SRY-related HMG-box (SOX) transcription factor family in the fetal temporal neocortex, including some possible human-specific genes, suggesting critical roles for these regulatory factors in the transcriptional orchestration of temporal neocortical development.

## 18

**Genome-Wide Combined Linkage/Association Scan Localizes Two QTLs Influencing Human Caudate Nucleus Volume.** *D. Glahn[1,3], A. Winkler[1], J. Curran[2], M. Carless[2], J. Charlesworth[2], M. Johnson[2], H. Göring[2], T. Dyer[2], E. Moses[2], L. Almasy[2], P. Fox[3], P. Kochunov[3], R. Olvera[4], R. Duggirala[2], J. Blangero[2].* 1) Olin Neuropsychiatry Research Center, Institute of Living, & Department of Psychiatry, Yale University School of Medicine, Hartford, CT; 2) Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX; 3) Research Imaging Center, UT Health Science Center, San Antonio, TX; 4) Department of Psychiatry, UT Health Science Center, San Antonio, TX.

The caudate nucleus, a brain region located within the basal ganglia, is associated with higher-order motor control, learning and memory, feedback processing, language and other executive functions. The caudate, which is innervated by dopaminergic neurons, is thought to be a locus of pathology for several neurological and psychiatric disorders (e.g. Schizophrenia and Bipolar Disorder). While the size and function of the caudate are thought to be under genetic control, the genes influencing the volume of this brain region are unknown. To better understand the genetic basis of caudate volume variation in the normal population, we undertook the first large-scale genome-wide search for genes influencing human brain structure. We collected high-resolution T1-weighted MRI scans from approximately 400 Mexican-American individuals from 30 extended pedigrees who participated in the Genetics of Brain Structure and Function study and who have Illumina HumanHap 550K SNP genotyping available. Imaging data was analyzed with FreeSurfer and genetic analyses were performed with SOLAR. The heritability of caudate volume was estimated to be 0.685 ($p=2.3\times10^{-10}$). Combined linkage/association analysis identified two regions exhibiting genome-wide significant evidence for harboring QTLs influencing caudate volume. One QTL was localized on chromosome 7p22 with a strongly associated SNP in the PRKAR1B gene (nominal p-value = $2.3\times10^{-8}$, genome-wide p-value = 0.02). This gene codes for the protein kinase, cAMP-dependent, regulatory, type I, beta, which is involved in brain development and linked to poor memory. Examination of the potential pleiotropic associations of this SNP with neurocognitive function revealed significant associations with processing speed ($p = 0.005$), episodic memory ($p = 0.04$) and working memory ($p=0.05$). Thus, this QTL appears to influence both brain structure and function. A second QTL influencing caudate volume was localized to chromosomal region 2p12 with strong association to a SNP near the LRRTM4 gene (nominal p-value = $1.3\times10^{-7}$, genome-wide p-value = 0.049) which codes for a neuronal protein of unknown function. Given the strength of these signals, we are currently deeply sequencing these genes to identify potential functional variants. Dissection of the genetic basis of normal human brain structural variation should lead to the identification of genes likely to be involved in disorders of brain structure/function.

## 19

**Genome-wide association of human amygdala volume identifies a QTL at *IRX2*.** *J. Charlesworth[1], D.C. Glahn[2], J.E. Curran[1], A. Winkler[2], M. Carless[1], M.P. Johnson[1], H.H.H. Göring[1], T.D. Dyer[1], E.K. Moses[1], L. Almasy[1], P.T. Fox[3], R. Olvera[3], P. Kochunov[3], R. Duggirala[1], J. Blangero[1]*. 1) Dept Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX; 2) Olin Neuropsychiatry Research Center, Institute of Living, & Department of Psychiatry, Yale University School of Medicine, Hartford, CT; 3) Research Imaging Center, UT Health Science Center, San Antonio, TX.

The brain is the most complex human organ with substantial structural variation amongst individuals, however little is known about the genetic basis of the anatomical variation. In this study we combined advanced imaging methods with modern genomic approaches to perform the first genome-wide search for genes influencing human brain structure, focusing on the amygdala, a critically important subcortical region. The amygdala plays a critical role in the detection and processing of emotional latent stimuli, emotional memory and fear conditioning. Abnormal amygdala volumes have been reported in a number of psychiatric and neurological disorders, including autism, bipolar disorder, major depression and Alzheimer's disease. The Genetics of Brain Structure and Function study involves 1,000 individuals from 40 Mexican American families previously investigated as part of the San Antonio Family Heart Study. To localize genes responsible for variation of amygdala volume in the normal population we collected high-resolution T1-weighted MRI scans. Imaging data was analyzed with FreeSurfer and genetic analyses were performed with SOLAR. The heritability of amygdala volume was estimated to be 0.76 ($p=2.6\times10^{-13}$) in our sample. We genotyped individuals using the Illumina HumanHap 550K beadchip and localized a QTL influencing amygdala volume on chromosome 5p15, with a strongly associated SNP in the *IRX2* gene (nominal p-value = $8.3\times10^{-8}$, genome-wide p-value = 0.03, based upon an empirically derived null distribution). This gene codes for the iroquois homeobox 2 protein which is known to be involved in brain development and acts as a transcription factor involved in neurogenesis. Focusing on this SNP, we identified multiple other brain regions that appear to be strongly pleiotropically influenced by the underlying QTL, including the rostral anterior division of the cingulate cortex that is known to be a main locus of *IRX2*. We are currently sequencing this gene in 192 founder individuals and will type all observed sequence variation in the complete sample to try to identify functional variants. Identification of the genetic determinates of amygdala volume in the normal population will provide empirically derived candidate genes for illnesses associated with amygdala dysfunction and the subsequent biological inferences that result from discovering a causal component of the pathological chain.

## 20

**Deep sequencing of VPS13A reveals effects on brain function and structure.** *C. Bellis[1], K.A Freed[1], J.E Curran[1], A. Winkler[2], M.A Carless[1], M.P Johnson[1], H.H.H Göring[1], T.D Dyer[1], L. Almasy[1], P.T Fox[3], P. Kochunov[3], R. Duggirala[1], D.C Glahn[2], E.K Moses[1], J. Blangero[1]*. 1) Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX; 2) Olin Neuropsychiatry Research Center, Institute of Living, & Department of Psychiatry, Yale University School of Medicine, Hartford, CT; 3) Research Imaging Center, UT Health Science Center, San Antonio, TX.

The genetic basis of variation in brain structure and function remains little known. This major set of endophenotypes is of great importance for a variety of mental and neurological disorders. In this study, we examine sequence variation in a candidate gene, VPS13A, using a sample of over 500 Mexican Americans in extended families who have been assessed for a number of brain anatomic and neurocognitive phenotypes. VPS13A is a 73 exon gene localizing to chromosome 9q21 which is responsible for a form of monogenic chorea-acanthocytosis, a neurodegenerative disease which is associated with schizophrenia, obsessive compulsive behaviors, impairments of frontal lobe executive skills, and multiple brain structure alterations. Genome-wide lymphocyte-based transcriptional analysis revealed significant correlations between VPS13A mRNA levels and aspects of executive function and multiple brain magnetic resonance imaging (MRI) measures including frontal lobe volume. These empirical correlations strengthened the support for VPS13A as a candidate gene potentially influencing normal variation in brain structure/function. In order to identify sequence variation, we extensively sequenced this large (over 200 kb) gene in 192 founder samples, focusing on coding regions, potential splice sites, conserved intronic regions, the 5' region and the 3' UTR. A total 320 variants were identified and genotyped in the whole sample. Of these, 270 variants were of sufficient frequency to permit reliable association analysis. Using eigenstructure methods, we determined that there were approximately 117 effective independent SNPs and employed this number to control for multiple testing. Using a variance component-based measured genotype approach that explicitly accounts for the non-independence amongst related individuals, we identified a SNP, rs13294928, which was highly significantly associated with multiple brain MRI measures. These included the frontal lobe volume ($p = 1.6\times10\text{-}5$), parietal lobe volume ($p = 8.6\times10\text{-}5$), and thalamus volume ($p = 1.1\times10\text{-}4$). Less strongly affected brain regions included the cingulate cortex ($p = 0.0027$), temporal lobe ($p = 0.0022$), accumbens ($p = 0.0064$), amygdala ($p=0.0029$), and putamen ($p = 0.0034$). These results suggest that deep sequencing of known candidate genes involved in monogenic neurological disorders can ultimately reveal potentially causal information relating to genes involved in both normal and common pathological brain structure/function.

## 21

**Genome-wide association study in partial epilepsies.** *D. Kasperaviciute[1], E.L. Heinzen[2], C. Depondt[3], C.P. Doherty[4], D. Leppert[5], R.A. Gibson[5], M.R. Johnson[5,6], R. Kälviäinen[7], K. Eriksson[8], G. Krämer[9], B.J. Steinhoff[10], D. Zumsteg[11], M. Ortega[11], N. Delanty[12], S.M. Sisodiya[1,13], D.B. Goldstein[2], GenEpA Consortium, EPIGEN Consortium.* 1) Department of Clinical and Experimental Epilepsy, UCL Institute of Neurology, Queen Square, London, WC1N 3BG, UK; 2) Institute for Genome Sciences and Policy, Center for Human Genome Variation, Duke University, Durham, NC 27708, USA; 3) Department of Neurology, Hôpital Erasme, Université Libre de Bruxelles, 1070 Brussels, Belgium; 4) The Department of Neurology, St James' Hospital Dublin, Ireland; 5) Genetics Division, Drug Discovery, GlaxoSmithKline; 6) Department of Clinical Neurosciences, Imperial College, Hammersmith Hospital, London W12 ONN, UK; 7) Kuopio Epilepsy Center, Kuopio University Hospital, Kuopio, Finland; 8) Pediatric Neurology Unit, Tampere University Hospital, Tampere, Finland; 9) Swiss Epilepsy Centre, Bleulerstrasse 60, 8008 Zurich, Switzerland; 10) Kork Epilepsy Centre, Kehl-Kork, Germany; 11) Department of Neurology, University Hospital Zurich, 8091 Zurich, Switzerland; 12) The Department of Clinical Neurological Sciences and Molecular and Cellular Therapeutics, RCSI Research Institute Royal College of Surgeons in Ireland, and Division of Neurology, Beaumont Hospital, Dublin, Ireland; 13) National Society for Epilepsy, Chalfont-St-Peter, Bucks, SL9 0RJ, UK.

Epilepsy is the commonest serious chronic neurological condition, and poses significant economic and personal burden. No genetic variants unequivocally associated with sporadic partial epilepsies have yet been identified despite the high heritability of the condition. To identify genetic risk factors for partial epilepsies we have performed genome-wide association study in 2814 patients with partial epilepsies of Caucasian origin from six centres (EPIGEN consortium www.epilepsygenetics.eu in collaboration with GenEpA Consortium) and 5300 controls (subjects without epilepsy and population controls from Wellcome Trust Case Control Consortium study). Subjects were genotyped using Illumina 610quad and 1Mduo chips. 523,940 SNPs that passed quality control criteria were included in the association analysis. Firstly, we looked for genetic variants increasing susceptibility to any focal-onset seizure type in the partial epilepsies shared across the partial epilepsy syndromes. Then, we performed subset analyses to look for genetic variants predisposing to temporal lobe epilepsy, mesial temporal lobe epilepsy with hippocampal sclerosis, susceptibility to secondary generalisation of partial seizures and susceptibility to febrile seizures. Three SNPs reached genome wide significance ($p<5x10^{-7}$) for association for susceptibility to focal-onset seizures (all partial epilepsy syndromes combined) and five SNPs for susceptibility to temporal lobe epilepsy. The strongest signal was detected in a gene-rich high LD region on chromosome 1p36 ($p=4x10^{-8}$, OR=1.26 [1.15-1.38]). Replication and follow-up analysis are under way. Interestingly, in our study, the strongest association signals were detected outside the channel genes, which are the most common causes of familial epilepsy syndromes, and therefore our work indicates new genes that may be important in understanding epilepsy. All above mentioned SNPs and susceptibility genes will be reported in the meeting.

## 22

**Genome-wide association study identifies UNC13A and 9p21.2 as susceptibility loci for sporadic Amyotrophic Lateral Sclerosis.** *M.A. van Es[1], J.H. Veldink[1], H.M. Blauw[1], P.W.J. van Vught[1], S. Schreiber[2], M.M. Nöthen[3], H.E. Wichmann[4], D. Rujescu[5], L.A. Kiemeney[6], A.C. Ludoplh[7], A.G. Uitterlinden[8], S. Cronin[9], O. Hardiman[9], J.E. Landers[10], A. Al-Chalabi[11], R.H. Brown Jr.[10], W. Robberecht[12], P.M. Andersen[13], R.A. Ophoff[14,15], L.H. van den Berg[1].* 1) Dept Neurology, University Medical Center Utrecht, Utrecht, Netherlands; 2) Institute for Clinical Molecular Biology, University of Kiel, Kiel, Germany; 3) Institute of Human Genetics, University of Bonn, Bonn, Germany; 4) Institute of Epidemiology, Helmholtz Center Munich, Germany; 5) 9Division of Molecular and Clinical Neurobiology, Department of Psychiatry, Ludwig-Maximilians-University, Munich, Germany; 6) Department of Urology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 7) Department of Neurology, University of Ulm, Ulm, Germany; 8) Department of Epidemiology, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands; 9) Department of Neurology, Beaumont Hospital, Dublin, Ireland; 10) Department of Neurology, University of Massachusetts School of Medicine, Worcester, USA; 11) Department of Clinical Neuroscience, King's College London, Medical Research Council (MRC) Centre for Neurodegeneration Research, Institute of Psychiatry, London, UK; 12) Department of Neurology, University Hospital Leuven, University of Leuven, Leuven, Belgium; 13) Institute of Clinical Neuroscience, Umeå University Hospital, Umeå, Sweden; 14) Department of Medical Genetics and Rudolf Magnus Institute, University Medical Center Utrecht, Utrecht, The Netherlands; 15) UCLA Center for Neurobehavioral Genetics, Los Angeles, USA.

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disorder characterized by the progressive loss of upper and lower motor neurons. Disease onset is usually in the late fifties and patients present with muscle weakness which spreads throughout the body leading to paralysis and eventually death due to respiratory failure. Currently the only medication to slow the disease is riluzole, which delays progression by 3 months. Despite the considerable heritability of the disease (0.35-0.85), the genetics of ALS remain largely unknown. Multiple GWAS have been performed and have not identified unambiguous associations. In this present study, we performed a genome-wide association study in sporadic ALS (2,323 patients and 9,013 controls) and evaluated all SNPs with $P<1.0x10^{-4}$ in a second, independent cohort of 2,532 patients and 5,940 controls. Analysis of the genome-wide data revealed genome-wide significance for one SNP, rs12608932 with $P=1.30x10^{-9}$. This SNP demonstrated robust replication in the second cohort ($P=1.86x10^{-6}$) and a combined analysis over the two stages yielded a $P=2.53x10^{-14}$. rs12608932 SNP maps to an LD block within the boundaries of UNC13A. UNC13A is a presynaptic protein found in central and neuromuscular synapses that regulates the release of neurotransmitters, peptides and hormones. Neurotransmitter release at synapses between nerve cells is mediated by calcium-triggered exocytotic fusion of synaptic vesicles with the plasma membrane. Before fusion, vesicles dock at the presynaptic release site, where they are primed to a fusion-competent state. It has been demonstrated that Munc13-1 (murine homologue) is required for synaptic vesicle priming and mice lacking Munc13-1 have disrupted glutamatergic neurotransmission, due to arrested synaptic vesicle maturation. In addition, Munc13-1 deficient mice display morphological defects in spinal cord motor neurons, muscle, and neuromuscular synapses. Follow-up of additional SNPs demonstrated genome-wide significance for two further SNPs (rs2814707 with $P=7.45x10^{-9}$ and rs3849942 with $P=1.01x10^{-8}$) in the combined analysis of both stages. These SNPs are located at chromosome 9p21.2 in a linkage region for familial ALS with fronto-temporal dementia found previously in several large pedigrees. The identification of 2 novel common variants that are robustly associated with susceptibility to sporadic ALS is an exciting step forward in understanding the genetics of this devastating disease.

## 23

**High Throughput Sequencing of Coding Regions in Alzheimer's Disease using Next Generation Technologies.** *L. Bertram[1,3], M. Kerick[1], B.-M.M. Schjeide[1], B. Thomson[2], M. Werber[1], M.R. Schweiger[1], R. Sudbrak[1], B. Herrmann[1], R.E. Tanzi[3], D. Burgess[2], H. Lehrach[1], B. Timmermann[1]*. 1) Department of Vertebrate Genomics, Max-Planck Institute for Molecular Genetics, Berlin, Germany; 2) Roche NimbleGen, Inc., Madison, WI, USA; 3) Genetics and Aging Research Unit, Massachusetts General Hospital, Charlestown, MA, USA.

Alzheimer's disease (AD) is a genetically complex disease whose pathogenesis is largely influenced by genetic factors. Three decades of intensive research have yielded four established AD genes, and hundreds of potential susceptibility loci, none of which has been unequivocally shown to modify disease risk using conventional methodologies. Systematic meta-analyses performed as part of the AlzGene database (www.alzgene.org), created and maintained by our group, highlight a number of potential AD susceptibility genes with consistent risk effects across all published datasets. However, the genetic effect sizes exerted by these loci are very small and more comprehensive experimental approaches are needed to identify novel AD genes, in particular those which contain rare, potentially disease-causing variants. To this end, we have begun "whole exome" sequencing in probands of 10 multiplex families affected by early-onset AD using microarray-based sequence capture of ~180,000 coding regions followed by massively parallel sequencing. To date, sequence capture and follow-up sequencing has been completed in 6 individuals. In each of these, we detected on average ~26,000 sequence variants, approximately 25% of which were not listed in dbSNP and/or the "1000 Genomes" project (1000GP). Of all detected variants showing minor-allele frequencies <1%, or those not reported at all in either dbSNP and/or 1000GP, nearly 15% represented non-synonymous substitutions, while another ~55% were located in adjacent non-coding sequences or micro-RNAs. An initial bioinformatic assessment of these variants implicated a total of 185 nucleotide-changes across 130 independent loci of possible relevance to the development of AD. This project represents one of the first systematic assessments of the "whole exome" using next-generation technologies in an attempt to identify novel disease-causing variants in a genetically complex disease. It highlights a number of potential AD mutations which warrant molecular genetic follow-up with high priority.

## 24

**Single cell expression profiling of dopaminergic neurons combined with association analysis identifies pyridoxal kinase as a Parkinson disease gene.** *H. Prokisch[1, 2], M. Elstner[1, 3], C. Morris[4], K. Heim[1], T. Gasser[5], C. Schulte[5], S. Goldwurm[7], M. Zeviani[8], D. Burn[9], I. McKeith[9], H.E. Wichmann[10], S. Schreiber[11], P. Chinnery[6], H. Campbell[12], J. Wilson[13], A. Wright[14, 15], T. Klopstock[3], D. Toniolo[16, 17], T. Meitinger[1, 2], D.M. Turnbull[6]*. 1) Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany; 2) Institute of Human Genetics, Technical University Munich, 81675 Munich, Germany; 3) Department of Neurology, Ludwig-Maximilians-University, 81377 Munich, Germany; 4) Medical Toxicology Centre, Wolfson Unit of Clinical Pharmacology, Newcastle University, Newcastle upon Tyne NE2 4AA, United Kingdom; 5) Hertie-Institute for Clinical Brain Research, Section for Neurodegenerative Diseases, University of Tübingen, 72076 Tübingen, Germany; 6) Institute for Ageing and Health, Newcastle University, Newcastle upon Tyne NE4 5PL, United Kingdom; 7) Parkinson Institute, Istituti Clinici di Perfezionamento, 20126 Milan, Italy; 8) O.U. Molecular Neurogenetics, IRCCS Foundation Neurological Institute "C. Besta", 20126 Milan, Italy; 9) Institute for Ageing and Health, Newcastle General Hospital, Newcastle University, Newcastle upon Tyne NE4 6BE, United Kingdom; 10) Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany; 11) Institute of Clinical Molecular Biology, Christian-Albrechts University, 24105 Kiel, Germany; 12) Community Health Sciences and Institute for Genetics and Molecular Medicine, University of Edinburgh Medical School, Edinburgh EH8 9AG, United Kingdom; 13) Centre for Population Health Sciences, University of Edinburgh, Edinburgh EH8 9AG, United Kingdom; 14) MRC Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, United Kingdom; 15) Institute of Genetics and Molecular Medicine, Colon Cancer Genetics Group, University of Edinburgh and MRC Human Genetics Unit, Edinburgh EH4 2XU, United Kingdom; 16) San Raffaele Scientific Institute, Division of Genetics and Cell Biology, Milano 20132, Italy; 17) Istitutio di Genetica Molecolare - Consilgio Nazionale delle Ricerche, 27100 Pavia, Italy.

Objective: The etiology of Parkinson disease (PD) is complex and multifactorial, with both hereditary and environmental factors contributing. Monogenic forms have provided molecular clues to disease mechanisms but genetic modifiers of idiopathic PD are still to be determined. Methods: We carried out whole genome expression profiling of isolated human substantia nigra neurons from patients with PD versus controls followed by association analysis of tagging SNPs in differentially regulated genes. Association was investigated in a German PD sample of 676 cases and 972 controls and confirmed in Italian and British cohorts (cases/controls = 353/704 and 203/1126). Results: We identified four differentially expressed genes located in PD candidate pathways, i.e. MTND2 (mitochondrial, p = 7.14 x 10-7), PDXK (vitamin B6/dopamine metabolism, p = 3.27 x 10-6), SRGAP3 (axon guidance, p = 5.65 x 10-6) and TRAPPC4 (vesicle transport, p = 5.81 x 10-6). Using these candidate genes we identified a DNA variant (rs2010795) in PDXK associated with an increased risk of PD in the German cohort. This association was confirmed in the British and Italian cohorts individually and reached a combined P-value of p = 1.2 x 10-7 (OR = 1.3; CI = 1.18 - 1.44). Interpretation: We provide an example how microgenomic genome-wide expression studies in combination with association analysis can aid the identification of genetic modifiers in neurodegenerative disorders. The detection of a genetic variant in PDXK, together with evidence accumulating from clinical studies, emphasizes the impact of vitamin B6 status and metabolism on disease risk and therapy of PD.

## 25

**The functional consequences of the Parkinson's Disease (PD)-associated SNCA-Rep1.** *O. Chiba-Falek[1], C. Linnertz[1], M.G. Schlossmacher[2], R.L. Nussbaum[3].* 1) IGSP and Dept Med/Neurology, Duke University Medical Center, Durham, NC; 2) OHRI/ Division of Neurosciences, University of Ottawa, Ottawa, Ontario; 3) Institute for Human Genetics/Division of Medical Genetics, University of California, San Francisco, California.

In the post-human genome era, a growing number of disease-associated genetic variants are being reported, increasing the need for a model system to pinpoint causal variants within an associated locus. Studies of familial PD show an effect of $\alpha$-synuclein gene (SNCA) dosage and expression levels in PD pathogenesis. Increasing number of association studies have repeatedly demonstrated that genetic variability across the SNCA locus is associated with an increased risk of developing sporadic PD. Nevertheless, the mechanism/s by which this happens and the identity of the causal variant/s are unclear. Expansion of Rep1, an upstream, polymorphic microsatellite of the SNCA gene, is associated with elevated risk for sporadic PD. Using mice transgenic for a PAC containing various Rep1 alleles in the context of the entire human SNCA locus, we demonstrated the regulatory effect of Rep1. Analysis of the effects of Rep1 alleles in the brains of 72 transgenic mice showed that human SNCA-mRNA and protein levels were increased 1.7- and 1.25-fold, respectively, in homozygotes for the expanded, PD risk-allele compared with homozygotes for the shorter, protective allele. Subsequently we studied the Rep1 effect in 228 healthy human brain samples from three structures differentially vulnerable to PD pathology (substantia-nigra, temporal cortex and frontal cortex). Homozygosity for the protective, Rep1 allele was associated with significantly lower levels of SNCA-mRNA relative to individuals who carried at least one copy of the PD-risk alleles, amounting to an average decrease of ~50% in temporal-cortex and substantia-nigra. In contrast, no difference in SNCA-mRNA level was seen in the frontal-cortex, indicating the brain-region specificity of the Rep1 regulation. Moreover, these results are supported by our published in vitro study using luciferase reporter assay in a neuronal cell culture. We therefore conclude that Rep1 is directly involved in contributing to overall disease susceptibility and, thus, to age-of-onset/progression rate, in a subset of sporadic PD subjects by modulating the expression of the wild-type SNCA gene. A similar approach is underway for studying the regulatory effect of PD-associated variants positioned in the 3'region of the SNCA gene. This novel approach serves as a prototype for future studies that seek to provide a biological mechanism for how the relevant variation within a disease-associated locus affects the risk for a complex disease.

## 26

**Replicating Main Effect and Interactions in Calcium Signaling and Cytoskeleton Regulation Incur Susceptibility to Multiple Sclerosis.** *W. Bush[1], J. Haines[1], P. DeJager[2], J. McCauley[3], M. Ritchie[1], International Multiple Sclerosis Genetics Consortium.* 1) Ctr Human Gen, Vanderbilt Univ, Nashville, TN; 2) Medical & Population Genetics Program Broad Institute, Boston, MA; 3) Miami Institute of Human Genomics University of Miami, Miami, FL.

Multiple sclerosis (MS) is an etiologically complex disorder characterized by demyelination and neurodegeneration in the central nervous system. We present a knowledge-driven multi-locus analysis of MS susceptibility using 931 affected trios from a genome-wide association study. By incorporating biological pathways, previous genetic associations, and other sources into our analysis, we have identified two replicating two-locus models of MS risk and a strongly replicating single-locus effect within the calcium signaling and cytoskeleton regulatory pathways. Using a case-control analysis of the 931 affected probands from the IMSGC GWAS study and 2,950 population controls, we identified 5,463 models with significant model fit statistics (mf) and a significant interaction term (it) by likelihood ratio test. Of these models, two two-locus models (PLC$\beta$4-PLC$\beta$1 and CHRM3-MYLK) replicate with both significant model fit and interaction term in an additional sample of 808 MS cases and 1720 controls. One interaction model from the GWAS screen replicated as a main effect in this sample (SCIN), and also replicated as a main effect in a third replication set of 2,330 MS cases and 2,110 controls, also ascertained from Brigham and Women's Hospital. PLC$\beta$4 and PLC$\beta$1 mediate signal transduction in the calcium signaling pathway (KEGG:ko00562), and SNPs from these two genes were in significant models in the IMSGC/WTCCC screen (mf $p = 0.0008$, it $p = 0.0009$) and in the first BWH replication set (mf $p = 0.0475$, it $p = 0.0095$). CHRM3 is a G-protein coupled receptor that can trigger calcium signaling, activating PLC$\beta$1 and PLC$\beta$4. Calcium signaling can activate MYLK, a kinase that triggers cytoskeletal rearrangement. SNPs from these two genes were in significant models in the IMSGC/WTCCC screen (mf $p = 0.0006$, it $p = 0.0002$) and in the first BWH replication set (mf $p = 0.0235$, it $p = 0.0026$). SCIN is a calcium-dependent actin severing protein that presumably plays a regulatory role in exocytosis and neurotransmitter release. A SNP from SCIN is in a significant interaction model with CYFIP in the IMSGC/WTCCC screen (mf $p = 0.0001$, it $p = 0.0001$) and appears as a main effect in both BWH replication sets ($p = 0.0003$ and OR = 1.243, $p = 0.0073$ and OR = 1.120). These results illustrate the utility of knowledge-based interaction analysis for the discovery of main effects and interactions, and present a possible neurodegenerative mechanism for MS involving calcium homeostasis.

## 27

**Functional annotation of GWAS hits.** *J. Knight[1], M.R. Barnes[2], G. Breen[3], M.E. Weale[1].* 1) Department of Medical and Molecular Genetics, King's College London School of Medicine, London, London, United Kingdom; 2) GlaxoSmithKline Research and Development, Harlow, UK; 3) MRC Social, Genetic and Developmental Psychiatric Centre, Institute of Psychiatry, King's College London, London, United Kingdom.

Genome wide association studies (GWAS) have provided new clues about the aetiology of complex genetic diseases. In addition to a small set of highly significant results, these studies have produced thousands of moderately suggestive results, requiring new systematic approaches to prioritise their follow up. A number of databases and weighting schemes have been developed to allow researchers to use annotation information, but none are empirically based. In order to calibrate a prioritization scheme, we have compared the functional annotations of a comprehensive list of GWAS hits against random expectation. We have explored three important annotation categories: non-synonomous SNPs, promoter SNPs and eQTLs. We investigated annotated SNPs plus their linkage disequilibrium proxies (i.e. SNPs with an r2 >0.8 in the HapMap CEU dataset). We found an increased proportion of annotation in the GWAS hits compared to the random SNPs in all three categories; as shown below for one of our analysis.

| | eQTL (%) | Non-synonomous (%) | Promoter (%) |
|---|---|---|---|
| GWAS hits | 13.7 | 7.9 | 3.6 |
| Random | 7.7 | 2.7 | 1.5 |

This pattern remained even when the SNPs were stratified by minor allele frequency; or when the extended MHC region on chromosome 6 was excluded; or when SNPs belonging to more than one annotation category were removed. Our study demonstrates that GWAS hits are enriched for these three functional categories, and hence it would be appropriate to provide a higher weighting for such SNPs when planning follow up studies for GWAS.

## 28

**SNPs reproducibly associated with complex human traits are disproportionately likely to predict gene transcript levels.** *N.J. Cox[1], E. Gamazon[1], W. Zhang[2], M.E. Dolan[2], A. Konkashbaev[1], D.L. Nicolae[1].* 1) Sec Genetic Medicine, Univ Chicago, Chicago, IL; 2) Section of Hem/Onc, Univ Chicago, Chicago, IL.

The SCAN database (SNP and Copy number variant ANotation; http://www.scandb.org) includes in addition to physical, LD, and publicly available functional annotation, results of GWAS for expression transcript levels measured using the Affymetrix Exon 1.0 ST Arrays on RNA obtained from lymphoblastoid cell lines from CEU and YRI HapMap trios (see, for example, Duan et al., Am J Hum Genet 82:1101-13). We annotated all unique SNPs reproducibly associated with complex human traits included in the NHGRI database summarizing results of all GWAS published to date (Hindorff LA, Junkins HA, Mehta JP and Manolio TA. A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/26525384) with information on whether the SNP was a significant predictor of expression transcript levels for one or more genes. We then used a permutation approach to determine whether SNPs reproducibly associated with complex traits are disproportionately likely to be eQTLs; from the set of SNPs included on all high throughput platforms, we randomly chose SNPs from the same minor allele frequency bins as the associated SNPs and determined the proportion of those SNPs characterized in SCAN as eQTLs. Results of these studies suggest that SNPs reproducibly associated with complex traits are significantly ($p<10^{-5}$) more likely to be eQTLs than expected for the observed MAF distribution of these SNPs, and further, that these associated SNPs more likely to be master regulators, defined as eQTLs that predict transcript levels of 10 more genes. These findings are robust across the entire range of human phenotypes included in the NHGRI database (neurological, psychiatric, cancer, other), for a range of thresholds for establishing expression associated SNPs ($10^{-4}$ to $10^{-8}$), and a range of thresholds for defining master regulators (10 - 100). These findings provide substantial support for the utility of LCL transcriptome studies, and highlight the utility of the SCAN db and the tools we have built to enable users to access the results of our studies on the LCL transcriptome.

## 29

**Expression quantitative trait loci (eQTL) analysis of subcutaneous human adipose tissue from Body Mass Index (BMI)-discordant sib-pairs.** *A. Walley[1], M. Falchi[1], J. Andersson[1], D. Meyre[2], C. Lecoeur[2], P. Jacobson[3], L. Sjöstrom[3], L. Carlsson[3], P. Froguel[1,2].* 1) Section Genomic Medicine, Imperial College London, London, United Kingdom; 2) CNRS 8090-Institute of Biology, Pasteur Institute, Lille, France; 3) Department of Molecular and Clinical Medicine and Center for Cardiovascular and Metabolic Research, The Sahlgrenska Academy, Gothenburg, Sweden.

**Background:** Genomewide association studies in obese subjects have recently identified many new genes but these are purely statistical correlations with the phenotype. By simultaneously mapping genetic variation and gene expression in a tissue of interest, genomic loci associated with gene expression can be identified, i.e. expression quantitative trait loci (eQTL). This combined analysis can provide substantial additional information about the genetic architecture of common non-syndromic obesity. **Subjects & Methods:** The SibPair cohort comprises 154 nuclear families (732 subjects) from Sweden, each containing an obesity-discordant sib pair (at least 10 kg/m2 difference in BMI). Gene expression and genetic variation were analysed in 359 siblings from the SibPair cohort, i.e. proband, the most discordant sibling and additional intermediate BMI siblings. Subcutaneous adipose tissue was obtained via punch biopsy. Gene expression was profiled using the Affymetrix Human U133Plus2.0 array. Genetic variation using the ABI SNPlex human linkage mapping set v4.0 (3922 markers) was analysed using genomic DNA prepared from whole blood samples. eQTL analysis was performed using RMA-normalised data for the transcripts and all SNP markers that passed QC criteria. Fine mapping of positive loci is being carried out using Illumina Infinium 610K SNP arrays. **Results:** A total of 27409 transcripts and 3877 SNP markers passed QC and were subject to analysis. A total of 353 transcripts were differentially expressed (fold-change>1.5) between obese and lean subjects, 328 upregulated and 25 downregulated. Analysis of all transcripts identified 578 and 1107 cis-eQTLs at FDR values of 5% and 10% respectively. A total of 53 individual trans-eQTLs were identified at an FDR of 10% (LOD score >8). Analysis of clustering of the peaks of linkage of all trans-eQTLs demonstrated statistically significant clustering on chromosomes 1 and 6. By comparison with rat data on adipose tissue expression, at an FDR of 10%, 16 human eQTLs corresponded to rat eQTLs. **Discussion:** This study provides the first eQTL analysis of human subcutaneous adipose tissue from BMI-discordant sib-pairs. Validation of the top cis-eQTLs identified in our dataset is being carried out as a subset of an ongoing GWAS analysis of these families.

## 30

**Use of genome-wide expression data to mine the "gray zone" of GWA studies leads to novel candidate obesity genes.** *J. Naukkarinen[1,2], I. Surakka[1], K.H. Pietiläinen[3,4], A. Rissanen[3], V. Salomaa[5], S. Ripatti[1], J. Kaprio[1,4,6], M-R. Taskinen[7], L. Peltonen[1,2,8,9].* 1) FIMM, Institute for Molecular Medicine Finland, Helsinki, Finland; 2) Department of Medical Genetics, University of Helsinki, Finland; 3) Obesity Research Unit, Department of Psychiatry, Helsinki University Central Hospital, Helsinki, Finland; 4) Finnish Twin Cohort Study, Department of Public Health, University of Helsinki, Finland; 5) Division of Welfare and Health Promotion, National Institute for Health and Welfare, Helsinki, Finland; 6) Division of Mental Health and Substance abuse services, National Institute for Health and Welfare, Helsinki, Finland; 7) Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland; 8) Broad Institute, MIT, Cambridge, Massachusetts, U.S.A; 9) The Wellcome Trust Sanger Institute, Cambridge, UK.

To get beyond the "low-hanging fruits" so far identified by genome-wide association (GWA) studies, new methods must be developed in order to discover the numerous remaining genes that estimates of heritability indicate should be contributing to complex human phenotypes, such as obesity. Here we describe a novel integrative method for complex disease gene identification utilizing both genome-wide transcript profiling of adipose tissue samples and consequent analysis of genome-wide association data generated by GWA:s. The difficulty in evaluating the role of differentially expressed transcripts lies in sorting out the mRNAs with causal relevance from those that merely represent reactive response to the disease state. We sought to infer causality of genes with obesity by employing a unique set of monozygotic (MZ)twins discordant for BMI representing the 5% of pairs most extremely discordant for BMI (n=13 pairs, age 24-28 years, 15.4 kg mean weight difference)and then contrasted their transcript profiles with those from a larger sample of non-related individuals (N=77).

Using a liberal threshold for significance (pair-wise t-test $P<0.05$) we identified 2,674 transcripts differentially regulated in the adipose tissue of the obese vs. lean twins. Next, in a data set consisting of adipose tissue samples from 77 unrelated individuals, using stringent criteria (Pearson correlation $P<10^{-4}$) we identified 84 transcripts correlated with BMI. Of these, 56 were shared with those identified in MZ pairs and thus we designated them as "reactive". The remaining 28 transcripts encoded 27 unique genes that, while strongly correlated with BMI, were not differentially regulated in the discordant MZ pairs and as such could represent candidates for being related to causal processes.

Testing for association of SNP variants in these 27 genes in the population samples of the large ENGAGE consortium (N=21,000) GWA dataset revealed a significant deviation of P-values from the expected ($X^2$ $P= 4 \times 10^{-4}$), while SNP variants in the genes classed "reactive" failed to show any deviation from the expected. Altogether 13 putatively causative genes harbored SNPs associating with BMI ($P<0.05$). Top hits were SNP variants in the coagulation factor *F13A1*, identifying it as a novel obesity gene; this was also replicated in a second GWA set of ~2000 individuals ($P=0.0004$ in the combined analysis).

## 31

**Elucidating Networks of eSNPs Associated with Type 2 Diabetes.** *H. Zhong, J. Beaulaurier, C. Molony, E. Schadt.* Department of Genetics, Rosetta Inpharmatics, seattle, WA.

Genome-wide association studies (GWAS) have demonstrated the ability to identify the strongest causal common variants in complex human diseases. However, to date, the massive data generated from GWAS have not been maximally leveraged and integrated with other types of data to identify true associations that fail to meet the stringent level of association required to achieve genome-wide significance. Genetics of gene expression (GGE) studies have shown promise towards identifying DNA variations associated with disease and providing a path to functionally characterize findings from GWAS. Here, we assemble a comprehensive set of SNPs associated with expression (eSNPs) identified in two GGE study cohorts representing liver, subcutaneous fat, and omental fat tissues to systematically characterize whether eSNPs are enriched for SNPs that associate with type 2 diabetes (T2D). We demonstrate that liver and adipose eSNPs are significantly more likely to associate with T2D in three large-scale GWAS than randomly selected SNPs. This enrichment for T2D association increases as we restrict to eSNPs that correspond to genes comprising gene networks constructed from adipose gene expression data isolated from a mouse population segregating a T2D phenotype. Finally, by restricting to eSNPs corresponding to genes comprising an adipose subnetwork strongly predicted as causal for T2D, we further increased the enrichment for SNPs associated with T2D and were able to identify a functionally related set of diabetes susceptibility genes. We identified and validated malic enzyme 1 (Me1) as a key regulator of this T2D subnetwork in both mouse and human cohorts. This is the first empiric study demonstrating that causal networks for diabetes constructed from experimental cross populations segregating T2D are enriched for genes that associate with T2d in human GWAS. We demonstrate an increased enrichment of T2D SNPs that in turn enhance our ability to assess T2D risk in human populations. This integration of eSNPs and networks provides a novel approach for extracting value from the wealth of data currently being generated by GWAS. The success of our approach emphasizes that large-scale molecular profiling data (i.e., functional data) in human and mouse populations segregating disease phenotypes provide increased power to connect genetic variations to susceptibility genes and the networks that drive disease phenotypes.

## 32

**Gene networks and microRNAs implicated in aggressive prostate cancer.** *L. Wang[1], H. Tang[2], A.L. Oberg[2], J.M. Cunningham[1], J.R. Cerhan[2], S.N. Thibodeau[1].* 1) Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 2) Health Sciences Research, Mayo Clinic, Rochester, MN.

Prostate cancer (PC) is believed to be caused by complex interactions between genes and environment factors. Genomewide association studies have identified over a dozen genetic loci that are associated with increased risk to prostate cancer, but none have been specifically linked to aggressive prostate cancer, which is the most clinically relevant form of the disease. A systems biology approach that organizes gene expression data into function-related modules has been successfully applied to multiple studies to identify co-expressed gene networks and possible candidate genes responsible for various complex traits. To identify genes involved in molecular pathways in aggressive PC, we evaluated transcriptional profiles from 13935 mRNAs and 366 miRNAs using lymphoblastoid cell lines from PC patients with different aggressive phenotypes (Gleason grade $\geq 8$ as aggressive and $\leq 5$ as non-aggressive phenotypes). We implemented an integrated analysis by examining co-expressed gene network and incorporating miRNA, a key regulator of gene expression, into the systems approach. Using a weighted gene co-expression network analysis algorithm, we identified four distinct co-expressed gene modules, one of which demonstrated significant association with an aggressive phenotype of PC. The module of interest contained a total of 266 genes and was characterized by over-representation of cell cycle-related genes. From this module, we further defined 20 hub genes that were highly connected to each other and to other non-hub genes. Most of these hub genes have been widely studied and show significant involvement in a variety of tumors. miRNA profiling analysis revealed 8 miRNAs whose expression levels were significantly associated with the aggressive phenotype of PC (false discovery rate $< 0.01$). Interestingly, four of the 8 miRNAs have been implicated in direct regulation of cell cycle and two are predicted to target three of the 20 hub genes. Our results suggest that cell cycle is likely to be a causal pathway whose mis-regulation is associated with an aggressive phenotype of PC. Through directly regulating cell cycle-related genes, some miRNAs may also influence PC aggressiveness. Further characterization of cell cycle-related genes (particularly, the hub genes) and miRNAs that regulate these hub genes could facilitate identification and characterization of candidate genes responsible for aggressive PC and lead to better understanding of PC etiology and progression.

## 33

**A genome-wide map of open chromatin in human pancreatic islets.**

*K.J. Gaulton[1], T. Nammo[2], L. Pasquali[2], J.M. Simon[3], P.G. Giresi[3], P. Mieczkowski[1], A. Secchi[6], T. Berny[7], D. Bosco[7], E. Montanya[8], K.L. Mohlke[1,5], J.D. Lieb[3,4,5], J. Ferrer[2].* 1) Department of Genetics, University of North Carolina, Chapel Hill, NC; 2) Hospital Clinic de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer, CIBERDEM, Barcelona, Spain; 3) Department of Biology, University of North Carolina, Chapel Hill, NC; 4) Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC; 5) Carolina Center for the Genome Sciences, University of North Carolina, Chapel Hill, NC; 6) San Raffaele Scientific Institute, Milan, Italy; 7) University of Geneva, Geneva, Switzerland; 8) Hospital de Bellvitge, CIBERDEM, Barcelona, Spain.

Common human genetic variants that occur in regulatory DNA may contribute to disease risk at many susceptibility loci. To detect the most likely regulatory variants at genomic intervals associated with type 2 diabetes (T2D) and to better characterize gene regulation in pancreatic islets, we sought to comprehensively identify regulatory sequences active in these cells. We used nucleosome loss, measured by FAIRE-seq (Formaldehyde-Assisted Isolation of Regulatory Elements with high-throughput sequencing) as a molecular tag to identify active regulatory DNA genome-wide. We identified ~80,000 discrete open chromatin regions. These regions were preferentially located proximal to known promoters, overlapped multi-species conserved regions and other marks indicative of regulatory function, and correlated with actively transcribed genes in islets. Regions of open chromatin detected in islets but not in five non-islet cell lines were predominantly located distal to known promoters and contained an over-representation of islet transcription factor binding sites. Throughout the genome, nucleosomes were coordinately evicted from local clusters of putative regulatory elements, which we term COREs (Clusters of Open Regulatory Elements). We identified islet-specific COREs and show that these were frequently flanked by CTCF-bound insulator elements, and were largely linked to single genes transcribed in islets. Strong regions of open chromatin were detected at previously described T2D susceptibility loci. We found that 10% of T2D-associated SNPs (HapMap CEU $r^2$>.8 with reference SNP) were contained within open chromatin regions; these variants are strong candidates to test for a role in islet transcriptional regulation. We identified a T2D-associated SNP with an allelic imbalance in FAIRE signal, suggesting a functional difference in chromatin openness in islets between alleles and highlighting the potential utility of FAIRE as a tool to identify functional SNPs. Our data demonstrate that FAIRE-seq can rapidly identify genome-wide regulatory elements from a primary human tissue and provide guidance for studies seeking regulatory SNPs underlying human disease.

## 34

**Importance of sequencing rare variants after a genome-wide association study (GWAS): the *MC1R* gene, 16q24 region and melanoma story.**

*F. Demenais[1,2], E. Corda[1,2], J. Barrett[3], M. Iles[3], E.M. Gillanders[4], A.M. Goldstein[5], P.A. Kanetsky[6], E. Bakker[7], D.T. Bishop[3], J.A. Newton-Bishop[3], N.A. Gruis[8], Melanoma Genetics Consortium (GenoMEL).* 1) INSERM Unit&eacute 946, Paris, France; 2) Fondation Jean-Dausset-CEPH, Paris, France; 3) Section of Epidemiology and Biostatistics, Leeds Institute of Molecular Medicine, Leeds, UK; 4) Inherited Disease Research Br., National Human Genome Research Institute, NIH, Baltimore, MD, USA; 5) Genetic Epidemiology Br., DCEG, National Cancer Institute, NIH, Bethesda, MD, USA; 6) Dept of Biostatistics and Epidemiology, University of Pennsylvania, USA; 7) Dept. of Clinical Genetics, Leiden University Medical Centre, Leiden, The Netherlands; 8) Dept. of Dermatology, Leiden University Medical Centre, Leiden, The Netherlands.

A GWAS of melanoma, performed by the GenoMEL consortium, identified the strongest association with the 16q24 region. This region includes candidate genes: *CDK10* (involved in cell-cycle regulation), *FANCA* (regulating genomic stability) and *MC1R* (previously associated with pigmentation phenotypes and melanoma). Three 16q24 SNPs had significant independent effects, rs258322 (*CDK10*), rs4785763 (*AFG3L1*) and rs8059973 (*DBNDD1*). However, none of the non-synonymous (NS) *MC1R* variants was present on the GWAS chips. To investigate whether the association signals in the 16q24 region might be accounted for by *MC1R* variants, this gene was sequenced in 1,805 GenoMEL GWAS subjects (918 cases, 887 controls). We first conducted univariate population-adjusted logistic regression with each NS *MC1R* variant. We then compared the strength of association when examining each SNP with and without each *MC1R* variant in the model and, conversely, when examining each *MC1R* variant with and without each SNP. This was followed by stepwise multiple regression and haplotype analysis (THESIAS program) with all significant SNPs and *MC1R* variants. A total of 75 *MC1R* variants were characterized, of which 9 NS variants had allele frequency between 1% and 13%. Univariate analysis showed significant effects of 3 *MC1R* variants, R151C, R160W and D294H (P ranging from $1.5x10^{-3}$ to $1.8x10^{-11}$). There was no longer evidence for association of melanoma with rs258322 after entering R151C in the model and decreased evidence for association with rs4785763 after entering either R151C or R160W (P decreased by a factor of $10^2$ to $10^4$) while the association with rs8059973 was barely modified. Conversely, the association with R151C was decreased in the presence of either rs258322 or rs4785763 and with R160W in the presence of rs4785763 while the association with D294H was unchanged. Stepwise regression showed independent effects of R151C, R160W and D294H (P ranging from $4.6x10^{-5}$ to $3.1x10^{-13}$) with a marginal effect of rs8059973 (P=0.06). Haplotype analysis confirmed these findings and demonstrated that the rs258322 signal was accounted for by R151C and the rs4785763 signal by both R151C and R160W. Any haplotype carrying rs8059973, which was never on the same haplotype as a significant *MC1R* variant, did not show a significant effect (P>0.20). This study clearly shows that ignoring rare variants can lead to incorrect inferences on the potential role of candidate genes carrying common SNPs identified by GWAS.

## 35

**Finemapping, Interaction Analyis and Rare Variant Patient Resequencing Results in novel QT interval loci identified by the QTSCD consortium.** *A. Pfeufer[1], D.E. Arking[2], S. Sanna[3], M. Müller[4], H. Prucha[1], M. Vieracker[1], S. Perz[5], M.F. Sinner[6], H. Prokisch[1], D. Mehta[1], T.W. Mühleisen[7], A.A. Hicks[8], P.P. Pramstaller[8], H.E. Wichmann[4], E. Boerwinkle[9], G. Tomaselli[10], S. Kääb[6], G. Abecasis[11], T. Meitinger[1], A. Chakravarti[2], ARIC, KORA, SardiNIA, GenNOVA and Heinz Nixdorf Recall Study Groups.* 1) Institute of Human Genetics, Technical University Munich and Helmholtz Center Munich, Germany; 2) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD, U.S.A; 3) Istituto di Neurogenetica e Neurofarmacologia, CNR, Monserrato, 09042 Cagliari, Italy; 4) Institute of Epidemiology, Helmholtz Center Munich, Germany; 5) Institute of Medical Informatics, Helmholtz Center Munich, Germany; 6) Department of Medicine I, Klinikum Grosshadern, Munich, Germany; 7) Institute of Human Genetics and Life & Brain Center, University of Bonn, Bonn, Germany; 8) Institute of Genetic Medicine, EURAC European Academy, Bolzano, Italy; 9) Genetics Center, University of Texas Health Science Center, Houston, Texas, U.S.A; 10) Department of Medicine, Johns Hopkins University, Baltimore, MD, U.S.A; 11) Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A.

The QT interval, a measure of cardiac repolarization, predisposes to ventricular tachycardia and sudden cardiac death (SCD) when prolonged or shortened (1). By genome-wide association we have identified ten genomic regions modifying QT interval in European population samples, four of which were not known from monogenic or functional studies: the NOS1AP (CAPON)-OLFML2B (1q23.3) region (2) and the RNF207 (1p36.31), LITAF (16p13.13) and NDRG4-GINS3-SETD6-CNOT1 (16q21) (3). For a more detailed characterization of these loci and to identify causal genes we have analyzed SNPxSNP interactions, performed eQTL analysis based on WBC expression data in n=302 KORA individuals and resequenced exons from regional genes in patients with monogenic Long-QT syndrome (LQTS). No significant interactions between any of the 10 QTL-SNPs identified were observed. eQTL analysis resulted in the identification of an eQTL for NOS1AP mRNA within ATP1A1 (p=1.41E-11). ATP1A1 encodes the $\alpha$-subunit of the sarcolemmal Na/K-ATPase, the $\beta$-subunit of which (ATP1B1) has been identified as a QTL for QT interval itself (3). In addition we found a significant overrepresentation of mutations within OLFML2B in patients with monogenic LQTS in comparison to population controls. These results provide new insights into myocardial electrophysiology and provide novel candidate genes for ventricular arrhythmias and SCD especially by pointing towards a mechanistic link between NOS1AP and the maintanance of the transmembrane electrochemical Na/K gradient and by implying OLFML2B in addition to NOS1AP as a contributor to the QT modifying effects in the 1q23.1 region. ---//--- 1. Tomaselli GF, Beuckelmann DJ, Calkins HG, et al. Sudden cardiac death in heart failure. The role of abnormal repolarization. Circulation 90, 2534-2539 (1994). --- 2. Arking DE, Pfeufer A, Post W et al. A common genetic variant in the NOS1 regulator NOS1AP modulates cardiac repolarization. Nature Genetics 38, 244-251 (2006). --- 3. Pfeufer A, Sanna S, Arking DE, et al. Common variants at ten loci modulate the QT interval duration in the QTSCD Study. Nat Genet . 41, 407-414 (2009).

## 36

**Non-Y402H SNPs and copy number at the CFH-RCA locus are associated with AMD.** *T. A. Sivakumaran[1], R. Igo, Jr.[1], J.M. Kidd[4], A. Itsara[4], L.J. Kopplin[2], W. Chen[5], S.A. Hagstrom[6,7], N.S. Peachey[6,7,8], P.J. Francis[9], M.L. Klein[9], E.Y. Chew[10], V.L. Ramprasad[11], G. Kumaramanickavel[11], A. Swaroop[12], G.R. Abecasis[5], R. Klein[13], B.E.K. Klein[13], D.A. Nickerson[4], E.E. Eichler[4], S.K. Iyengar[1,2,3].* 1) Department of Epidemiology & Biostatistics, Case Western Reserve Univ, Cleveland, OH; 2) Department of Genetics, Case Western Reserve University, Cleveland, OH; 3) Department of Opthalmology, Case Western Reserve Univesity, Cleveland, OH; 4) Department of Genome Sciences, University of Washington, Seattle, WA; 5) Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI; 6) Cole Eye Institute, Cleveland Clinical Foundation, Cleveland, OH; 7) Department of Ophthalmology, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, OH; 8) Research Service, Louis Stokes Cleveland Department of Veterans Affairs Medical Center, Cleveland, OH; 9) Casey Eye Institute, Oregon Health & Science University, Portland, OR; 10) Division of Epidemiology and Clinical Applications, National Eye Institute, Bethesda, MD; 11) Sankara Nethralaya, Chennai, India; 12) Neurobiology Neurodegeneration and Repair Laboratory, National Eye Institute, Bethesda, MD; 13) Department of Ophthalmology and Visual Sciences, University of Wisconsin School, of Medicine and Public Health, Madison, WI.

Age-related macular degeneration (AMD) is the leading cause of visual dysfunction and blindness in developed countries, and a rising cause in underdeveloped countries. In the recent past, many genes involved in the complement pathway have been implicated in disease pathogenesis. A nonsynonymous SNP, Y402H, in complement factor H (CFH) is more strongly associated with AMD, whereas a common deletion involving nearby genes CFHR3 and CFHR1 protects against AMD. However, several variants in CFH are more significantly associated with AMD than Y402H. To refine the role of causal variants, we systematically narrowed down the critical region at the CFH-RCA locus using SNPs and CNV analysis. Analysis of large case-control sets (N = 4498 cases and N = 3341 controls) revealed one risk (with shared core SNPs), three closely-related protective and two neutral haplotypes. The risk-bearing haplotype (H1) with risk allele C at codon 402 and no common deletions conferred the greatest risk by increasing the odds of AMD by 2.22 fold (95% CI = [1.97, 2.5]; p = $1.0 \times 10^{-39}$). Our results show that two critical regions spanning ~60kb in CFH play an important role in AMD susceptibility. Trans-ethnic mapping using data from the Indian population showed a haplotype (H47), lacking the risk allele at codon 402, increased the odds of AMD by 2.58 fold (p = 0.012). Further analysis revealed a potential causal variant, segregating with the risk haplotype. Bioinformatic analyses show that the variant affects transcriptional binding sites. Preliminary studies using the gene expression and genotype data from HapMap samples indicate that variation in gene expression is significantly ($p < 0.05$) associated with the variant. Additional studies are underway to further confirm the role of this variant in pathogenesis of AMD.

## 37

**Population genetics and genomics of human gene expression.** *B.E. Stranger[1], S.B. Montgomery[2,3], A.S. Dimas[2,3], C.E. Ingle[3], M. Sekowska[3], C. Beazley[3], P. Deloukas[3], E.T. Dermitzakis[3].* 1) Division of Genetics, Harvard Medical School, Boston, MA; 2) Division of Medical Genetics, University of Geneva Medical School, Geneva Switzerland; 3) The Wellcome Trust Sanger Institute, Hinxton, UK.

Gene expression is a heritable, quantitative phenotype that can be used to assess the impact at the cellular level of naturally-occurring functional genetic variation. It is increasingly used to assist in the interpretation of disease association signals. We previously described a comprehensive analysis of the genetic basis of gene expression variation in the original 270 individuals of the HapMap project using both SNPs and CNVs (Stranger et al. Science 2007 and Nat Genet. 2007). We have now extended this analysis to 830 unrelated individuals from 8 global populations of the extended HapMap samples. We interrogated whole-genome gene expression in lymphoblastoid cell lines using array-based technolgies, and performed association analyses with these expression data and 1.6M SNP genotypes from the HapMap3 Consortium, with the goal of identifying and characterizing cis eQTLs in these populations. The increased power from analyzing nearly double the sample size of each of the initial four HapMap populations allows us to detect approximately 2.5 times the number of previously-reported genes with SNP-expression eQTLs in each of these populations, corresponding to a cis-eQTL for approximately 16% of expressed genes from samples of around 100 unrelated individuals. Together with the additional populations, in total we detect cis-eQTLs for approximately 70% of expressed genes, with 56% of detected eQTL observed in at least 2 populations. We will present analyses of population differentiation at the gene expression level, describing those categories of genes with high differentiation, and estimating the contribution of true genetic effects and environmental factors. We will also describe the effect of natural selection on variants that influence gene expression using haplotype-based methodologies designed to detect recent selection. We believe this is the largest analysis of gene expression in population samples from around the world to date.

## 38

**1st Generation Genome-Wide Map Of *Cis*-Regulatory SNPs In Human Cells.** *T. Pastinen[1,2], T. Kwan[1,2], B. Ge[2], D.J. Verlaan[1,2], V. Koka[1], M. Ouimet[3], V. Gagné[3], N. N'Diaye[3], K. Lam[1], E. Grundberg[1,2], A. Montpetit[2], E. Harmsen[2], D.K. Pokholok[4], K.L. Gunderson[4], M. Blanchette[5], D. Sinnett[3].* 1) Department of Human Genetics, McGill University, Montréal, Québec, Canada; 2) McGill University and Genome Québec Innovation Centre, Montréal, Québec, Canada; 3) Hôpital Ste-Justine, Université de Montréal, Montréal, Québec, Canada; 4) Illumina Inc, San Diego, CA; 5) School of Computer Science, McGill University, Montréal, Québec, Canada.

The expanding lists of functional elements in the genome and disease-associated SNPs call for hypothesis-free assessment of regulatory variation. The *cis*-acting components of expression variation can be identified through mapping allelic expression (AE) differences. A map of AE-associated SNPs in HapMap CEU lymphoblasts (LCLs) using genome-wide measurements of AE on Illumina® Human 1M BeadChips shows that common SNPs are associated with AE in 30% of measured RefSeq genes. A large fraction of these impact full-length primary transcripts and, on average, explain 60% of population variance in transcription of associated genes. The large effect sizes observed by this method provide a straightforward approach for pinpointing causal *cis*-regulatory SNPs (*cis*-rSNPs). A map of AE-associated SNPs in YRI LCLs was recently generated, identifying ~500 primary transcripts showing regions of shared AE association in both populations, containing potential *cis*-rSNPs altering transcription rates. The CEU/YRI AE mapping data was intersected with the 1000 Genomes pilot data, yielding 6497 candidate *cis*-rSNPs altering regulation of 483 full-length transcripts. Candidate sites are defined by the shared AE-associated intervals and heterozygosity in samples showing differential AE. Precisely mapped *cis*-rSNPs from 1000 Genomes and AE-association intersection (i.e. cases with <=3 candidate sites) show highly significant enrichment with regulatory and ENCODE ChIP-seq functional data over control sites (heterozygous SNPs from same regions in samples not showing AE): e.g. 12.3x, 6.7x, 4.9x, 4.6x, and 3.2x enrichment of junD, CpG islands, c-Fos, Max, and RNA pol2 signals, respectively, is observed. Intersecting the AE, 1000 Genomes, and ENCODE functional datasets enables the generation of new hypotheses on gene regulation for functional testing and follow-up. Current fine-mapping experiments are ongoing, along with in vitro characterization of these candidate sites for validating allele-specific effects on promoter/enhancer activity, with >20 causal *cis*-rSNPs isolated to date. These results are anticipated to yield a genome-wide collection of common *cis*-rSNPs based on in vivo differences in transcriptional control. This first catalog of causal regulatory variation in the human genome will pave the way for large-scale annotation of naturally occurring *cis*-regulatory variation, which will allow for the direct interrogation of their role in human phenotypic variance.

## 39

**Results from HapMap 3: Rare and common variation in diverse human populations.** *S. Schaffner, The HapMap 3 Consortium.* The Broad Institute of MIT and Harvard, Cambridge, MA.

Patterns of human genetic variation have already been mapped and analyzed for millions of individual DNA markers. This has greatly advanced our understanding of population structure and genome history, and has led to the discovery of risk loci for many diseases. Most variants have not yet been studied, however, and most hereditary risk remains unexplained. There is a growing appreciation that rare alleles and structural variants make major contributions to phenotypic effects. To bridge the gap between our current knowledge and the deep understanding we need, we have extended the large-scale study of diversity toward rarer alleles, toward inclusion of copy-number variants (CNVs) and into a wider range of populations. We exhaustively genotyped 1.6M SNPs and 850 common copy-number variant (CNV) sites in 1,300 individuals from 11 global populations, and completely resequenced ten 100-kb regions chosen from ENCODE regions. We find that for rare variants, SNP discovery in one population varies greatly in how well it captures variation in another population, and that FST is of limited value as a predictor. The expanded resource reported here improves imputation of rare SNPs, with 25% of them showing a large improvement compared to HapMap 2. For new SNPs discovered in sequencing, we find that imputation works well ($r^2$ between true and imputed genotype > 0.7) as long as at least two copies of the minor allele are included in the reference panel, and even has considerable success for a single reference copy ($r^2$ = ~0.65). Imputation of biallelic CNVs has a similar success rate as for frequency-matched rare SNPs. Finally, we confirm signals of recent natural selection from past studies and report 60 new selection signals in East African samples. These data greatly extend our knowledge of the relationship between different classes of variation, in different populations and highlight the need for deep interrogation of human variation from the perspective of density of nature of markers as well as the diversity of population sampled.

## 40

***Cis*-regulatory variants exert stable effects upon environmental perturbation in human primary cells.** *E. Grundberg[1,2], T. Kwan[1,2], V. Adoue[1,2], B. Ge[2], K.C.L. Lam[2], V. Koka[2], A. Kindmark[3], H. Mallmin[4], O. Ljunggren[3], O. Nilsson[4], T. Pastinen[1,2,5].* 1) Department of Human Genetics , McGill University, Montreal , Quebec, Canada; 2) McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada; 3) Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 4) Department of Surgical Sciences, Uppsala University, Uppsala, Sweden; 5) Department of Medical Genetics , McGill University , Montreal , Quebec, Canada.

Genetic variants altering *cis*-regulation of normal gene expression (*cis*-eQTLs) have been extensively mapped in human cells and tissues, but *cis*-eQTLs underlying differences in induced gene expression in primary cells have not been studied to date. We carried out large-scale induction experiments using primary osteoblasts derived from ~100 unrelated Swedish donors cultured under 18 different conditions (BMP-2, dexamethasone, 1.25VidD3, IGF-1, PGE2, PTH, TNF-$\alpha$, untreated control-1 and untreated control-2 at t=2h and t=24h). The treatments with the most drastic effect on gene expression, as assessed by Affymetrix HG-U133 arrays, were selected and included BMP-2 (t=2h, N=96), dexamethasone (t=24h, N=101), PGE2 (t=24h, N=100) and untreated control-1 (t=24h, N=94). We performed genome-wide expression profiling (Illumina HumanRef8) in biological replicates ($n_{total}$=782) and genotyping (Illumina HumanHap550) in order to map *cis*-eQTLs underlying differences in induced gene expression. We verified robust responses upon treatment by principal component analysis showing distinct clusters corresponding to the different treatments. Co-expressed genes from the different clusters were enriched in well-known treatment-specific pathways, e.g. IGF-1 and glucocorticoid receptor signaling pathways were up-regulated upon dexamethasone stimulation. Next we assessed the proportion of shared *cis*-eQTLs (defined as ±250kb window flanking the gene) across treatments and found that 98% of *cis*-eQTLs at P<3.5E-8 were replicated in at least one additional treatment at P<0.0005. In fact, on average only 0.3% of the *cis*-eQTLs were considered as high confidence treatment-specific eQTLs. Using less stringent criteria we observed slight enrichment of putative inducible *cis*-eQTLs among up-regulated genes: up to 1.5% of *cis*-eQTLs overlapped genes whose expression was at least 1.5-fold higher upon treatment. We then selected ten genes that overlapped a treatment-specific *cis*-eQTL and were expressed in all treatments and carried out allelic expression (AE) measurements of unspliced pre-mRNA transcripts. We were however not able to validate any of the treatment specific *cis*-regulatory effects seen in the eQTL analysis. On-going genome-wide AE experiments upon treatment are designed to validate the surprising invariability of *cis*-regulatory landscapes in induced vs. resting cells and these results have widespread impact on design and interpretation of functional genomic studies in human cells.

## 41

**"Simple" Disorders, Complex Traits: A Search for Genetic Modifiers in Neurofibromatosis Type 1.** *A. Pemov[1], J.L. Sloan[2], D.R. Stewart[1].* 1) Dept GDRB, NHGRI/NIH, Bethesda, MD; 2) Dept GMBB, NHGRI/NIH, Bethesda, MD.

**Background.** The cause of the variation in phenotypic severity in neurofibromatosis type 1 (NF1) is unknown and may be due to genetic modifiers. A differentially expressed transcript that correlates in a statistically significant way with a measurable phenotype is referred to as a "quantitative trait transcript" (QTT). Identification of QTTs (and any corresponding expression quantitative trait loci (eQTL)) is a novel and emerging technique to identify modifier genes in animal models and humans (Passador-Gurgel et al., *Nat Genet* 2007; 39: 264-268; Morley et al., *Nature* 2004; 430: 743-747). We hypothesized that variation in *germline* gene expression of certain genes correlates with variation in the severity of quantifiable phenotypic features ("sub-phenotypes") of NF1. **Methods.** We performed whole-genome transcriptional profiling (Illumina HumanRef-8 arrays) in lymphoblastoid cell lines from 79 individuals affected with NF1 and 23 controls. A single observer quantified severity in multiple NF1 sub-phenotypes, including height, head circumference (OFC), burden of cutaneous neurofibromas (CNF), café-au-lait macules (CALM), Lisch nodules (LN), and cherry hemangiomas (CH). We examined the correlation of the 6 NF1 sub-phenotypes with the level of each of the 22,177 transcripts. To control for multiple testing, we calculated a False Discovery Rate (FDR), in addition to a nominal *P*-value of the significance of the regression. We filtered for FDR (< 0.3), expression range (~2X) and expression level (mean $\log_2 > 6.0$). **Results.** From the statistical analysis, we identified 32 unique transcript-phenotype pairs (QTTs). These included 6 genes whose expression level significantly correlated with CALM burden, 2 with CH burden, 8 with LN burden, 1 with CNF burden, 8 with height and 7 with OFC. We then validated 22 QTTs by quantitative PCR on low-density microfluidic arrays (ABI). By qPCR, 9 QTTs remained statistically significant (nominal *P*-value < 0.05). Many QTTs were gender-specific, even for traits not known for sexual dimorphism (e.g. café-au-lait macule burden). **Conclusions.** We identified 9 putative genetic modifiers (QTTs) of severity in NF1. On-going investigation of these 9 QTTs includes genotyping known eQTLs and gene-specific functional validation.

## 42

**Genome-wide association reveals master eQTL regulators of microRNA expression variation in human fibroblasts.** *C. Borel[1], S. Deutsch[1], A. Letourneau[1], E. Migliavacca[1], A.S. Dimas[2], C.E. Vejnar[1], H. Attar[1], M. Gagnebin[1], C. Gehrig[1], E. Falconnet[1], Y. Dupré[1], S.E. Antonarakis[1].* 1) Dept Genetic Medicine, Univ Geneva Medical Sch, Geneva, Switzerland; 2) Wellcome Trust Institute, Wellcome Trust Genome Campus, Cambridge, UK.

MicroRNAs (miRNA) have emerged as an important class of regulatory non-coding RNAs modulating the processing of over a third of human transcripts. Inter-individual variation of miRNA levels is likely to influence expression of their target genes, and may therefore contribute to phenotypic differences, including susceptibility to common disorders. The genetic control of microRNA expression levels is currently unexplored. In this study, we have performed a whole-genome, quantitative association study of 365 microRNA expression phenotypes in 180 primary fibroblasts from Caucasian newborns of the GenCord project, genotyped with the Illumina HumanHap550 array. We find extensive variation in microRNA expression levels and estimate that 33% of microRNAs are differentially expressed. For 121 selected expressed microRNAs, GWAS yielded highly significant cis- (9%) and trans- (11%) associations. Furthermore, we characterized large gene desert regions containing master eQTL regulators that influence the expression of multiple miRNAs, thus providing a previously unknown mechanism suggesting co-regulation of miRNA expression. This is the first attempt to characterize the genetic regulation of miRNA expression levels. eQTLs identified through this approach are likely to be important determinants of human phenotypes. C.B and S.D contributed equally to this work.

## 43

**Human-specific loss of regulatory DNA and the evolution of human-specific traits.** *G. Bejerano[1,2], C. McLean[2], P. Reno[1,3], A.A. Pollen[1], A. Bassan[1], T. Capellini[1], C. Guenther[1,3], V. Indjeian[1,3], X. Lim[1], D.B. Menke[1], A. Wenger[2], D.M. Kingsley[1,3].* 1) Department of Developmental Biology, Stanford University; 2) Department of Computer Science, Stanford University; 3) Howard Hughes Medical Institute.

The availability of several primate whole genome sequences has spurred great excitement for the prospect of understanding the molecular basis of what makes us human. Recent investigations have discovered dozens of conserved non protein coding genomic loci which have experienced accelerated basepair changes in the human lineage, as well as some proteins that show evidence of positive selection. Many previous genome surveys have focused on small base pair changes in otherwise well aligned sequences. Here we expand these studies to look for a type of event particularly likely to produce functional effects: complete deletion in humans of sequences that are otherwise highly conserved in other organisms. By searching for regions of the chimpanzee genome highly conserved over mammalian evolution that are clearly missing in humans, we discover 583 human-specific losses of putatively functional ancestral DNA. PCR and computational validation show that roughly 80% of the deletions are fixed in human populations, while others are polymorphic in different individuals. Most of the deletions removed conserved non-coding sequences rather than protein-coding regions, and many lie in proximity to genes involved in development, morphogenesis, neural function, and steroid hormone signaling. We have functionally tested a subset of human specific deletions in transgenic mice, and have found intriguing examples of regulatory alterations in humans that appear to be associated with evolution of specific anatomical differences between humans and other animals.

## 44

**Nucleosomes with specific histone marks are positioned in exons - A new Darwinian force with an epigenetic twist?** *C. Wadelius[1], R. Andersson[2], S. Enroth[2], A. Rada-Iglesias[2], F. de la Vega[3], K. McKernan[4], J. Komorowski[2, 5].* 1) Dept Gen & Pathology, Uppsala Univ, Uppsala, Sweden; 2) Linnaeus Centre for Bioinformatics, Uppsala University, Sweden; 3) Life Technologies, Foster City, CA; 4) Life Technologies, Beverly, MA; 5) Interdisciplinary Centre for Mathematical and Computer Modelling, Warsaw University, Poland.

It is known that nucleosomes are well positioned over the first exon in active genes, with histone modifications reflecting the transcription rate. We reanalyzed public nucleosome position data for man and C. elegans, histone modification data from man and mouse as well as gene and exon expression data from man to look for distinct patterns also in internal exons. We found one well-positioned nucleosome at internal exons with a signal clearly higher than at the TSS. The peaks are centered at +94 (human) and +101 (C. elegans) relative to the exon start meaning that the average 5' end of a nucleosome is positioned at +20 and +27 in man and worm, respectively. We found no positioned nucleosome in exons <50 bp, which comprise <5% of human exons, but in long exons there is a nucleosome positioned at the start and end. Nucleosomes were positioned at internal exons regardless of transcription level, in contrast to the situation at the TSS. We systematically screened 38 histone modifications to see if the nucleosomes had distinct patterns related to gene and exon expression. The H3K36me3 signal was significantly higher in exons than in the following introns in highly expressed human and mouse genes, p<10e-4 for each exon-intron comparison. This applies from the third exon and onwards. On the other hand H3K27me2 and me3 were associated with gene silencing. In highly expressed genes, high exon usage was associated with high H3K36me3 and low H3K27me2 and the opposite was found in exons with low usage. These data suggest that H3K36me3 might facilitate exon inclusion during co-transcriptional splicing and that splicing is under epigenetic control. We have generated deep data sets by sequencing nucleosomes, RNA and ChIP-DNA from HepG2 cells to further evaluate the findings. Our results show that exons are functional units not only defined by their coding capacity but also by the way they are packaged in nucleosomes. The factors controlling nucleosome positioning at internal exons must be under strong evolutionary constraint given the strikingly similar pattern in man and worm, with a common ancestor around 1 billion years ago. An effect on evolution is also supported by recent reports showing lower substitution rates in linker regions than in nucleosomal DNA as well as higher rates of insertions and deletions longer than 1 bp in linker regions.

## 45

**A gene network regulating lysosomal biogenesis and function.** *A. Ballabio, M. Sardiello, M. Palmieri, A. di Ronza, D.L. Medina, C. Di Malta, V. Embrione, G. Parenti.* TIGEM, Telethon Inst Gen & Med, Naples, Italy.

The lysosome is a specialized cellular organelle dedicated to degradative processes. As degradative requirements of the cell may vary depending on tissue type, age and environmental conditions, we postulated the presence of a system allowing the coordination of lysosomal activity. By using the g:profiler tool on publicly available microarray data we observed that genes encoding lysosomal proteins tend to have coordinated expression behavior. Pattern discovery analysis of the promoter regions of 96 known lysosomal genes resulted in the identification of a palindromic 10-bp GTCACGTGAC motif, corresponding to a specific E-box type regulatory element, highly enriched in this promoter set. We named this regulatory network CLEAR (Coordinated Lysosomal Expression And Regulation). CLEAR elements are also associated to promoters of genes encoding non-lysosomal proteins involved in lysosomal biogenesis and degradative pathways, such as autophagy. We also found that the CLEAR network can be predictive for the identification of novel lysosomal proteins. By chromatin immunoprecipitation analysis we demonstrated that the HLH-leucine zipper transcription factor TFEB binds to CLEAR sites. Overexpression and microRNA-mediated silencing experiments showed that TFEB positively regulates the expression of target genes, thus acting as a master regulator of the network. Intralysosomal storage of undegraded material determines TFEB activation by promoting its translocation from the cytoplasm, where it resides in resting cells, to the nucleus. An expansion of the lysosomal compartment was detected in HeLa transfectants stably overexpressing TFEB. Accordingly, ultrastructural analysis revealed a significant increase in the number of lysosomes per cell, indicating the involvement of TFEB in lysosomal biogenesis. TFEB stable transfectants displayed a faster rate of GAG clearance compared to controls. Furthermore, by using a rat striatal cell model HD43 that carries an inducible transgene for expanded huntingtin, we demonstrated an increased clearance of the mutant protein in TFEB overexpressing cells. An appealing perspective would be the use of the CLEAR network as a therapeutic target to enhance cellular response to intracellular pathogenic accumulation in neurodegenerative diseases.

## 46

**Initial Map of Core Regulatory Circuitry of Human Hematopoietic Stem Cells.** *L.N. Lawton[1], G.M. Frampton[1,2], M.C. Cole[1], R. Kumar[1], A. Drake[3], J. Chen[2,3], R.A. Young[1,2].* 1) Whitehead Institute for Biomedical Research, Cambridge, MA; 2) Dept. of Biology, Massachusetts Institute of Technology, Cambridge, MA; 3) Koch Institute, Massachusetts Institute of Technology, Cambridge, MA.

An understanding of the transcriptional regulatory circuitry that establishes and maintains gene expression programs in mammalian cells is fundamental to understanding development and should provide the foundation for improved diagnosis and treatment of disease. Although it is not yet feasible to map the entirety of this circuitry in all mammalian cells, recent work in embryonic stem (ES) cells by us and others has demonstrated that core features of the circuitry can be discovered through studies involving selected transcription factors, chromatin regulators and signaling pathways. Here we describe an initial map of the core regulatory circuitry of human hematopoietic stem cells (HSCs) using the concepts and technologies that have recently been used to discover this circuitry in mammalian ES cells (reviewed in Cole and Young 2008). HSCs are the best-characterized adult stem cells and are responsible for the long-term survival of patients treated clinically with bone marrow transplantation. In the adult, these cells have the capacity to self-renew in the bone marrow while retaining the ability to continually produce all of the mature cell-types in the blood and immune systems. The majority of adult HSCs are thought to be quiescent, while cells from the small active stem cell pool face two alternative fates: self-renewal or differentiation. Because these cells are in limited quantities, compromising as few as 1 million cells per umbilical cord blood or bone marrow sample, determining how key transcription factors contribute to core regulatory circuitry is challenging. In order to obtain sufficient numbers for such analysis, we utilized the 20-fold culture expansion of CD133+ HSCs from human umbilical cord blood that results in an increase in SCID-repopulating cells (SRCs) in xenotransplantation studies (Zhang and Lodish 2008). We have identified the direct targets of key regulators of HSC function genome-wide using chromatin immunoprecipitation coupled with sequence analysis (ChIP-Seq). These data, and functional evidence from transcription factor knockdown, have allowed us to generate an initial model of the core regulatory circuitry of human HSCs. We will present this model and discuss its implications for the control of HSC functions.

## 47

**A statistical approach for identifying differentially methylated CpG sites with DNA methylation array data.** *K.N. Conneely[1], J.F. Cubells[1,2], D.J. Newport[2], Z.N. Stowe[2], A.K. Smith[1,2].* 1) Dept Human Genetics, Emory Univ, Atlanta, GA; 2) Dept Psychiatry, Emory Univ, Atlanta, GA.

Emerging evidence suggests a potential role for DNA methylation in the etiology of numerous complex diseases. Variation in DNA methylation can lead to substantial changes in gene expression and can change in response to environmental stimuli, such as diet or toxins. The identification of CpG sites where methylation varies with disease phenotype or environmental background could implicate genes which act in concert with the environment to impact disease risk. The advent of microarray-based methods to interrogate CpG sites across the genome in a hypothesis-free manner has created a need for analytical methods to identify CpGs differentially methylated between groups. While standard methods of assessing differential gene expression may be adapted to this goal, differences in the measurement and distribution of DNA methylation may warrant methylation-specific statistical methods. The Infinium HumanMethylation27 BeadChip interrogates 27,578 CpG sites from 14,495 genes with replicate probes for both methylated and unmethylated alleles of each CpG. The ratio of the average intensity of methylated probes to the average intensity of both methylated and unmethylated probes for a specific CpG provides an estimate of the fraction of DNA methylated that has been suggested to follow a beta distribution due to its potential bimodality. We propose a score statistic for assessment of differential methylation assuming a beta distribution in which precision differs across samples according to the number of replicate probes for each sample and CpG. We investigate the applicability of this method to data from the HumanMethylation27 BeadChip. For DNA samples extracted from the umbilical cord blood of 10 male and 9 female neonates, we exploit the expected methylation differences due to X-inactivation to assess performance of our method, testing all CpG sites for differential methylation between male and female neonates. We assume that all CpGs on X, and none of the autosomal CpGs, are truly differentially methylated in males and females, a conservative assumption that will actually underestimate power and overestimate FDR. Nevertheless, at an estimated FDR of 5%, our method successfully detects 72% of all CpGs on X, and 81% of all CpG islands on X, outperforming methods for gene expression analysis such as Significance Analysis of Microarrays (SAM). Supported by the Emory Univ. Sch. Med. Dept. of Human Genetics and grant P50 MH MH077928.

## 48

**Detecting chromosomal aberrations and DNA mixtures using genome-wide SNP intensity data.** *T. Bhangale, C. Laurie, C. McHugh, F. Boehm, B. Weir, The GENEVA Investigators.* Univ Washington Seattle, Seattle, WA.

Large scale chromosomal abnormalities are often found in tumor DNA samples and are routinely used in predicting prognosis. Cell line and blood samples commonly used in genome-wide association studies (GWAS) can also develop such abnormalities, possibly due to a survival advantage conferred to the cells by multiple copies. In the GWAS samples, these abnormalities manifest in several forms viz. cells with two or more copies of a regions or of the entire chromosome; mixture of cells of two types: those with >2 copies and those with <2 copies. A related problem for GWAS is sample contamination e.g. mixture of DNA from multiple individuals, which produce characteristic patterns in the SNP intensity data. Detecting and characterizing chromosomal abnormalities play an important role in tumor study and in the QC process for GWAS. We developed a hidden Markov model (HMM) based algorithm for automating the detection of chromosomal abnormalities and DNA contaminations using the allele intensity data from genome-wide SNP chip: BAlleleFrequency (normalized angle sustained by the allele intensity vector of a sample for a SNP with the abscissa) and LogRRatio (total allele intensity at the SNP). BAlelleFrequency is modeled using a HMM where the states in the Markov chain correspond to normal/abnormal regions of the chromosome and the observed data is modeled using a mixture of normal distributions. First, Viterbi algorithm is used to determine the optimum hidden state sequence for each chromosome and thereby find those with abnormalities, as well as to determine the boundaries of the abnormalities. Further analysis is then done on the abnormal segments to distinguish between different types of abnormalities and to infer the number of copies. We apply the method to HapMap 3 intensity data and find 45 abnormalities in these samples. The method had a false discovery rate of 0.067 and a missed abnormality rate of 0.044 suggesting that it can be used to analyze GWAS samples or tumor DNA samples in clinical settings in a highly automated manner. Entire chromosome is affected in 11 abnormalities and >1/4th chromosome is affected in 8 of the remaining abnormalities. Locations of the discovered abnormalities may shed light on the biologic processes involved. As HapMap data is widely used by the community, careful QC of these data is important. Genotypes of SNPs in these abnormalities are likely to be incorrect. Thus our results also suggest additional QC steps for these data.

## 49

**Convincing identification of over 1,000 human trans-eQTLs by a meta-analysis of ~2,000 samples.** *L. Franke[1,2], R. Fehrmann[1,3,4], J. Fu[1], E. Festen[1], A. Smolonska[1], G. Trynka[1], M. Platteel[1], M. Bruinenberg[1], D.A. van Heel[2], C. Wijmenga[1].* 1) Genetics, University Medical Centre Groningen, Groningen, Groningen, Netherlands; 2) Institute of Cell and Molecular Science, Barts and The London School of Medicine and Dentistry, London, UK; 3) Department of Gynaecologic Oncology, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands; 4) Department of Medical Oncology, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands.

It is well known that genetic variants can exert effects on gene expression (eQTLs). However, their effect sizes are usually small (likely comparable to variants that affect complex traits such as lipid levels, adult height and many diseases). Thus, in order to be adequately powered to detect these eQTLs, considerable amounts of samples are required and expression levels should be quantitated as accurately as possible. To achieve sufficient statistical power, we conducted a meta-analysis on 11 'genetical genomics' datasets (comprising over 2,000 samples). To quantitate expression levels as accurately as possible we used a two step approach: We first applied principal component analysis to unrelated expression data (over 20,000 samples) from the Gene Expression Omnibus, resulting in the identification of 50 principal components that reflect environmental factors, some of which exert strong influences on global gene expression levels. Subsequently, by superimposing these principal components on the various genetical genomics datasets, we could correct per individual sample or these factors, enabling considerably improved eQTL mapping. Over 15,000 unique SNPs and CNVs (imputed using TriTyper, Franke et al, AJHG 2008) were identified that affect gene expression levels (FDR 0.05). Among these, over 1,000 trans-eQTLs were convincingly detected. Subsequent biological interpretation was conducted by using a gene network (Franke et al, AJHG 2006, constituting over 80,000 known interactions derived from BIND, HPRD, IntAct, Reactome and KEGG). This analysis indicates that the genes within the eQTL locus can usually be connected to the trans-gene with a fewer number of steps through the network than within permuted data. Most detected eQTLs are tissue specific, whereas eQTLs that are detected in all tissues often reflect probe hybridization artefacts.

## 50

**Model selection via penalized logistic regression.** *K. Ayers, H. Cordell.* Institute of Human Genetics, Newcastle University, Newcastle Upon Tyne, United Kingdom.

Penalized regression methods are an attractive alternative to single marker regression in association analysis. In underdetermined problems, regression methods are overwhelmed as the model becomes saturated and over fitting occurs. Penalized regression methods shrink the size of the coefficient of a marker with little effect on the trait of interest, ideally leading to a subset of markers most associated with the disease. Here we explore the advantages and pitfalls of penalization in selecting predictors in genetic association studies. Penalization parameters are data dependent in that the same penalty can lead to a much larger set of predictors in one data set over another. The penalization parameter values that lead to the best model may be selected through computationally expensive cross validation. The best model may also be chosen through a maximum likelihood based model selection criteria, such as the BIC, but in practice we find these scores allow far too many variables. Drawbacks of penalized methods include: (1) there is no efficient way to obtain a confidence interval or p-value for a coefficient, and (2) typically, a variable may only enter the model if it significantly improves prediction. Thus, if other variables better explain an effect, variables with a strong marginal effect may be overlooked. We investigated the performance of several software packages for performing penalized logistic regression, using computer simulations. We simulated data under a variety of disease locus effect size and linkage disequilibrium patterns. We compared several penalized methods, such as the elastic net, the lasso, the MCP, and the Bayesian inspired NEG shrinkage prior, to standard single locus analysis (the ATT), and simple forward stepwise regression. We investigated how markers enter the model as penalties and p-values are varied, and report the sensitivity and specificity for each of the methods. Preliminary results show similar performance in penalized methods and the ATT, with the main difference being penalized methods generally do not allow correlated variables to enter the model, leading to a sparser model in which most of the explanatory markers are accounted for. Forward stepwise regression performs poorly with a higher false positive rate.

## 51

**A unified mixed effects likelihood framework for detecting associations with rare variants using sib and unrelated individuals with extreme quantitative phenotypes: application to next generation sequencing data.** *D.J. Liu[1,2], S.M. Leal[1,2].* 1) Dept Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Dept Statistics, Rice University, Houston, TX.

There is strong support that quantitative phenotypes can be due to rare variants. Next generation sequencing makes it possible to identify rare variants, which is the first step to carry-out direct mapping association studies of quantitative trait loci (QTL). In designing association studies to identify rare QTL (rQTL), sampling individuals with extreme quantitative trait values (QTVs) can be used to enrich for rare causal variants. In particular rare causal variants are more likely to be aggregated in related individuals with extreme phenotypes. Although extreme trait ascertainment in related individuals has had only limited success in linkage and association studies, this design is highly advantageous in the analysis of rare variants, because the success of rQTL mapping is highly dependent on causal rare variant enrichment. Collecting a specific family type (e.g. discordant sib-pairs) can be difficult, therefore it is desirable to be able to combine different family structures as well as unrelated individuals in rQTL analysis. Although methods do exist to analyze QTL due to common variants these methods are underpowered when applied to the analysis of rQTL data, and inflexible for modeling relative phenotype correlations due to multiple shared rare variants. Additionally these methods cannot combine the analysis of QTL data from related and unrelated individuals. We propose a flexible likelihood framework with mixed effects for modeling extreme trait genetic associations with rQTL (MEGA-rQTL) for the analysis of related and unrelated individuals with extreme QTVs. MEGA-rQTL detects associations with rQTL through likelihood ratio tests. A unique feature of MEGA-rQTL is that parameters of genetic interests such as heritability and sibling residual correlation can be efficiently estimated. We investigated the power of the MEGA-rQTL method, for 7 commonly used prospective selective sampling strategies. Simulation was carried out via coalescence theory using parameters estimated from population genetic data and heritabilities from clinically relevant quantitative traits. We demonstrate that analyzing sib-pairs with extreme QTVs or using one sib per sib-pair with extreme QTVs are consistently more powerful than using unrelated individuals with extreme QTVs. In conclusion, MEGA-rQTL is a powerful approach to analyze next generation sequence data to map QTL due to rare variants by combining data from related and unrelated individuals with extreme phenotypes.

## 52

**A statistical framework for genotype/phenotype analysis by re-sequencing.** *G.V. Kryukov[1], A.L. Price[2], P.I. de Bakker[1], S. Purcell[3], L.-J. Wei[4], S.R. Sunyaev[1].* 1) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115; 2) Department of Epidemiology, Harvard School of Public Health, 665 Huntington Avenue, Boston, MA 02115; 3) Center for Human Genetic Research, 185 Cambridge Street, CPZ-N, Simches Research Center, Massachusetts General Hospital, Boston, MA 02114; 4) Department of Biostatistics, Harvard University, 677 Huntington Avenue, Boston, MA 02115.

Continuing reduction in cost of DNA sequencing will enable human geneticists to relate full sequence information in genes and, soon, complete genomes to human traits of clinical relevance. Deep sequencing in large samples promises to reveal a vast trove of rare alleles, a significant fraction of which may be important determinants of complex traits. Although knowledge of all rare variants segregating in the population would seem to increase the power of genetic analysis, this prospect faces daunting statistical challenges. A larger pool of sequence variants would require a more stringent multiple testing correction, while the power to detect an association for less common variants is reduced. This problem can be surmounted by pooling rare allelic variants by gene, functional non-coding region or pathway. Statistical analysis based on pooled allelic variants considers composite information from individual genes rather than individual variants as the unit of the association test.

We developed two statistical methods to detect association of multiple rare variants in protein coding genes with a quantitative trait. The key feature of these methods is that they utilize both computational predictions of the functional effects of missense variants and allele frequency distribution.

The first method is a fully parametric likelihood ratio test based on a population genetic model. The test compares the likelihood that observed phenotypic values are independent of individuals' genotypes with the likelihood that individuals carrying variants predicted to be functionally significant have a distribution of phenotypic values with a shifted mean.

The second method is based on regression of phenotypic values on individuals' genotype scores. Genotype scores are calculated on the basis of functional predictions for variants present in an individual that have an allele frequency below a variable gene-specific threshold. This method optimizes the gene-specific threshold by permutation testing, making it robust with respect to properties of allelic distributions in individual genes.

Both these methods were extensively evaluated using both simulated and real deep re-sequencing data.

## 53

**Bayesian methods of estimating ancestry using whole-genome SNP data.** *C. Churchhouse, J. Marchini.* Department of Statistics, University of Oxford, Oxford, United Kingdom.

Estimation of the genetic ancestry of an individual is useful for association studies, disease risk prediction, population genetic analyses and is of inherent interest for the individual themselves. We have investigated methods of estimating ancestry using whole-genome SNP data on each individual. We focus on the scenario where the goal is to determine ancestry in relation to a set of genotype or haplotype data that is available from a set of distinct source populations, for example, the HapMap 2, HapMap 3 or 1000 Genomes datasets. Inference in this setting can focus either on the estimation of global ancestry, in which an overall estimate of the proportion of ancestry from the source populations is needed, or local ancestry, which aims to partition an individual genome into distinct segments of ancestry from the source populations. We have compared 2 models based on the estimated allele frequencies in the source populations at a set of unlinked SNPs. Model 1 only models global admixture, whereas Model 2 models both global and local admixture. Using simulated individuals with differing proportions of CEU and YRI admixture (based on HapMap3 data) we find that there is a relatively small difference in the mean square error of the estimates of global admixture from the 2 methods ($1.16\ 10^{-4}$ and $8.88\ 10^{-5}$ respectively). Since Model 1 is much faster to fit that Model 2 these results suggest that Model 1 can be used to estimate the level of global ancestry, or at the very least will be useful as an initial estimate for use in Model 2. Further investigation is required to see how these results hold for more genetically similar source populations. In contrast, the mean square error for the estimates of local admixture from the 2 methods is 0.298 and 0.0861 respectively, suggesting that an explicit model of local ancestry is needed to carry out this level of inference. We are also investigating the utility and practicality of using linked SNP data to estimate global and local admixture.

## 54

**The advance of Bayesian methods for genetic association analysis.** *D.J. Balding.* Institute of Genetics, University College London, London, United Kingdom.

Bayesian statistical methods have revolutionised many scientific areas in the past 15 years, but human genetics has to a large extent resisted this advance. The publication of Bayesian analyses by the Wellcome Trust Case Control Consortium in 2007 may have signalled a tidal change, and there has been a flurry of developments since then. Awareness of the limitations of classical methods is also growing, for example that p-values do not indicate the credibility of a putative association, and can be seen as implicitly assuming a prior with serious drawbacks. One of the barriers to the adoption of Bayesian methods is computation: association datasets are getting larger and potential analyses are getting more complex. However, shortcuts have been developed that make some Bayesian analyses as fast as classical methods. Adoption of Bayesian methods can bring many benefits, for example in more rational ways to deal with the problem of multiple testing and to combine results via meta-analyses. I will briefly review general developments in the field, and then focus on specific novel methods that illustrate the benefits of being Bayesian. In the analysis of quantitative traits, attention usually focusses on differences in means across genotypes, but differences in variances can also be very informative, particularly when the observed weak association at a common variant is due to a stronger but rarer causal variant. In the classical paradigm one can test means or test variances, but it is difficult to do both simultaneously in a principled manner; this is straightforward in a Bayesian analysis, and computationally fast. Bayesian methods are also advantageous for prediction of disease state or drug response, since a careful choice of prior can lead to appropriate values for the number of predictors and linkage disequilibrium among them. Bayesian methods have not to date been able to deal effectively with population structure, and I will introduce a new and very promising method for this. All the methods will be illustrated using simulated and real genome-wide association data.

## 55

**Analyses conditional on established type 2 diabetes loci reveal putative novel associations.** *T. Ferreira[1], E. Zeggini[2], A.P. Morris[1].* 1) Genetic and Genomic Epidemiology Unit, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Wellcome Trust Sanger Institute, Hinxton, United Kingdom.

With the success of genome-wide association (GWA) studies of type 2 diabetes (T2D) in the last 3 years, there are now 19 established risk loci that have been replicated in samples from multiple populations. Despite this enormous step forward, much of the genetic contribution to the variation in T2D risk remains unexplained. One way to potentially increase the power of existing GWA studies to identify further variants contributing to susceptibility to T2D is to perform conditional analyses, accounting for genotypes of individuals at leading SNPs at established risk loci.

We consider 2 conditional analysis paradigms: (i) model the linear trend in the main effect of a SNP, conditional on the linear trend of the main effect of the leading SNP at an established risk locus; and (ii) the same model, but incorporating an interaction between the 2 SNPs, equivalent to a test of association within each genotype group at the leading SNP of the established risk locus.

We have analysed data from the Wellcome Trust Case Control Consortium (WTCCC) consisting of ~2000 cases of T2D and ~3000 controls. We applied the same QC filters as employed by the WTCCC and applied the two tests described above, genome-wide (GW), conditioning on the effects of each established T2D locus. When searching for main effects only, conditional on all established loci simultaneously, only moderate evidence of association is obtained ($p<10^{-5}$, but not exceeding GW significance). The strongest novel signals of association were observed on chromosomes 8p21 ($p=3.4\times10^{-6}$ compared with unadjusted $p=8.9\times10^{-5}$) and 4q31 ($p=5.7\times10^{-6}$ compared with unadjusted $p=6.9\times10^{-5}$). When allowing for interaction effects, considering each established locus in turn, the strongest signal of association is observed on chromosome 5p15 by conditioning on the lead SNP at the *FTO* locus ($p=2.3\times10^{-7}$ compared with unadjusted $p=2.5\times10^{-2}$). Weaker evidence of association is observed on chromosome 11q14 by conditioning on the lead SNP at the *IGF2BP2* locus ($p=4.0\times10^{-6}$ compared with unadjusted $p=2.5\times10^{-3}$). These signals warrant follow-up in studies from other populations, and are currently being undertaken as part of the Diabetes Genetics Replication and Meta-Analysis consortium, with the hope that they might provide further insight into the genetic mechanisms underlying T2D risk.

## 56

**Pleiotropic effects on T cell activation are associated with PDAY (*Pathobiological Determinants of Atherosclerosis in Youth*) risk score in the San Antonio Family Heart Study.** *A. Vinson[1], V.P. Diego[2], H.H.H. Goring[2], J.E. Curran[2], M.P. Johnson[2], T.D. Dyer[2], E.K. Moses[2], D.L. Rainwater[2], A.G. Comuzzie[2], J.W. MacCluer[2], J. Blangero[2], M.C. Mahaney[2].* 1) Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR; 2) Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX.

$CD4^+$ effector T cells modulate inflammation in atherosclerosis, but the extent of pleiotropic effects on these pathways and their influence on risk of atherosclerosis have not been characterized. This study aimed 1) to estimate pleiotropic effects on heritable expression levels of 56 candidate genes implicated in multi-cellular pathways of T helper-1 (Th1) and T helper-2 (Th2) cell activation, differentiation, and cytokine secretion, and 2) to assess whether these effects influence a composite index score of risk for advanced atherosclerotic lesions (PDAY, *Pathobiological Determinants of Atherosclerosis in Youth*) measured in members of the San Antonio Family Heart Study (SAFHS). Using the results of a transcriptional profile of lymphocytes in 1,240 members of the SAFHS, we conducted bivariate analyses employing a maximum likelihood variance decomposition approach to estimate all pair-wise additive genetic correlations among candidate gene transcript levels. Results of these analyses revealed evidence supporting substantial shared additive genetic (i.e., pleiotropic) effects on transcript levels (208 of 1,540 pairwise estimates of $\rho_G$ at FDR≤0.05, involving 55 of 56 genes). To reduce these data to underlying orthogonal factors, we conducted a principal components analysis and scored individuals on the first 10 components (accounting for ~80% of the total variance in candidate gene expression levels). We treated these components as covariates in polygenic models and employed a Bayesian model averaging approach to assess evidence supporting hypotheses of effect on PDAY scores (including no effect) due to single or multiple components. Results of this analysis found the second component (accounting for 14% of total variance) to be the best predictor of PDAY scores (N=1,171, BIC=-4.01, effect size=-0.15, posterior probability=0.89). Expression levels of 12 genes (*CD28*, *CD3D*, *CD3E*, *CD3G*, *CSF2RB*, *CTLA4*, *ICOS*, *IL10RB*, *LAT*, *LTB*, *NFKB1*, and *TRAT1*) were highly correlated with this component (rotated component loadings ≥|0.40|). The majority of these genes code for structural components of the T cell receptor complex, or for molecules required for co-stimulation and signal transduction in T cell activation. We conclude that pleiotropic effects on gene expression implicated in Th1 and Th2 effector pathways are substantial, and that pleiotropic effects on pathways of T cell activation in particular influence risk for severe atherosclerosis in this population.

## 57

**GM1 buildup at the ER-mitochondrion microdomains links the unfolded protein response to mitochondrial apoptosis in a neurodegenerative disease.** *A. D'Azzo[1], I. Annunziata[1], R. Sano[1, 2], A. Patterson[1].* 1) Department of Genetics, St Jude Children's Research Hospital, Memphis, TN; 2) Burnham Institute for medical research, La Jolla, CA.

Glycosphingolipids are emerging as determinants of apoptosis in physiological and pathological conditions. In the neurodegenerative lysosomal disorder GM1-gangliosidosis, progressive accumulation of GM1-ganglioside (GM1), due to deficiency of acid β-galactosidase, is associated with neuronal apoptosis. We have shown previously that in the mouse model of this disease, abnormal buildup of GM1 at the ER membrane induces depletion of Ca2+ stores and activation of the unfolded protein response that ultimately leads to cell death (Tessitore et al, Mol. Cell 2004). In this study we show that prior to neuron loss GM1 accumulates specifically in the glycosphingolipid-enriched microdomains (GEMs) of the mitochondria associated membranes (MAMs), the sites of juxtaposition between ER and mitochondria that buffer Ca2+ between these organelles. In affected brains, GM1 interacts with the phosphorylated, active form of the IP3R-1 favoring Ca2+ flux from the ER to the mitochondria. The ensuing depletion of ER Ca2+ is followed by mitochondrial uptake leading to Ca2+ overload that unleashes a cyclosporine A-sensitive opening of the permeability transition pore (PTP), cytochrome c release and activation of the downstream caspase cascade. These studies propose a new mechanism of Ca2+-mediated apoptotic signaling at the GEMs where GM1 acts as molecular effector of both ER stress- and mitochondria-mediated apoptosis. (This work was supported in part by NIH grants DK52025 and GM60905, the NIH Cancer Center Support Grant CA021765, the Assisi Foundation of Memphis, and the American Lebanese Syrian Associated Charities (ALSAC) of SJCRH).

## 58

**Novel missense mutations in the Menkes disease copper transporter, ATP7A, cause isolated, progressive, adult-onset distal motor neuropathy.** *M. Kennerson[1, 2], G. Nicholson[1, 2], B. Kowalski[1], S. Chu[1], R. Takata[3], C. Speck-Martins[3], J. Baets[4], V. Timmerman[4], S. Kaler[5], J. Tang[5], J. Mercer[6], R. Llanos[6], P. De Jonghe[4], M. Shy[7], J. Garbern[7].* 1) Northcott Neuroscience Laboratory, ANZAC Research Institute, Concord, New South Wales, Australia; 2) Molecular Medicine Laboratory, Concord Hospital, Concord, New South Wales, Australia; 3) Sarah Network Rehabilitation Hospitals, Brasilia, DF Brazil; 4) Department of Molecular Genetics, VIB and University of Antwerpen, Antwerpen, Belgium; 5) Unit on Pediatric Genetics, Molecular Medicine Program, National Institute of Child Health and Human Development, National Institutes of Health. Bethesda, MD; 6) Centre of Cellular and Molecular Biology, School of Life and Environmental Sciences, Deakin University, Burwood, Victoria, Australia; 7) Department of Neurology, Wayne State University School of Medicine, Michigan, USA.

**Background:** The distal hereditary motor neuropathies (distal HMN) comprise a clinically and genetically heterogeneous group of disorders affecting the distal motor neurons. Recently, one form (DSMAX) was mapped to chromosome Xq13.1-q21 (*J Med Genet* 2004; 41:224-229). DSMAX presents as a distal muscular atrophy affecting the lower legs and feet with minimal sensory involvement. Symptoms typically commence in the second or third decade of life. We identified a U.S. X-linked dHMN family mapping to and refining the DSMAX locus (*Neurology* 2009; 72:246-252). Among the candidate genes included within the DSMAX interval is ATP7A, which encodes a copper-transporting p-type ATPase mutated in Menkes disease, a severe neurodegenerative condition with early infantile onset. **Methods:** We used high resolution melt analysis, DNA sequencing, reverse-transcriptase polymerase chain reaction, western blotting, yeast complementation, and immunohistochemical analysis of patient fibroblasts to identify and characterize novel ATP7A mutations associated with adult-onset isolated distal motor neuropathy. In selected affected individuals, we assessed biochemical markers of copper metabolism and performed electrophysiological studies. **Results:**We identified novel missense ATP7A mutations (T994I, and P1386S) in two families of Brazilian and U.S./European origin. The mutations, which do not interrupt ATP7A functional domains, segregate with the disease phenotype in each family and were absent in 800 control chromosomes. ATP7A screening in additional unrelated individuals with distal HMN disclosed three other novel sequence variations whose pathogenicity is currently being confirmed. Affected individuals with the P1386S mutation show no physical stigmata of Menkes disease or its known allelic variants, nor biochemical evidence of systemic abnormal copper metabolism. Western blots of P1386S fibroblast protein show normal ATP7A quantity as well as molecular mass compared to normal controls. **Conclusions:** While ATP7A mutations are typically associated with severe Menkes disease or its milder allelic variant, occipital horn syndrome, we report here that subtle missense mutations at this locus can cause a distinct syndrome of progressive, isolated distal motor neuropathy with onset in the second decade of life or later. This novel genotype-phenotype correlation reveals a potentially new and important role for the ATP7A copper transporter in lower motor neuron maintenance and function.

## 59

**27-hydroxy-7-dehydrocholesterol is an endogenous teratogen in Smith-Lemli-Opitz syndrome (SLOS) that decreases cholesterol levels and increases phenotypic severity.** *C. Wassif[1], E. Merkel[1], N. Javitt[1], E. Lietersdorf[2], F. Porter[1].* 1) Section on Molecular Dysmorphology, Eunice Kennedy Schriver National Institute of Child Health and Human Development, NIH, DHHS , Bethesda, MD; 2) Department of Pathology, Hadassah University Hospital, 91120 Jerusalem, Israel.

SLOS is a multiple malformation, mental retardation syndrome with a variable phenotype caused by mutations of 7-dehydrocholesterol reductase gene (*DHCR7*). Aberrant sterols and steroids synthesized from 7-dehydrocholesterol (7DHC) may have altered biological function and contribute to the SLOS pathology. We identified a novel oxysterol, 27-hydroxy-7-dehydrocholesterol (27-7DHC), and found a significant negative correlation between plasma 27-7DHC and cholesterol in SLOS patients. We thus hypothesized that 27-7DHC contributes to the pathology of SLOS. Specifically that increased 27-7DHC during development would suppress sterol synthesis, further impairing hedgehog signaling. To test this we produced SLOS mice (Dhcr7-/-) expressing a CYP27 transgene to elevate endogenous 27-7DHC. CYP27Tg mice have elevated CYP27 expression and 27-C levels in all tissues but normal cholesterol levels; Dhcr7-/-:CYP27Tg embryos have elevated 27-7DHC. Dhcr7-/- mice have growth retardation, cleft palate (9%), neurologic abnormalities, and die within 1 day. Consistent with the hypothesized detrimental effects of 27-7DHC during embryogenesis, Dhcr7-/-:CYP27Tg embryos have a more severe mutant phenotype. Dhcr7-/-:CYP27Tg embryos are stillborn. Malformations include growth retardation, micrognathia, cleft palate (77%), lingual and dental hypoplasia, ankyloglossia, umbilical hernia, cardiac defects, presence of a cloacae, curled tails, limb defects, and autopod defects. Brain malformations include ventricular dilatation, absence of the corpus callosum, and a hypoplastic cortical plate. Consistent with our hypothesis total sterol levels were decreased 2-fold in liver and 20-fold in brain tissue of Dhcr7-/-:CYP27Tg compared to littermate Dhcr7-/- embryos. We also observed alterations in tissue sterol content in littermates, suggesting that 27-7DHC could have a hormonal-like activity. Consistent with this proposal, significant amounts of 27-7DHC were detected in the pregnant dams serum if Dhcr7-/-: CYP27Tg embryos were present. This could explain why human mothers carrying SLOS affected fetuses fail to have a normal rise in cholesterol during pregnancy. Reversal of 27-7DHC effects as demonstrated in Dhcr7-/-:Cyp27-/- embryos revealed increased cholesterol synthesis and restoration of intrauterine growth. The pathological role of 27-7DHC in SLOS may explain some of the phenotypic variability.

## 60

**Mutant glucocerebrosidase as a parkin substrate: A possible molecular link between Gaucher disease and Parkinson disease.** *M. Horowitz, I. Ron.* Cell Res & Immunology,Wise Fac, Tel Aviv Univ, Ramat Aviv, Israel.

Gaucher disease, a sphingolipidosis characterized by impaired activity of the lysosomal enzyme glucocerebrosidase, results from mutations in the glucocerebrosidase encoding gene, GBA. We have shown that mutant glucocerebrosidase variants present variable degrees of ER retention and undergo ER associated degradation (ERAD) in the proteasome. Furthermore, the degree of ERAD of mutant glucocerebrosidase variants correlates and therefore determines Gaucher disease severity. An association between Gaucher disease and Parkinson disease has been demonstrated by the concurrence of Parkinson diseases in patients with Gaucher disease and the identification of glucocerebrosidase mutations in probands with sporadic Parkinson disease. One of the genes involved in Parkinson disease is parkin, encoding an E3 ubiquitin-protein ligase. Parkin also functions in the ERAD of misfolded ER proteins, and it is up-regulated by unfolded-protein response (UPR). Loss of parkin function leads to accumulation of its substrates, which is deleterious to dopaminergic neurons. We, therefore, tested the possibility that the association between Gaucher disease and Parkinson disease reflects the fact that parkin is an E3 ligase of mutant glucocerebrosidase variants. Our results showed that mutant glucocerebrosidase variants associate with parkin. Normal parkin, but not its ring fingar mutants, affects the stability of mutant glucocerebrosidase variants. Parkin is involved in K48 mediated polyubiquitination of mutant glucocerebrosidase forms and their degradation, indicating its function as an E3 ligase of mutant glucocerebrosidase variants. We suggest that involvement of parkin in degradation of mutant glucocerebrosidase explains the concurrence of Gaucher disease and Parkinson disease.

## 61

**Collaborative use of a long-term follow-up database after newborn screening: initial outcomes for medium-chain acyl CoA dehydrogenase deficiency.** *S. Berry for the Region 4 Genetics Collaborative - Priority 2 Workgroup IBEM-IS for Long-term Follow-up.* Michigan Public Health Institute, Okemos MI.

To improve outcomes from newborn blood spot screening (NBS), metabolic clinicians in states participating in the HRSA-funded Region 4 Genetics Collaborative (IL, IN, KY, MI, MN, OH, WI) agreed that they would collaborate in gathering data about children after NBS. Our initial goals from this activity were twofold: we wanted to show that metabolic practitioners could efficiently gather information about patient outcomes in a systematic fashion and we wanted longitudinal information about clinical outcomes after NBS. We initiated data collection in January 2007 with a focus on MCAD deficiency. We wanted to test the hypotheses that children with the highest C8 screening values would be most symptomatic with a secondary question regarding the nature of mutations found in these children with the hypothesis that the highest C8 values would be found in children who are 985 A>G homozygotes. Results: As of May 2009, 42 infants ascertained by NBS (with C8 as the informative value) had entries with information about lab abnormalities or other symptoms at the time of initial metabolic presentation. Of the 21 patients with the lowest C8 values (range 0.4-8.69, "lo") 17 had 2 mutations ascertained; 5 were 985 A>G homozygotes. Of the 21 patients with the highest C8 values (range 8.97-38.8, "hi") 16 had two mutations ascertained; 11 were 985 A>G homozygotes. Of the 21 lo range patients, one was noted to have abnormal liver function tests, one had respiratory distress due to prematurity, one was admitted for possible apnea but apnea was not confirmed. All others had no labs done, no abnormal labs and/or had no symptoms. Of the 21 hi range patients; six had either a lab abnormality or symptom (lab findings: (lab values suggesting) dehydration, hypoglycemia, elevated liver function tests; symptoms: loose stools, fever, pallor, limp, poor breastfeeding.) Of these, three had both a lab abnormality and symptom. We conclude that a long-term follow-up database can be successful in ascertaining systematic data about short-term follow-up for children with inborn errors of metabolism. Further, though preliminary, higher C8 values on NBS for MCAD deficiency are associated with a higher risk of being a 985 A>G homozygote and having a symptom or lab abnormality at the time of NBS. Supported by HRSA MCHB U22MC03963 - Priority 2 Project: Follow-up Activity.

## 62

**Prevalence of autism spectrum disorders in subjects with definite diagnosis of mitochondrial cytopathies.** *F. Scaglia[1,2], S. Zhang[1], Z. Tan[1], N. Fouladi[1,2], E. Schmitt[1], L-J. Wong[1].* 1) Dept Molec & Human Gen, Baylor Col Med, Houston, TX; 2) Texas Children's Hospital, Houston, TX.

Mitochondrial encephalomyopathies exhibit a wide range of central nervous system abnormalities including focal brain necrosis, dementia, encephalopathy, cognitive impairment, depression and schizophrenia. Mitochondria play an important role in regulating neurite outgrowth, axonal polarity, and synaptic plasticity. Synaptic and dendritic organizations of brain structures are impaired in autism spectrum disorders. There is cumulative evidence that mitochondrial dysfunction may be linked to autistic features, however; no systematic studies have been conducted thus far to assess the occurrence of autism in a large population of patients suspected of having mitochondrial disease. We investigated the clinical endophenotype of autistic features observed in patients with definite mitochondrial dysfunction and further characterized these patients from a biochemical and molecular standpoint. We reviewed the records of 4,194 subjects suspected of mitochondrial dysfunction who were evaluated through the Pediatric Genetics Clinic at Texas Children's Hospital and the Mitochondrial Diagnostic Laboratory at Baylor College of Medicine between April of 2005 and February of 2009. Among this cohort, 276 patients exhibited autistic features. Within this last group, fourteen patients met definite diagnostic criteria for mitochondrial disease by using the modified Walker criteria. There was a male predominance with a gender ratio of 2.5. The predominant neurological features accompanying the autistic features were developmental delay, ataxia, dystonia and seizure disorder. Two patients harbored the common m.3243 A>G mutation associated with MELAS syndrome, two patients had POLG mutations, two carried primary LHON mutations, one had mitochondrial DNA depletion on skeletal muscle, and interestingly one patient was homozygous for a SCO2 mutation and demonstrated COX deficiency. In addition four subjects presented with complex IV deficiency, one with complex I deficiency and another one with complex II deficiency. This study provides further evidence that there is disturbed energy metabolism as an underlying pathological mechanism in a specific subset of patients within the spectrum of autism spectrum disorders.

## 63

**Transaldolase deficiency : 4 new patients and new pathophysiological insights on the pentose-phosphate pathway.** *V. Valayannopoulos[1,6,7], M. Rio[2,6,7], M. Wamelink[3], D. Rabier[4,6], D. Habes[5, 8], G. Salomons[3], A. Slama[8,9], P. de Lonlay[1,6,7], O. Bernard[5,8], C. Jakobs[3].* 1) Metabolic Departement, Necker-Enfants Malades Hosp, Paris, France; 2) Genetics Department, Necker-Enfants Malades Hosp, Paris, France; 3) Dept. Clinical Chemistry VU University Medical Center, Amsterdam, The Netherlands; 4) Biochemistry Lab,Necker-Enfants Malades Hosp, Paris, France; 5) Pediatric Hepatology Dept, Bicetre Hospital, Le Kremlin-Bicetre, France; 6) Universite Paris Descartes, Paris, France; 7) INSERM U781, Paris , France; 8) Universite Paris Sud 11, Le Kremlin-Bicetre, France; 9) Biochemistry Lab, Bicetre Hospital, Le Kremlin-Bicetre, France.

Background: Transaldolase deficiency an inborn error of the pentose phosphate pathway (PPP) has been diagnosed so far in 10 patients from 6 different families. All presented as neonates, or even prenatally, with liver disease and the clinical courses have been diverse. Patients: we present 4 new patients from 2 families of African origin. The 2 siblings (boy and girl) from the first non-consanguineous family presented with low birth weight, cutis laxa, anemia, thrombopenia, neonatal cholestasis, elevated transaminases and alpha fetoprotein and eventually liver failure. Transient cardiac hypertrophy and pulmonary hypertension were also found. Anemia, thrombopenia, splenomegaly and hepatomegaly persisted in the first sibling and cirrhosis was diagnosed on a liver biopsy at age 4 while in the second sibling the liver fibrosis was found. The clinical presentation of the 2 siblings from the second family was milder. They were both diagnosed with hepatosplenomegaly, high transaminases and failure to thrive within the first months of life. The first child also displayed transient cardiomyopathy. A liver biopsy was performed at age 2 years because of progressive liver failure revealing cirrhosis. The second child had a similar presentation with hepatosplenomegaly and liver cirrhosis at age 9 months. Biochemical investigations: polyols were measured in urine, revealing elevation of erythritol, arabitol and ribitol associated with sedoheptulose and sedoheptulose-7 P strongly suggestive of TALDO deficiency. In one patient, respiratory chain studies were performed in fibroblasts and liver tissue revealing complex I deficiency. Enzyme studies and molecular investigations are currently performed to confirm diagnosis. Discussion: the new patients confirm that TALDO deficiency is panethnic and the constant hallmarks include liver impairment with cirrhosis associated to hematological abnormalities, and various heart symptoms. Elevation in urinary polyols and sugar phosphates has been constantly found in all patients confirming that this method constitutes a valuable diagnostic tool. So far sugar phosphate toxicity and osmotic imbalance have been proposed as pathophysiological mechanisms in this disorder. In TALDO deficient mice impaired mitochondrial function was recently reported leading to sterility. Our findings in the liver of one of the TALDO deficient patients may suggest similar mechanisms for liver damage that yet have to be confirmed in other patients.

## 64

**Mutation analysis of homogentisic acid oxidase (*HGD*) in alkaptonuria: Exploring excessive homozygosity.** *T. Vilboux[1], M. Kayser[1,2], C. Ciccone-Stevens[1], T. Markello[1], W. Introne[3], P. Suwannarat[1,4], I. Bernardini[1], R. Fischer[1], K. O'Brien[3], R. Kleta[1,5], M. Huizing[1], W. Gahl[1,3].* 1) MGB, NHGRI, NIH, Bethesda, MD; 2) Warren Clinic Center for Genetics, Saint Francis Health System, Tulsa, OK; 3) Office of Rare Diseases, Intramural Program, Office of the Director NIH, Bethesda, MD; 4) Department of Pediatrics, Children's Hospital of Pittsburgh and University of Pittsburgh School of Medicine, Pittsburgh, PA; 5) UCL, Centre for Nephrology, London, UK.

Alkaptonuria (AKU) is a rare autosomal recessive metabolic disorder characterized by accumulation of homogentisic acid leading to darkened urine, pigmentation of cartilage and other connective tissues (ochronosis), joint and spine arthritis and destruction of cardiac valves. AKU is due to mutations in the homogentisate dioxygenase gene, *HGD*, which converts homogentisic acid to maleylacetoacetic acid in the tyrosine catabolic pathway. We sequenced *HGD* in 79 unrelated affected individuals and 14 affected siblings. We identified 52 *HGD* variants (22 represented novel variants), including 36 missense, 7 splice site, 6 frameshift, 2 nonsense and a no-stop mutation. Most variants reside in exons 3, 6, 8 and 13. Adding the variants found in our study to all previously reported variants results in a total of 89 potential disease-causing sequence variations in AKU patients. Of these variants, 61 are missense variations, for which we assessed the potential effect on the protein using 5 different bioinformatic tools designed for interpretation of missense variants (SIFT, POLYPHEN, PANTHER, PMUT and SNAP). We also analyzed the potential effect of splice site variants using two additional tools (BDGP and NetGene2). Interestingly, 30 patients (24 probands) appeared homozygous for 16 different variants; this high level of homozygosity cannot be explained by consanguinity levels of the probands. In some patients we suspect hemizygosity, which we are confirming using different approaches, including SNP-array analysis and allele quantification by real-time PCR. A potential HGD deletion may be common due to a founder effect or a because of susceptibility for deletion in this area. Such deletion(s) would explain the high level of homozygous AKU patients found in our study and in previously reported works.

## 65

**Hypophosphatasia: Natural History Of 177 Pediatric Patients.** *D. Wenkert[1], F. Zhang[1], M.C. Benigno[1], J.A. Zerega[1], K.E. Mack[1], V. Lim[1], S.P. Coburn[2], K. Ericson[2], W.H. McAlister[3], S. Mumm[1,4], M.P. Whyte[1,4].* 1) Ctr Metabolic Bone Disease, Shriners Hosp Children, St Louis, MO; 2) Indiana University-Purdue University, Fort Wayne, IN; 3) Mallinckrodt Inst Radiology, Washington University School of Medicine, St Louis, MO; 4) Div Bone and Mineral Diseases, Washington University School of Medicine, St Louis, MO.

Hypophosphatasia (HPP) is the inborn-error-of-metabolism caused by deactivating mutation(s) within the gene that encodes the tissue nonspecific isoenzyme of alkaline phosphatase (TNSALP). Consequently, inorganic pyrophosphate (PPi) and pyridoxal 5'-phosphate (PLP), natural substrates for this ectoenzyme, accumulate extracellularly. Excess PPi blocks skeletal mineralization causing rickets or osteomalacia. Pediatric HPP spans stillbirth (profound skeletal hypomineralization) to premature loss of primary teeth without skeletal disease. During the past 26 years, 177 pediatric HPP patients had 1 or more inpatient evaluations when fasting blood and 24-hour urine collections were obtained while we matched ad libitum dietary calcium intakes. Disturbances in endochondral bone formation were followed radiographically using single views of wrists and knees. We evaluated the following pediatric forms of HPP: 65 odonto-HPP (dental disease only), 99 childhood HPP (41 mild), 12 infantile HPP, and 1 perinatal/infantile HPP. Principal complications identified were premature tooth exfoliation, joint hypermobility, lower extremity malalignment, skeletal pain, muscle weakness, craniosynostosis, chest deformity, scoliosis, clubfoot, and fractures. To date, 52 of the 177 patients had 4-14 admissions (usually occurring once every few years). Serum total ALP activity, analyzed for ages ≤10 yr, correlated with HPP severity; lower values reflected more severe disease, but also showed 'physiological' decrements after puberty. Elevated plasma PLP levels correlated with skeletal disease severity and the number of teeth lost prematurely, but in contrast to serum ALP did not change over time according to a 2-stage statistical analysis. This indicates that plasma PLP is a good diagnostic test that correlates with HPP severity even when only measured once. Height z-scores were without gender differences, reflected HPP severity, and showed no significant changes over time; height assessments at >2 yrs of age predicted adult heights. Radiographic features were highly individualized (physeal widening, irregularity of zones of calcification, "tongues of radiolucency", metaphyseal osteopenia or osteosclerosis, and cortical thinning), but typically changed little during follow-up. Accordingly, we have defined major features of the natural history of HPP which will serve as a basis for assessment of future therapies.

## 66

**Argininosuccinate Lyase is an essential regulator of nitric oxide synthase.** *A. Erez[1], Y. Chen[1], O. Shchelochkov[1], S. Nagamani[1], H. Zeng[2], A. Mian[1], M. Jiang[1], H. Garg[4], M. Summar[2], W. O'Brien[1], J. Marini[3], J. Aschner[2], N. Bryan[4], B. Lee[1,5].* 1) Molecular and Human Genetic Department, Baylor College of Medicine, Houston, TX; 2) Department of Pediatrics and Center for Human Genetics Research, Vanderbilt; 3) Department of Pediatrics/ Nutrition, USDA/ARS Children's Nutrition Research; 4) Brown Foundation Institute of Molecular Medicine, University of Texas, Houston Health Science Center; 5) Howard Hughes Medical Institute, Houston, TX.

Nitric Oxide (NO) is an essential signaling molecule for diverse physiological and disease processes. The regulation of NO flux has focused on the study of the three NO synthases (NOS), but their respective genetic deficiencies exhibit only modest phenotypes. In humans, the urea cycle disorder argininosuccinic aciduria (ASA) caused by deficiency of argininosuccinic acid lyase (ASL) shows systemic features that may reflect global dysregulation of NO homeostasis due to deficiency of intracellular arginine synthesis and/or inability to utilize extracellular arginine hence; ASA patients have decreased biochemical markers and dynamic measures of NO production. In contrast to ASS fibroblasts, ASA patient's fibroblasts cannot utilize extracellular arginine for NO synthesis emphasizing the importance of ASL for NO production. We show that NO synthesis from arginine requires formation of a protein complex that includes HSP90, ASL, ASS, and NOS that is decreased in hypomorphic Asl mice. When ASL is deficient, the hypomorph mice and the ASA patients have increased NOS uncoupling resulting in the production of more free radicals which contribute to the cellular damage. Together, these data show that ASL regulates systemic NO production within a novel complex and helps explain further the unique features seen only in ASA compared to other urea cycle disorders.

## 67

**Development and validation of a disease targeted SNP-based array for clinical personal genomic profiling.** *A.S. Willis[1,2], P. Fang[1,2], M. Strivens[1,2], L. Franco[2], A. Jinnah[2], S. Masri[1,2], R. Pace[1,2], D. Vo[1,2], M. Maheshwari[1,2], J.W. Belmont[2], C.M. Eng[1,2].* 1) Medical Genetics Laboratories, Baylor College of Medicine, Houston, TX; 2) Dept of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Recent advances in high throughput genotyping have made possible large scale analysis of an individual's genomic profile. When properly designed to give information of clear clinical relevance, these platforms have the potential to individualize medical care by providing prior knowledge of future disease risk and predicted response to drug therapies. Towards this goal, the Medical Genetics Laboratories of Baylor College of Medicine have undertaken to design, validate, and implement a custom-designed SNP-based array, using the Illumina Infinium platform, to detect a large number of DNA variants or mutations that increase risk for a number of diseases. The selected mutations and variants are grouped into two categories: mutations associated with adult-onset Mendelian disorders including cardiovascular, neurological, gastrointestinal, and inherited cancer syndromes, and SNPs associated with risk for selected common disorders or with pharmacogenetic implications. Validation of this personal genomic profiling array consisted of testing the array and individual SNP performance in approximately 190 samples including samples with known mutations for selected adult-onset Mendelian disorders, Caucasian CEPH HapMap samples, and samples from the SNP500 panel (Coriell Cell Repository). For the initial phase of the validation and to generate cluster positions for future use, we evaluated the clustering and performance of each individual SNP within Illumina's BeadStudio software. There was complete concordance between the positive control samples with known mutations and the results of SNP analysis. For the Mendelian disorders, most HapMap and SNP500 samples were homozygous for the reference allele; samples located outside the main cluster for a particular SNP were sequenced. 41 mutations were confirmed to be heterozygous or in rare instances homozygous for the change; in addition, 3 outliers identified had a different change at the same or adjacent base. To validate the risk SNPs, we selected a subset of 10 HapMap samples and compared genotypes generated by our array to those published in public databases. The genotypes for 123 risk SNPs were concordant to those available in public databases. In conclusion, our laboratory has designed and validated a SNP-based array for personal genomic profiling targeted to clinically relevant disorders that has the potential to provide patients and their physicians with a resource to personalize medical care.

## 68

**Using lifetime risk estimates in personal genomic profiles: estimation of uncertainty.** *Quanhe. Yang[1], W.Dana Flanders[2], Ramal. Moonesinghe[3], JohnP.A Ioannidis[4,5], Idris. Guessous[2], MuinJ. Khoury[1].* 1) Office of Public Health Genomics, Centers for Disease Control, Atlanta, GA; 2) Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA; 3) Office of Minority Health, Centers for Disease Control and Prevention, Atlanta, GA, USA; 4) Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine and Biomedical Research Institute, Foundation for Research and Technology-Hellas, Ioannina, Greece; 5) Tufts Clinical and Translational Science Institute and Center for Genetic Epidemiology and Modeling, Tufts Medical Center, and Department of Medicine, Tufts University School of Medicine, Boston, Massachusetts, USA.

Background: Personal genomic testing is now offered directly to consumers using genetic variants identified by genome-wide association studies (GWAS) for common diseases. Tests report estimates of age-specific and lifetime risks for various diseases based on genotypes at multiple loci. However, the uncertainty surrounding such risk estimates has not been thoroughly examined. Objectives: Using breast cancer as an example, we examined the combined effect of uncertainties in population incidence rates, genotype frequency, effect sizes, and models of joint effects among genetic variants on the lifetime risk estimates. Design: We performed simulations to estimate the lifetime breast cancer risk for carriers and non-carriers of genetic variants. We derived population-based cancer incident rates from the Surveillance, Epidemiology, and End Results (SEER) program and comparative international data. We used data for non-Hispanic white women from 2003-2005. We derived genotype frequencies from HapMap, and effect sizes from published GWAS and meta-analyses thereof. Results: For a single genetic variant in the FGFR2 gene (rs2981582), the combination of uncertainty in these parameters produced risk estimates where the upper and lower 95% credibility intervals differed by over 4-fold. For a panel of five genetic variants, estimated lifetime risk of developing breast cancer before age 80 for a woman that carried all five risk variants ranged from 5.9% to 20%, depending on the assumptions of additive or multiplicative joint effects and the breast cancer incidence rates. Conclusion: Epidemiologic parameters involved in computation of disease risk have substantial uncertainty, and the cumulative uncertainty should be properly recognized. Reliance on point estimates alone could be seriously misleading.

## 69

**A Template for Molecular Genetic Test Reports for Heritable Conditions: Results from an Expert Panel.** *I. Lubin[1], L. Hilborne[2,3,4], M.T. Scheuner[2,5,6].* 1) Div Lab Systms, Ctr Disease Control/Prevention, Atlanta, GA; 2) RAND Corporation, Health Unit, Santa Monica, CA; 3) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA; 4) Quest Diagnostics, West Hills, CA; 5) VA Greater Los Angeles Healthcare System, Los Angeles, CA; 6) Department of Health Services, UCLA School of Public Health, Los Angeles, CA.

Background: Health professionals lacking formal training in genetics are ordering molecular genetic tests with increasing frequency. Quality of care could be compromised if the clinicians using genetic tests do not properly understand the results. Methods: We convened a meeting of 28 experts representing molecular and clinical genetics, primary care, information technology, the insurance industry and consumers to establish principles, concepts, and terminology important to communicating test results. Their input was based on professional and personal insight and previously published work. A modified Delphi panel process was used to refine development of a molecular genetic test report. This process also informed the development of structured surveys administered by Email that addressed the report format, specific data elements, element grouping, and grouping order. The results of surveys were discussed by phone to reach consensus or achieve a majority opinion. Results: The expert panel agreed that the report should present information in sections that flow in a logical manner, consisting of patient and ordering clinician information, the test ordered, results & interpretation, guidance, and supplemental information. There was lack of consensus about the specific data elements to include, the grouping of elements, and the order of the groupings. Disagreement appeared to arise from divergent opinions relating to the purpose and comprehensiveness of the report. Some wanted to include elements such as family history, ethnicity/race and test indication only when relevant to the test interpretation, while others wanted to consistently present these elements, as relevance could evolve and differ depending on stakeholder. Some preferred a more comprehensive, integrated report that could more readily support management decisions, whereas others wanted a concise representation of the test and result. Despite these differences, a report template acceptable to all was developed. The template will be presented including the order of report sections; the data elements to always, sometimes or never include; and the preferred, acceptable and not acceptable grouping of elements. Conclusions: A molecular genetic test report template designed to improve communication between the laboratory and ordering clinician has been developed. An evaluation of the template is underway that will assess feasibility, acceptability, ease of use and effectiveness.

## 70

**A comparison of electronic, self-directed family history collection to counselor-supplemented family history collection.** *F.M. Facio[1], W.G. Feero[1,2], A. Linn[1,3], L.G. Biesecker[1].* 1) Natl Human Genome Res Inst, NIH, Bethesda, MD; 2) Maine-Dartmouth Family Medicine Residency Program, Augusta, ME; 3) Case Western Reserve University, Cleveland, OH.

Introduction: ClinSeq is a study designed to investigate medical sequencing as a clinical research tool for evaluating risk for common conditions. Participants complete the U.S. Surgeon General's My Family Health Portrait (MFHP), a web-based tool for self-directed collection of family history information for six common disorders (heart disease [HD], stroke [ST], diabetes [DM], and colorectal [CRC], breast [BR], and ovarian [OV] cancers). Family Healthware (FHW) is a research tool created by the CDC for the electronic collection of family history that also provides familial risk categorization (weak, moderate, strong). The purpose of this study was to determine the analytic and clinical validity of patient-entered data into MFHP using the genetic counseling interview as a reference. Methods: ClinSeq participants completed MFHP and were subsequently interviewed by a genetic counselor, who supplemented the MFHP family history. The specificities and sensitivities for 1st and 2nd degree relative family histories were calculated. To measure the clinical validity of MFHP, patient entered data from MFHP and data from the genetic counseling interview were then entered separately into FHW to calculate risk scores for the 6 common conditions. Risk scores were compared using kappa statistics as a measure of correlation. The data was analyzed for sources of inaccuracy of family history data. Results: 150 probands (avg. age 58, F:M 1.3, 96% white) with 888 1st and 2,282 2nd degree relatives were studied. Specificities and sensitivities for 1st degree relatives were for HD (92.0%, 95.7%), ST (99.7%, 98.3%), DM (100%, 97.7%), CRC (99.9%, 90.0%), BR (100%, 96.1%), and OV (99.8%, 100%). Specificities and sensitivities for 2nd degree relatives were for HD (93.9%, 66.7%), ST (99.4%, 82.1%), DM (99.9%, 69.5%), CRC (99.8%, 78.1%), BR (99.9%, 73.5%), and OV (99.7%, 85.7%). Comparison of FHW risk scores derived using MFHP data versus those derived using data from the genetic counseling interview yielded kappa statistics showing substantial agreement, with the exception of HD where agreement was moderate. Sources of error included user mis-categorization of HD cases and omitted relatives. Conclusions: MFHP has a high analytic and clinical validity for collecting family history for five common conditions when compared to MFHP supplemented with a genetic counseling interview. Improving the specificity of diagnostic choices for HD may improve tool performance.

## 71

**Using Family Health History as a Genomic and Clinical Tool in Primary Care.** *N. Arar.* Dept Medicine, Div Nephrology, Univ Texas Health Sci Ctr, San Antonio, TX.

We conducted a cross-sectional, observational study to examine the content and process of patient-provider exchange about FHH during medical encounter. Fifty encounters involving six physicians at a VA primary care outpatient clinic were videotaped in San Antonio, TX. Patients were recruited as part of a convenience sample of patients attending the clinic. The videotapes were coded and analyzed for verbal content and processes regarding patient-provider communication during clinic visits. Content analysis showed that during an average 17-minute outpatient visit, the mean time spent discussing FHH was less than 2 minutes. Patients came to the clinic for a wide variety of reasons, such as diabetes, high blood pressure, cardiovascular diseases, and depression. The mean number of years the patients had seen their providers was 4.8 years (range: 3-5.75; SD = 0.74). We investigated the relationship between the proportion of time providers spent on EMR, FHH discussion and the total number of self-care management (SCM) elements discussed during encounters. FHH discussion included diseases such as cancer, diabetes and CVD. The average number of self-care elements discussed was 6 (sd=2.9; range:2-13). Average number of self-care elements was 7(sd=2.7; range:2-13) when encounters featured use of EMR, compared to 3.4 (sd=1.8; range:2-9) for encounters with low EMR usage (t-test, $p< 0.05$). These findings indicated that patient-provider communication regarding the discussion of self-care management plan was not adequate, though a significant improvement was observed with the EMR-users (t-test, $p< 0.05$). Patient-provider exchanges about SCM were related to discussion regarding FHH information (t-test, $p< 0.05$). Physician awareness of FHH may increase the use of preventive measures and encourage early screening for diabetes. These observations suggest that an easy-to-use tool for systematically providing the FHH has the potential to increase preventive service delivery in primary care. Primary care providers can play a major role in chronic diseases prevention by reviewing their patients' family histories and making recommendations for early detection or intervention strategies and counseling on lifestyle. Developing system-based interventions that will enhance the adoption and delivery of FHH as genomic tool may improve health care for individuals at risk or with common chronic diseases.

## 72

**Newborn screening in Fragile X syndrome: a pilot study.** *K.P. Iong[1], J. Patel[1], R.J. Hagerman[2,3], L.W. Gane[2], P. Sorensen[2], K. Urban[4], C. Hervey[4], E. Berry-Kravis[4], F. Tassone[1,2].* 1) Department of Biochemistry and Molecular Medicine, University of California, School of Medicine, Davis, CA, USA; 2) M.I.N.D. Institute, University of California Davis Medical Center, Sacramento, CA, USA; 3) Department of Pediatrics, University of California, School of Medicine, Davis, CA, USA; 4) Rush University Medical Center, Chicago, IL, USA.

Technological development, new treatments and advocacy efforts are contributing to a push for rapid expansion of newborn screening. However, expanded screening raises a number of concerns and in the case of the FMR1 mutations has been a topic of considerable discussion since the gene was identified. Because phenotypic features are not evident at birth, Fragile X (FXS) must be discerned through abnormalities in development or behavior. The average age of diagnosis is 30-36 months for full mutation males; consequently, children miss the opportunity to participate in early intervention and parents often have additional children with FXS without knowing reproductive risks. Thus, screening for FXS will allow the identification of a greater number of individuals at risk for the disorder or transmitting the disorder. We have recently begun a pilot study of Newborn Screening in FXS aimed to the determination of allele frequencies in the general population and to the assessment of clinical involvement in the wide variety of fragile X-related phenotypes in the primary and extended families of the newborn probands identified by newborn screening. Using our recently developed PCR method for the identification of premutation and full mutation alleles in the FMR1 gene, our preliminary data, based on over 2500 newborn blood spots, indicates that the frequency of occurrence of premutation alleles is greater than that previously reported. We have also started to document the degree of clinical involvement in the newborn proband and the extended family members and our experience will be discussed.

## 73

**GenetiKit: a randomized controlled trial to enhance delivery of genetics services by family physicians.** *J.C. Carroll[1,2], B.J. Wilson[3], J. Allanson[4,5], J. Grimshaw[6], S.M. Blaine[1], W. Meschino[7,8], J. Permaul[6], I.D. Graham[3,6,9].* 1) Department of Family and Community Medicine, University of Toronto, Canada; 2) Granovsky Gluskin Family Medicine Centre, Mount Sinai Hospital, Toronto, Canada; 3) Department of Epidemiology and Community Medicine, University of Ottawa, Canada; 4) Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, Canada; 5) Department of Pediatrics, University of Ottawa, Canada; 6) Department of Medicine and Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa, Canada; 7) Genetics Program, North York General Hospital, Canada; 8) Department of Pediatrics, University of Toronto, Canada; 9) Canadian Institutes of Health Research, Ottawa, Canada.

Context: Patients look to their family physicians (FPs) for credible information and guidance in making informed choices about genetic testing. FPs are willing to play a significant role in delivering genetics services but are challenged by lack of knowledge and the rapid pace of genetic discovery, making it difficult to keep current. There is an urgent need for effective interventions to facilitate the integration of genetics into family medicine. Objective: To determine if a multi-faceted knowledge translation intervention would improve skills and knowledge in primary care genetics including referral decisions, confidence in core genetics competencies and knowledge. Design: Randomized controlled trial Setting: 4 communities in Ontario, Canada: 2 urban, 2 Northern Participants: Family physicians in active practice in study communities Intervention: Interactive educational workshop, portfolio of clinical genetics tools, knowledge service called Gene Messenger. Outcome Measures: Appropriate referral decisions to genetics in response to 10 breast cancer scenarios, decisional conflict, self-reported confidence in 11 genetics core competencies, 3 knowledge questions and questions evaluating the intervention components 6 months post-intervention Results: 125 FPs randomized, 80 (64%) completed the study (32/64 control, 48/61 intervention). Intervention FPs had significantly higher appropriate referral decision scores (6.35/10 control, 7.84/10 intervention, $p<0.001$), self-reported confidence on core genetics competencies (33.9/55 control, 42.6/55 intervention, $p<0.001$) and significantly more correct answers to one of three knowledge questions. The intervention group showed a significant increase in self-reported confidence across all genetics core competency items. Over 90% of FPs wanted to continue receiving Gene Messengers and would recommend them to colleagues. No differences in decisional conflict were detected. Conclusions: This study demonstrated that a complex educational intervention including an interactive workshop, portfolio of genetics tools and practical knowledge service (Gene Messenger), was able to significantly improve practice intent for a variety of clinical genetic scenarios that could be found in primary care, as well as confidence in skills and knowledge in genetics.

## 74

**Putting Knowledge into Practice: Well Informed Physicians Are Twice as Likely to Order or Recommend Pharmacogenetic Testing.** *C.L. Sanders[1], E.J. Stanek[1], J.R. Teagarden[1], B.C. Agatep[1], K.A. Johansen[2], R.E. Aubert[1], M. Khalid[1], A. Patel[1], F.W. Frueh[1], R.S. Epstein[1].* 1) Medco Health Solutions, Inc., Franklin Lakes, NJ; 2) American Medical Association, Chicago, IL.

Background: Numerous factors may influence the adoption of new technology and knowledge, such as pharmacogenomics (PGx), into medical practice. We sought to explore the adoption of PGx testing among physicians who feel adequately informed about its application in practice. Methods: We conducted an anonymous, cross-sectional, survey of US physicians in late 2008. The fax-based survey solicited demographic and professional profile elements, as well as educational background, beliefs, practices, and preferred PGx information resources. Respondent characteristics were compared to the American Medical Association MasterFile to assess generalizability. Associations between physicians' level of PGx testing knowledge versus demographics and practice characteristics, past/future PGx testing, and preferred information sources were assessed by multivariate logistic regression analysis. Results: We surveyed 397,832 physicians and 10,303 (3%) completed surveys were returned. Respondent characteristics were similar to the overall US physician population. While almost all surveyed physicians think that genetic variations influence drug therapy, only 1048 (10.3%) believed they were adequately informed about PGx testing availability and applications. Those who believed that they were adequately informed were more likely to be male, have had prior PGx education, be federal/state government employees, specialize in oncology, and believe that PGx tests mainly benefited patients through reduced drug toxicity. Preferred PGx information sources used by these physicians included drug labels and/or information from genetic testing labs. However, FDA approval/recommendations of PGx testing were less likely to be regarded as important or very important. In practice, physicians who believed that they were adequately informed about PGx were more than twice as likely to have ordered or recommended a PGx test in the past six months or to do so in the future than those who did not report feeling well-informed. Conclusion: Even though almost all US physicians agree that genetic information may play a role in drug therapy, only one in ten physicians report feeling adequately informed about PGx testing. Importantly, they are more than two times more likely to adopt PGx testing, indicating that knowledge is key for translation into clinical practice. The results of this survey may prove helpful in planning and executing initiatives to foster broader adoption of PGx testing.

## 75

**Clinical Genetic Testing for Patients with Autism Spectrum Disorders.** *Y. Shen[1,2,3,4], K. Dies[1,4], I. Holm[1,4,5], J. Picker[1,6], J. Milunsky[1,7], B. Wu[1,2,4], D. Miller[1,2,4,6], Boston Autism Consortium Clinical Genetics/CHB DNA Diagnostics Collaboration.* 1) Autism Consortium, Boston, MA; 2) Dept of Laboratory Medicine, Children's Hospital Boston, Boston, MA; 3) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 4) Harvard Medical School, Boston, MA; 5) Program in Genomics, Children's Hospital Boston, Boston, MA; 6) Division of Genetics, Children's Hospital Boston, Boston, MA; 7) Clinical Genetics, Boston University School of Medicine, Boston, MA.

Multiple lines of evidence indicate a strong genetic contribution to Autism Spectrum Disorders. Current guidelines for clinical genetic testing recommend a G-banded karyotype to detect chromosomal abnormalities and fragile X DNA testing, but guidelines for chromosomal microarray analysis have not been established. To evaluate and compare the diagnostic yield of laboratory genetic tests, a cohort of 933 patients with a clinical diagnosis of autism spectrum disorder (ASD) were subjected for G-banded karyotype, fragile X testing, and chromosomal microarray analysis (CMA). As a result, karyotype identified 19 abnormal cases among 852 patients (2.23%), fragile X testing was abnormal in 2 out of 861 patients (0.23%), and CMA identified potentially significant deletions or duplications in 204 out of 848 patients (24.1%) of which 59/848 (7.0%) were considered clinically significant pathogenic variants or variants of possible significance. CMA results were normal in 8/852 (0.94%) patients with abnormal karyotype due to balanced rearrangements or unidentified marker chromosome. CMA with whole genome coverage and targeted CMA detected clinically relevant copy number variants (CNVs) in 7.3% of patients (51/697) and 5.3% of patients (8/151), respectively, both higher than karyotype. With the exception of recurrent deletion and duplication of chromosome 16p11.2 and 15q13.2q13.3, most copy number changes were unique or identified in only a small subset of patients. As a conclusion, CMA had the highest detection rate among clinically available genetic tests for patients with ASD. Interpretation of microarray data is complicated by the presence of both novel and recurrent copy number variants of unknown significance. Despite these limitations, CMA should be considered as part of the initial diagnostic evaluation of patients with ASD.

## 76

**Utilization of prenatal serum screening versus invasive diagnostic testing for previous pregnancy with Down syndrome.** *N. Nakata, S. Bhatt.* Genzyme Genetics, Los Angeles, CA.

Women with a history of previous pregnancy with autosomal trisomy are at increased risk for fetal aneuploidy in subsequent pregnancies with recurrence risks between 1-2%. Current ACOG and ACMG guidelines recommend that these women be offered prenatal diagnostic testing. We reviewed the testing decisions of patients referred to Genzyme Genetics for genetic counseling during 2008 due to a previous pregnancy with Down syndrome. Data were analyzed from our clinical database in order to evaluate patients' utilization of screening and diagnostic testing. The study included 296 patients; all were offered amniocentesis and 119 (57%) were also offered CVS. Depending upon availability and gestational age, all patients were also offered screening options of first trimester, sequential, integrated and/or quadruple marker screening. For patients <35 years, 30% (67) proceeded with CVS/amniocentesis, 48% (106) declined either procedure, 20% (44) were undecided pending screening results and the remaining 2% (5) were undecided about invasive testing/screening or declined screening. The corresponding decisions for patients ≥35 years revealed 37% (72) choose CVS/amniocentesis, 44% (84) declined either, 17% (33) were undecided pending screening and 2% (4) were undecided about any testing or declined screening. Among patients who declined invasive testing, 65% (69) of the younger group and 38% (32) of the older group opted for screening. Even among patients who proceeded with invasive testing, 21% (14) of the younger group and 13% (9) of the older group had had prior screening. We observed that the majority of our patients considered to be high risk due to a previous pregnancy with Down syndrome chose screening instead of proceeding directly to invasive testing. With the improved detection rates for prenatal serum screening for Down syndrome, a similar trend away from invasive testing has occurred for high risk pregnancies due to AMA. Since the level of recurrence risks for non-translocation Down syndrome is within the range of risks due to AMA, recommendations for screening and invasive testing should be consistent for these two groups. Patients with a previous pregnancy with Down syndrome, similar to AMA patients, can refine their risks for fetal trisomy 21 through prenatal screening prior to deciding whether or not to pursue invasive testing. Collection of pregnancy outcomes should be undertaken to support the efficacy of this approach.

## 77

**First signs of alterations in RNA editing in autistic brains.** *A. Eran[1,2], K. Vatalaro[2], I.S. Kohane[1,3], L.M. Kunkel[2,3,4].* 1) Harvard-MIT Health Sciences and Technology, Cambridge, MA; 2) Program in Genomics, Children's Hospital Boston, Boston, MA; 3) Department of Pediatrics, Harvard Medical School, Boston, MA; 4) Howard Hughes Medical Institute.

Autism spectrum disorders (ASDs) are common neurodevelopmental disorders characterized by a spectrum of social deficits, communication impairments, stereotyped interests, and repetitive behavior. Twin and family studies show that ASD is among the most heritable of complex disorders. However, even with substantial recent advances in autism genetics, the known genetic risk factors explain only 15-20% of ASDs. As RNA mediated regulatory mechanisms gain increasing recognition for their key role in orchestrating brain development, they become plausible candidates for the elusive underlying cause of ASD. Specifically, RNA editing, in the form of site-specific adenosine-to-inosine (A-to-I) base conversions, has been shown to exhibit precise regional specificity in the brain and is essential for normal behavior. A-to-I editing of ionotropic receptors directly affect neurotransmission and editing of miRNAs and their targets alters gene expression in the brain. We hypothesized that RNA editing may be important to the development of autism. To test this hypothesis, we use 454 sequencing to directly interrogate editing sites at a single molecule scale, and to accurately quantify the levels of editing present in autistic and control RNAs, with respect to their genomic sequences. The 454 Genome Sequencer FLX System is used for deep sequencing of small RNAs and select A-to-I edited functional candidate mRNAs isolated from autistic and control cerebella. Preliminary results from 25 brains show significant differences in RNA editing between autistics and controls. Such alterations in upstream regulatory mechanisms could account for the broad phenotypic spectrum and complex inheritance pattern observed in ASD.

## 78

**Genomewide Joint Tests Reveal Associations and Gene by Environment Interactions in Multiplex Autism Families.** *A. Lu[1], R.M. Cantor Chiu[1,2].* 1) Dept Human Genetics, Univ California, Los Angeles, Los Angeles, CA; 2) Center for Neurobehavioral Genetics, Department of Psychiatry, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles.

Joint tests of genetic association (G) with a gene by environment interaction (GxE) have been found to boost statistical power for G, as well as identify significant GxE (Kraft, 2007). This analytic approach is straightforward in the analysis of case and control samples, but more difficult to implement in the analysis of multiplex families. The case/pseudocontrol approach (Cordell and Clayton, 2002)is appropriate for families. It is based on conditional logistic regression and derives the control genotypes from the combinations of the alleles not transmitted to the affected by their parents. The power of this method and the high frequency of Autism in males encouraged us to test for male specific associations on a genomewide level. That is, since Autism is four times more frequent in males than females, some risk alleles are likely to be penetrant, and therefore associated in the male 'environment' while not in the female. This heterogeneity may be viewed as a GxE, and evidence of such a GxE in Autism has been identified by linkage and targeted association studies. To search for Autism G and GxE, we applied the case/pseudocontrol approach to 218,682 tagging SNPs of 483,075 cleaned Illumina HumanHap550 SNPs genotyped in 548 multiplex families (2,236 individuals) with complete nuclear family information from the Autism Genetics Resource Exchange (AGRE)( http://www.agre.org/index.cfm). Using a Bonferroni correction, 6 SNPs were highlighted. However, while all had a significant G, none showed a significant GxE. When the threshold was set less stringently, at 4.57E-06 to allow 1 false positive finding, 11 SNPs passed the criterion. Each was associated with Autism, and 2 of them exhibited a GxE. One of the GxE on Chromosome 9 is within an intron of Dopamine Beta Hydroxylase (DBH), a gene expressed in the brain and implicated in a number of neuropsychiatric disorders. As expected, preferential transmission was seen in the subset of families having only affected boys, but not in those families with affected girls. This powerful joint test that combines the analysis of G and GxE in nuclear families reveals a highly plausible candidate gene for additional studies and exhibits the power of this approach to conduct analyses on a genomewide level.

## 79

**Identification of a novel adenosine A3 receptor (A3AR) variant (Leu90-Val) in autism families: Gain-of-function effects on serotonin transporter trafficking and function.** *N.G. Campbell[1], C.B. Zhu[2], K. Lindler[2], R.D. Blakely[2,3,5], J.S. Sutcliffe[3,4,5].* 1) Neuroscience Graduate Program, Vanderbilt University, Nashville, TN; 2) Department of Pharmacology, Vanderbilt University, Nashville, TN; 3) Department of Psychiatry, Vanderbilt University, Nashville, TN; 4) Department of Molecular Physiology & Biophysics, Vanderbilt University, Nashville, TN; 5) Center for Molecular Neuroscience, Vanderbilt University, Nashville, TN.

Autism is a common neuropsychiatric disorder expressed across a spectrum of abnormalities that includes deficits in reciprocal social interaction and communication, and patterns of repetitive behaviors and restricted interests. Autism affects approximately 1 in 150 individuals, and evidence indicates a highly complex genetic etiology. Recent discoveries of causal and susceptibility variants involving multiple loci in particular pathways and/or protein complexes highlight the potential utility of taking a network-based approach to identify autism genes. Serotonin dysregulation has long been implicated in autism, and rare autism-associated variants in the serotonin transporter (SERT; gene symbol: *SLC6A4*) lead to gains of function involving increased activity and abnormal regulation. We hypothesized that genes for SERT regulatory proteins harbor autism-related genetic variation that leads to molecular endpoints similar to what is seen with SERT variants. Thus, we invoke a network-oriented variant analysis (NOVA) paradigm to identify autism-related variation within the SERT regulatory network. Testing our hypothesis, we screened the adenosine A3 receptor locus (*ADORA3*) for sequence variants, as the A3AR has an important role in modulating SERT to regulate 5-HT clearance at the synapse. Sequencing ADORA3 exons in autism probands identified carriers (3/176) of an undocumented variant (Leu90Val) that was not found in controls (0/281). This nonsynonymous change affects a conserved position in the protein and is predicted to affect the ligand binding site. To address functionality of this variant, we co-expressed variant A3AR with SERT in CHO cells and found that stimulation of SERT activity with a selective A3AR agonist leads to prolonged SERT activation relative to the wild type allele. Extending previous observations of a physical association between A3AR and SERT, we found enhanced interaction for Val90 A3AR and increased SERT surface expression following A3 stimulation. The observed gain of function effects of Val90 *ADORA3* on SERT activity and trafficking mimic those previously described for multiple SERT coding variants and consistent with disrupted 5-HT homeostasis as an important neurochemical alteration in autism.

## 80

**Direct Observations of Elevated De novo Mutation Rates and Excess Rare Deleterious Alleles in Schizophrenia and Autism.** *P. Awadalla[1,2,5], R.A. Myers[1,2,3], F. Cassals[1,2], A. Boyko[3], F. Hamdan[2,5], D. Spiegelman[5], E. Henrion[5], J. Gauthier[5], R. Lafreniere[5], C. Bustamante[4], E.A. Stone[4], P. Drapeau[6], G.A Rouleau[2,5].* 1) Dept. of Pediatrics, University of Montreal, Montreal; 2) CHU Ste Justine, Montreal; 3) Bioinformatics Research Centre, North Carolina State University, Raleigh, North Carolina, USA; 4) Department of Biostatistics, Cornell University, Ithaca, NY, USA; 5) Centre of Excellence in Neuromics of Université de Montréal, Centre Hospitalier de l'Université de Centre Centre of Excellence in Neuromics of Université de Montréal, Centre Hospitalier de l'Université de Montréal, and Department of Medicine,; 6) Department of Pathology and Cell Biology and Groupe de recherche sur le système nerveux central, University of Montreal.

The role of de novo mutations (DNM) in common diseases remains unknown. Nonetheless, the rate of novel deleterious mutations and the strength of selection against novel mutations are critical to the genetic architecture of disease. Discovery of such high impact de novo point mutations requires substantial high-resolution interrogation of partial or complete genomes of families through re-sequencing. We hypothesized that deleterious de novo point mutations may play a role in cases of Autism Spectrum Disorder (ASD) and Schizophrenia (SCZ), both etiologically heterogeneous disorders with significantly reduced reproductive fitness. We present the first deep re-sequencing effort of ~ 0.5 Gigabase (Gb) of DNA from two large cohorts of ASD and SCZ trio families (n=285), and a large cohort of randomly selected families (n=285), to directly observe DNM, and to estimate mutation rates ($\mu$) and selective constraints. A survey of 401 synapse-expressed genes found 28 candidate DNM, 13 of which were cell-line artifacts, but 15 of which were validated in blood DNA. The direct neutral mutation rate is similar to previous indirect estimates but a significant excess of de novo point mutations that are protein-disruptive and highly deleterious occurs in individuals with ASD and SCZ. The rate of spontaneous mutation in schizophrenia patients is 3-fold higher than that for either autism patients or a random cohort. The overall distributions of nucleotide variation indicates a significant excess of rare deleterious alleles relative to random populations. Deleterious DNM were discovered in genes encoding synapse structure proteins and kinesins. We identified four genes, GRIN2B, GRIN2C, GRM8, and CACNA1F, that exhibit extreme excesses of rare missense mutations with a high predicted functional impact in schizophrenia and autism. At GRIN2B we also observed one de novo mutation in a schizophrenia patient. Both de novo mutations and multiple rare segregating mutations are required to explain the complex aetiology of these difficult disorders.This study also emphasizes the need for careful validation of rare mutations. Using DNA from archived cell lines can overestimate mutation rate by at least two-fold.

## 81

**Next generation sequencing of chromosome 5q33 and Xp22 in the PIC schizophrenia sample.** *Z. Deng[1], O.V. Evgrafov[1], H.J. Yoon[1], F.A. Middleton[2], R. Ajman[3], M.T. Pato[1], C.N. Pato[1], J.A. Knowles[1].* 1) Psychiatry & Behavioral Scienc, Univ Southern California, Los Angeles, CA; 2) Department of Psychiatry, Upstate Medical University, Syracuse. NY; 3) Department of Veterans Affairs Medical Center, Washington, DC.

Schizophrenia is a highly heritable disorder and the discovery of susceptibility genes for the disorder has been challenging. Geographically and genetically isolated populations been useful for have played key roles in disease-gene identification in other disroders. We have previously performed both genome-wide linkage association (GWA) scans of the familial form of schizophrenia, from a Portuguese population isolate. These studies have identified several chromosomal regions conferring risk, with the promising loci being located on the chromosome 5q33 (rs1946487, p= 1.7X10-7) and Xp22 (rs6638512, p= 4.6X10-7). We have re-sequenced a portion of these two risk-chromosome regions (200 kb Xp22 and 400 kb 5q33) with Next Generation DNA sequencing technology. The targeted regions (600 kb) were amplified in overlapping long-range PCR amplicons from 10 PIC schizophrenia samples. Seven samples were sequencing with Illumina 1 G Genetic Analyzer with 36 bp single-end sequencing, and three samples with paired-end sequencing (36 bp X 2). We identified both common and rare single nucleotide variants, as well as small insertion or deletion events in these samples. Several deletions (around 2 -4 kb) were shared by several samples. These were tested further in a larger sample set of 100 familial cases and 100 controls, and confirmed that both chromosome 5q33 and Xp22 might be associated with schizophrenia in the PIC sample. Additionally, in a ChIP-Seq experiment using an antibody to RNA Polymerase II, we find a dramatic difference in Pol II binding in exactly in the 5q33 area of interest. This convergence of data from linkage, GWAS, DNA sequencing and now epigenetics has increased our interest in the region.

## 82

**The Psychiatric Genetics Consortium: Genome-wide association analysis of common genetic variation in bipolar disorder in 6875 bipolar case and 9114 European American control samples.** *L.J. Scott for the Psychiatric Genetics Consortium Bipolar Group.* Dept Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI.

Bipolar disorder (BD) is a disabling and often life-threatening disorder that affects ~1% of the population worldwide. The Psychiatric Genetics Consortium (PGC) aims to analyze ~60,000 samples from bipolar disorder, schizophrenia, autism, major depressive disorder, and attention deficit hyperactivity disorder. As part of the PGC we have performed an initial combined analysis of 6,875 bipolar cases and 9,114 bipolar controls from 10 European ancestry samples sets (Bonn, Dublin, Edinburgh, GAIN, GSK, NIMH/Pritzker, WTCCC, Step-BD, TOP, UCL, WTCCC) using non-overlapping genome-wide genotyped and imputed data from each sample. We included study sample and the top 5 multi-dimensional scaling components as covariates. The ANK3 locus, identified by Ferreira et al. (2008), showed genome-wide significance in the combined sample (including the Ferreira et al. samples) with the strongest association at rs10994359 (genomic controlled p-value = $7.8 \times 10^{-9}$). Ankyrin G, the protein encoded by ANK3, is a cytoskeletal linker protein enriched at axon initial segments and nodes of Ranvier, where it anchors membrane channels, such as the voltage-sensitive sodium channel and cell adhesion molecules such as Neurofascin and NrCAM. Of the five next most strongly associated loci, at least one contains a gene involved in brain development. These results further confirm the ANK3 locus as a bipolar disorder associated locus.

## 83

**Comprehensive GWAS follow-up study for nicotine dependence replicates association with nicotinic receptors and implicates several novel genes.** *A. Goate[1], V. Stevens[2], J.C. Wang[1], N. Saccone[3], S. Saccone[1], A. Hinrichs[1], L. Fox[1], J. Rice[1], N. Breslau[4], E. Johnson[5], L. Bierut[1].* 1) Psychiatry, Washington Univ Sch Med, St Louis, MO; 2) Department of Epidemiology, American Cancer Society, Atlanta, GA; 3) Department of Genetics, Washington Univ Sch Med, St Louis, MO; 4) Department of Epidemiology, Michigan State University, East Lansing, MI; 5) Research Triangle Institute International, Research Triangle Park, NC.

Smoking has detrimental effects on physical health, increasing risk for cancer, heart disease, stroke, and chronic lung disease. Evidence from twin studies indicates that genetic factors contribute to the development of smoking, smoking persistence, and nicotine dependence. Our group performed the first genome-wide association study (GWAS) and candidate gene study of nicotine dependence using a case control sample of unrelated individuals. The most compelling evidence for genetic factors influencing risk for nicotine dependence was within the CHRNA5-CHRNA3-CHRNB4 gene cluster at chromosome 15q24-25.1. This finding has been replicated in several recent GWAS. In this study, we performed a comprehensive survey to replicate our original GWAS in an independent case control series. The replication dataset consists of 1386 light smokers and 1458 heavy smokers who were participants in the American Cancer Society CPS-II Nutrition Cohort. Individuals who reported smoking at least 30 cigarettes per day for at least five years were defined as heavy smokers. Light smokers were defined as individuals who reported smoking for at least one year during their lifetime and in 1982 and 1992, reported always smoking fewer than 10 cigarettes per day. A total of 1760 single nucleotide polymorphisms (SNPs) were genotyped including 255 SNPs for population stratification. Using logistic regression analysis we confirmed the association of 69/1505 SNPs genotyped with nicotine dependence in the replication dataset. The top 19 SNPs in this series ($1.11 \times 10\text{-}8 \leq p \leq 8.38 \times 10\text{-}6$) are located in the chromosome 15q24-25.1 region within and flanking the CHRNA5-CHRNA3-CHRNB4 gene cluster and represent at least two uncorrelated findings. These SNPs ranked by p-value from 12-817 in the original GWAS. In addition, 24 SNPs in other chromosomal regions showed consistent association with nicotine dependence in both datasets. Twenty of these SNPs are in genes including ADCY5, FBXL-17, FMO1, GABRR1, which have all previously been linked to substance dependence through functional and /or genetic studies. Among the 20 replicated SNPs the rank (by p-value) in the original GWAS ranged from 10-895. Only 7 of the 43 replicated SNPs were in the top 100 SNPs from the original GWAS, underlining the importance of carrying out extensive follow up genotyping in replication studies.

## 84

**Variants in *CHRNA5-CHRNA3-CHRNB4* and other nicotinic receptor subunit genes affect risk for nicotine dependence in African-Americans as well as in European-Americans.** *N.L. Saccone[1], J.C. Wang[2], N. Breslau[3], E.O. Johnson[4], D. Hatsukami[5], S.F. Saccone[2], R.A. Grucza[2], L. Sun[2], W. Duan[1], T.-H. An[6], J. Budde[2], R.C. Culverhouse[7], L. Fox[2], A.L. Hinrichs[2], J.H. Steinbach[8], M. Wu[6], J.P. Rice[1,2], A.M. Goate[1,2], L.J. Bierut[2].* 1) Dept of Genetics, Division of Human Genetics, Washington University School of Medicine, St. Louis, MO; 2) Dept of Psychiatry, Washington University School of Medicine, St. Louis, MO; 3) Dept of Epidemiology, Michigan State University, East Lansing, MI; 4) Research Triangle Institute International, Research Triangle Park, NC; 5) Dept of Psychiatry, University of Minnesota, Minneapolis, MN; 6) Divison of Biology and Biomedical Sciences, Washington University, St. Louis, MO; 7) Dept of Medicine, Washington University School of Medicine, St. Louis, MO; 8) Dept of Anesthesiology, Washington University School of Medicine, St. Louis, MO.

Multiple independent studies of populations of European descent have shown that the *CHRNA5-CHRNA3-CHRNB4* gene cluster on chromosome 15q harbors variants associated with risk for nicotine dependence, smoking, lung cancer, and chronic obstructive pulmonary disease. Other cholinergic nicotinic receptor subunit (*CHRN*) genes have been implicated in nicotine dependence in a European-ancestry sample. We have now genotyped SNPs covering these genes in both European-Americans (EAs) (N=2062) and African-Americans (AAs) (N=710) recruited by the Collaborative Genetic Study of Nicotine Dependence. Cases are nicotine-dependent and controls are non-dependent smokers. Our primary analysis uses the full sample of both EAs and AAs with gender and race as covariates. This analysis will highlight SNPs that show consistent effects across the two populations. Within the chromosome 15q24-25.1 region, we tested 76 SNPs. Of these, the non-synonymous *CHRNA5* SNP rs16969968 is the most significantly associated with nicotine dependence in the full sample of 2772 subjects ($p=4.5x10^{-8}$, OR = 1.42 (1.25-1.61)) as well as in AAs only (p = 0.015, OR = 2.04 (1.15-3.62) and EAs only ($p = 4.1x10^{-7}$, OR = 1.40 (1.23-1.59)). Other SNPs in this region that have been shown to affect mRNA levels of *CHRNA5* in EAs are associated with nicotine dependence in AAs but not in EAs. Outside the chromosome 15q24-25.1 region, variants in or near *CHRNB1*, *CHRNB3*, and *CHRND* show consistent evidence of association with nicotine dependence in African-Americans and in European-Americans. In analysis of the AA sample alone, p-values are < 0.1 and odds ratios are consistent with those in the larger EA sample. Our data demonstrate that rs16969968 - a confirmed, highly replicated risk variant in European-descent populations - is also associated with risk for nicotine dependence in African-Americans. The risk allele is the same in both European- and African-Americans, and the odds ratios in the two groups are not significantly different in a formal heterogeneity test. Because of the different population history in African-Americans, this result is important. In African-Americans, this allele has a markedly lower allele frequency and the surrounding linkage disequilibrium patterns differ from those in European-Americans, and yet it is still associated with risk. Our data also indicate that additional *CHRN* genes harbor variants that are consistently associated with nicotine dependence risk in both populations.

## 85

**Acyl-CoA synthetase long-chain family member 6 (ACSL6) and nicotine dependence.** *J. Chen[1], D. Brunzell[2], K. Jackson[1, 2], K. Kendler[1], X. Chen[1].* 1) Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, 800 East Leigh Street, Biotech I Suite 110 Richmond VA 23219-1534; 2) Department of Pharmacology and Toxicology, Virginia Commonwealth University, 410 North 12th Street, Robert Blackwell Smith Building, Richmond, VA 23298-0613.

It is well known that schizophrenia is highly comorbid with tobacco smoking and nicotine dependence; however, the underlying mechanism for this comorbidity is not clear. One hypothesis is that schizophrenia and nicotine dependence share some common liability genes. This study was designed to test this hypothesis using the acyl-CoA synthetase long-chain family member 6 (ACSL6) gene that was reported to be associated with schizophrenia in our previous studies. In human association studies (n = 2139), we tested 9 tagged SNPs in the ACSL6 region. Two SNPsin the ACSL6 gene, rs667437 and rs477086, were significantly associated with nicotine dependence (p = 0.0022 and 0.0066 respectively) and the risk alleles were the same as shown for schizophrenia. We next used in vivo and in vitro assays to test whether nicotine regulates ACSL6 expression. Real-time PCR and Western blot revealed that both ACSL6 mRNA and protein were dose-dependently increased in rat cortical cultures following acute exposure to nicotine. Chronic exposure to nicotine also regulated levels of ACSL6 expression in the hippocampus and ventral tegmental area of mice. The chronic effects of nicotine on ACSL6 were suppressed by the administration of the nicotinic antagonist, mecamylamine, suggesting that nicotine-associated increases of ACSL6 protein are specific to nicotine exposure. Together these data suggest that ACSL6 expression is altered by nicotine in brain areas associated with tobacco addiction. The fact that the same alleles at ACSL6 were associated with increased risk for both schizophrenia and nicotine dependence suggests that ACSL6 may be a shared liability gene for schizophrenia and nicotine dependence.

## 86

**The Gamma-Aminobutyric Acid Receptor Genes Are Strongly Associated With Drug Dependece.** *D. Li[1], H. Zhao[2,3].* 1) Department of Psychiatry, School of Medicine, Yale University, New Haven, CT 06511, USA; 2) Department of Epidemiology and Public Health, School of Medicine, Yale University, New Haven, CT 06511, USA; 3) Department of Genetics, School of Medicine, Yale University, New Haven, CT 06511, USA.

Gamma-Aminobutyric Acid (GABA) is the major inhibitory neurotransmitter in mammalian brain. Studies support the hypothesis that GABA receptor subunit genes are involved in a number of common disorders including drug dependence, which is an important public-health and social problem. Since the first positive association study, a number of subsequent independent studies have attempted to replicate the association, with some being successful. Furthermore, no obviously functional or pathogenic variants have been identified, and the relationship between the subunit genes and drug dependence has remained inconclusive. To reconcile these conflicting findings and to give a comprehensive picture of the genetic architecture of these important genes, we have performed a systematic meta-analysis of the association studies of the GABA receptor genes (*GABRB2*, *GABRA6*, *GABRA1*, and *GABRG2*) on chromosome 5q and the *GABRA2* gene on chromosome 4p12 based on 2814 cases with alcohol, heroin, or methamphetamine dependence and 2637 controls published up to June 2009. Strong evidence of association was found between the *GABRA6*, *GABRG2*, and *GABRA2* genes and alcohol, heroin, and methamphetamine dependence (P = 0.0005 for rs3219151, P = 0.002 for rs211014 P = 0.0003 for rs279858 after Bonferroni correction for multiple testing, respectively). To our knowledge, this is the first meta-analysis study between the gamma-aminobutyric acid receptor genes and drug dependence.

## 87

**New loci on chromosomes 3q25 and 3q21 are associated with size at birth.** *R.M. Freathy[1], U. Sovio[2], I. Prokopenko[3], D.O. Mook-Kanamori[4], D. Berry[5], N.J. Timpson[6], J.J. Hottenga[7], L. Coin[2], A. Hofman[4], E. Hypponen[5], D.M. Evans[6], G. Willemsen[7], T.M. Frayling[1], M.R. Jarvelin[2,8], M.I. McCarthy[3], V. Jaddoe[4], C. Power[5], G. Davey Smith[6], D.I. Boomsma[7] for the Early Growth Genetics Consortium.* 1) Genetics of Complex Traits, Peninsula Medical School, Exeter, UK; 2) Department of Epidemiology and Public Health, Imperial College London, London, UK; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 4) Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 5) MRC Centre of Epidemiology for Child Health, University College London Institute of Child Health, London, UK; 6) MRC Centre for Causal Analyses in Translational Epidemiology, University of Bristol, Bristol, UK; 7) Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands; 8) Institute of Health, University of Oulu, Oulu, Finland.

Identifying genes involved in birth outcomes presents some particular challenges. Notably, the maternal intrauterine environment is often considered to be of primary importance, while two genomes (maternal and fetal) may interact with it, and with one another, to influence any given trait. Despite this, it is important to understand the etiology of fetal growth. High and low birth weights are associated clinically with considerable perinatal risk and epidemiologically with later life chronic disease. The identification of such genes is likely to provide important insights into fetal growth processes. We analyzed genome-wide association data from 10623 participants, born of singleton pregnancies at ≥37 weeks' gestation, from 5 European studies: Avon Longitudinal Study of Parents and Children (N=1418); Northern Finland 1966 Birth Cohort (N=4333); 1958 British Birth Cohort (N=3264); Netherlands Twin Registry (N=414); Generation R (N=1194). After quality control, a total of 2427548 directly-genotyped and imputed SNPs were available for meta-analysis. We tested the association of each SNP with birth weight, assuming an additive model and adjusting for sex and gestational age. Common variants on chromosome 3q25, between *LEKR1* and *CCNL1*, were strongly associated with birth weight at genome-wide significance ($P$=1x10$^{-10}$), with an estimated per-allele effect size of 0.09SD (approx. 90g difference between the two homozygous groups). However, further work is needed to establish the causal gene. A second locus on chromosome 3q21 also provided strong evidence of association ($P$=8x10$^{-8}$), with a similar effect size. Our study provides evidence that genetic factors contribute to normal variation in birth weight. A difference in birth weight of 90g between homozygous groups is equivalent to the effect of a mother smoking 3 cigarettes per day in the third trimester of pregnancy. Replication efforts are underway, and further studies will be necessary to assess the role of maternal genotype at these loci and to investigate potential maternal-fetal genotype interactions.

## 88

**The identification of over 135 loci involved in adult height variation provides important insights into the contribution of common variation to a model complex trait.** *H. Lango Allen[1], G. Lettre[2], K. Estrada[3], S.I. Berndt[4], M.N. Weedon[1], G.R. Abecasis[5], M. Boehnke[6], C. Gieger[7], D. Gudbjartsson[8], N.L. Heard-Costa[9], A.U. Jackson[6], M.I. McCarthy[10], F. Rivadeneira[3], A. Smith[11], N. Soranzo[12], A.G. Uitterlinden[3], T.M. Frayling[1], J.N. Hirschhorn[13], The GIANT Consortium.* 1) Genetics of Complex Traits, Peninsula Medical School, Exeter, United Kingdom; 2) Montreal Heart Institute (Research Center), Université de Montréal, Montréal, Québec, Canada; 3) Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands; 4) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA; 5) Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA; 6) Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA; 7) Institute of Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany; 8) deCODE genetics, Reykjavik, Iceland; 9) Department of Neurology, Boston University School of Medicine, Boston, Massachusetts, USA; 10) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 11) Icelandic Heart Association, Kopavogur, Iceland; 12) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 13) Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Boston, Massachusetts, USA.

Genome-wide association studies (GWAS) have identified >200 variants involved in common diseases and traits, but the biological implications of these findings have recently been questioned. As part of the Genetic Investigation of ANthropometric Traits (GIANT) Consortium, we used adult height as a model of a highly heritable, polygenic trait and meta-analysed summary association statistics from 2.8 million SNPs from 50 individual GWAS, including a total of 133,800 individuals of European ancestry. We tested the extent to which SNPs associated with height clustered within loci, and the clustering of implicated genes within pathways. Preliminary analyses identified 318 independent signals ($P<5x10^{-8}$, HapMap $r^2<0.05$, 1Mb window) in ~210 distinct genomic loci (1Mb window). Conservatively adjusting for the overall GC of 1.42 reduced this to 164 signals in 136 loci. The 318 signals explain ~14% of height variation, based on 4162 non-discovery samples. Many genes at the associated loci are excellent candidates for influencing height, including unreported variants in the growth hormone secretagogue receptor (*GHSR*, P=1x10$^{-13}$), insulin-like growth factor receptor (*IGF1R*, P=1x10$^{-12}$), and two independent signals ($r^2<0.01$) in the estrogen receptor 1 (*ESR1*, P=1x10$^{-9}$ and P=8x10$^{-9}$) genes. In contrast, other loci implicate genes with no obvious role in development, and represent novel insights into the biology of normal growth. Using the text-mining tool GRAIL on genes located within the recombination hotspot boundaries of the associated signals, we found a highly significant enrichment of biologically connected genes, with "growth" and "development" amongst the top keywords (18 had a GRAIL P<0.001, whereas <2 were expected by chance), suggesting that the height loci identified are not randomly placed in the genome. Furthermore, a comparison with the OMIM database showed that at 8% of the signals, the nearest gene, when mutated, causes a human monogenic skeletal development or stature disorder, compared to <2% expected by chance. In conclusion, we have identified many new loci influencing adult height. Many of the implicated genes cluster within biological pathways and processes relevant to human growth and development. The clustering of variants at loci suggests that allelic heterogeneity may be integral to the genetic architecture of polygenic traits. Our data suggest that GWAS meta-analyses of >100,000 individuals will be useful in understanding complex human traits.

## 89

**Chromosome 9q31.2 locus is associated with onset of puberty in both males and females.** *A. Murray[1], J. Perry[1], L. Stolk[2,3,4], N. Franceschini[5], K.L. Lunetta[6,7], G. Zhai[8], P.F. McArdle[9], A.V. Smith[10], E. Streeten[9], T.B. Harris[11], T.D. Spector[8], E. Demerath[12], A.G Uitterlinden[2,3,4], J. Murabito[6,13], D. Lawlor[14], K. Northstone[14], M. Maisonet[15,16], M. McGeehin[15], C. Rubin[15], M. Marcus[15,16].* 1) Genetics of Complex Traits, Peninsula Medical School, Exeter, UK; 2) Department of Internal Medicine, Erasmus MC, Rotterdam, the Netherlands; 3) Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands; 4) Member of the NGI (Netherlands Genomics Initiative)-sponsored Netherlands Consortium for Healthy Aging (NGI-NCHA); 5) Department of Epidemiology, University of North Carolina Gillings School of Global Public Health, Chapel Hill, NC, USA; 6) The National Heart Lung and Blood Institute's Framingham Heart Study, Framingham, MA, USA; 7) Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; 8) Department of Twin Research & Genetic Epidemiology, King's College London, UK; 9) Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, MD, USA; 10) Icelandic Heart Association, Kopavogur, Iceland; 11) Laboratory of Epidemiology, Demography, and Biometry, Intramural Research Program, National Institute on Aging, Bethesda, Maryland, USA; 12) Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA; 13) Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA; 14) Department of Social Medicine, University of Bristol, Bristol UK; 15) National Center for Environmental health, Centers for Disease Control and Prevention, Atlanta, GA, USA; 16) Epidemiology Department, Rollins School of Public Health, Emory University, Atlanta, GA, USA.

The timing of the start of female reproductive life influences growth and health in later life. Twin and family studies suggest a significant genetic component determining age at menarche, with at least 50% heritability. Recent genome-wide association studies have identified 2 loci associated with age at menarche. Four independent studies identified variation in the LIN28B gene on chromosome 6q21 to be associated with a per allele effect on age at menarche of 5 weeks. A second locus at 9q31.2, but no obvious candidate gene, was identified by 2 studies, including our own. Both loci have previously been associated with variation in adult height in males and females and are therefore likely to be genes regulating pubertal development in both sexes. An association between LIN28B at 6q21 and male pubertal development has been described. We aimed to explore the role of variants associated with age at menarche, in puberty and growth in both males and females. We genotyped the rs2090409 SNP at 9q31.2 in two independent population based UK cohorts: ~3,500 women from the British Women's Health and Heart Study and ~7,000 women plus ~8,000 of their children from the population based ALSPAC study of pregnant women and their offspring. Tanner staging of puberty in ALSPAC children was determined by questionnaire on 5 separate occasions, between ages 8 and 13. The 9q31.2 locus was identified in 2 independent GWAS, both with similar effect sizes and p values (effect = 0.09 and 0.1 years per allele; p=3.4 x10-9 and 1.7 x 10-9). Combining replication data with discovery GWAS gave a meta-analysis p value = 2 x10-13 for association of rs2090409 with age at menarche, with effect sizes similar to the discovery set (0.048-0.1 years per allele). The SNP was associated with more advanced Tanner staging for pubic hair growth in boys (p=0.000389, N=1453) and girls (p=0.019, N=2053), genital development in boys (p=0.015, N= 1321) and earlier development of breasts in girls (p=0.007, N=2171). These data provide robust evidence confirming that the 9q31.2 locus is associated with age at menarche, with effect sizes in independent populations of a similar magnitude to that seen in the discovery cohorts. In addition to the association with age at menarche, the 9q31.2 locus is also associated with onset of puberty in both sexes and the association with adult height reported for this locus is likely to be explained by variation in pubertal timing.

## 90

**Genome-wide association meta-analysis of waist-to-hip ratio (WHR) across 35 studies comprising over 74,000 individuals identifies common variants influencing body fat distribution.** *J.C. Randall[1,15], V. Steinthorsdottir[2], G. Thorleifsson[2], L. Qi[3], R. Magi[1,16], E.K. Speliotes[4,17], I.M. Heid[5,18], T. Winkler[5], T.L. Assimes[6], A.U. Jackson[7], M.C. Zillikens[8], T.B. Harris[9], L.A. Cupples[10], I. Barroso[11], R.J.F. Loos[12], K.L. Mohlke[13], C.S. Fox[14], M.I. McCarthy[1,16], C.M. Lindgren[1,16], GIANT (Genetic Investigation of ANthropometric Traits) Consortium.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) deCODE genetics Inc, 101 Reykjavik, Iceland; 3) Departments of Nutrition and Epidemiology, Harvard School of Public Health, Boston, MA; 4) Department of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 5) Institute of Epidemiology, University of Regensburg, Germany; 6) Department of Medicine, Stanford University School of Medicine, Stanford, CA; 7) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 8) Department of Internal Medicine, Erasmus Medical Centre, Rotterdam, the Netherlands; 9) National Institute on Aging, Bethesda, MD; 10) Boston University Schools of Medicine and Public Health, Boston, MA; 11) Metabolic Disease Group, Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 12) Medical Research Council Epidemiology Unit, Strangeways Research Laboratories, Cambridge, United Kingdom; 13) Department of Genetics, University of North Carolina, Chapel Hill, NC; 14) The National Heart Lung and Blood Institute's Framingham Heart Study, Framingham, MA; 15) Department of Statistics, University of Oxford, Oxford, United Kingdom; 16) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Oxford, United Kingdom; 17) Broad Institute at MIT and Harvard, Boston, MA; 18) Institute of Epidemiology, Helmholtz Zentrum München, Germany.

Obesity is an increasing public health issue, but not all forms of obesity carry the same risk. Those with high waist-to-hip ratio (WHR) - or those who tend to be more apple than pear shaped - have an increased risk for developing obesity-related complications, including diabetes, hypertension, heart disease, stroke, certain cancers, and even death. WHR is used as the primary measure of central obesity in the World Health Organization's definition of metabolic syndrome. Central obesity is highly heritable (heritability~50%), but the genetic variants that predispose to it have not yet been well-characterized. To increase power to detect common variants with modest effect sizes, we performed a meta-analysis of 35 genome-wide association studies comprising over 74,000 individuals of European ancestry that contributed data on anthropometric measures as part of the GIANT consortium. WHR data was inverse-normal transformed separately in each gender before performing gender-stratified association testing of ~2.8 million imputed and genotyped SNPs using an additive model with age and BMI as covariates. Our preliminary analysis has identified 11 independent loci associated with central obesity at levels of genome-wide significance ($p<5x10^{-8}$). Among these, we confirm a previously-reported locus between *SLC30A10* and *LYPLAL1* (1q41; $p=3.5x10^{-14}$), and identify 10 novel associations, including *RSPO3* (6q22; $p=1.6x10^{-16}$), *TBX15* (1p11; $p=2.4x10^{-15}$), near *VEGFA* (6p12; $p=7.0x10^{-11}$), and *DNM3* (1q24; $p=1.6x10^{-9}$). *RSPO3* may regulate embryonic vasculogenesis, angiogenesis, and tumor genesis in certain cancers. *TBX15* is differentially expressed between subcutaneous and visceral fat and its expression is correlated with WHR. *VEGFA* is a growth factor that may play a role in diabetic nephropathy and retinopathy, as well as some cancers. None of the 11 loci are significantly associated with height, weight, or BMI in our analyses of those traits of similar sample size, highlighting the unique nature of the central fat depot. Interestingly, a second independent locus at *DNM3* is significantly associated with height, but not WHR in our data (rs678962; p=0.49; $r^2=0.03$). These results confirm the previously reported WHR association at common variants and identify a number of novel associations for central obesity. Many of the identified loci may also play a role in disease susceptibility, possibly through their effect on central fat distribution, though this is an area for further study.

## 91

**META-ANALYSIS OF GENOME-WIDE ASSOCIATION DATA FROM OVER 126,000 INDIVIDUALS REVEALS 18 INDEPENDENT LOCI THAT ASSOCIATE WITH BODY MASS INDEX.** *E.K. Speliotes[1,2], G. Þorleifsson[3], C.J. Willer[4], E. Ingelsson[5], S.I. Berndt[6], K.L. Monda[7], J. Luan[8], S. Vedantam[9], L. Qi[10], C.M. Lindgren[11], I.M. Heid[12], H. Stringham[4], K.E. North[7,13], M. Boehnke[4], U. Thorsteinsdottir[3], I. Barroso[14], G.R. Abecasis[15], M.I. McCarthy[11], J.N. Hirschhorn[16], R.J.F. Loos[8], the GIANT Consortium.* 1) Dept Gastroenterology, Massachusetts Gen Hosp, Boston, MA; 2) Broad Institute at MIT and Harvard, Boston MA 02115, USA; 3) deCODE Genetics, Reykjavik, Iceland; 4) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA; 5) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 6) Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, Maryland; 7) Department of Epidemiology University of North Carolina at Chapel Hill Chapel Hill,NC 27516-3997, USA; 8) MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK; 9) Division of Genetics, Children's Hospital, Boston, MA 02115, USA; 10) Department of Nutrition, Harvard School of Public Health, Boston, MA 02115, USA; 11) Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK, Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Oxford OX3 7LJ, UK; 12) Institute of Epidemiology,University of Regensburg and Institute of Epidemiology, Helmholtz Zentrum München, Germany; 13) Carolina Center for Genome Sciences University of North Carolina at Chapel Hill Chapel Hill, NC 27516, USA; 14) Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK; 15) Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan; 16) Department of Genetics, Harvard Medical School; Department of Pediatrics, Harvard Medical School; Division of Endocrinology, Children's Hospital, Boston, MA 02115, USA.

Obesity and its complications have reached epidemic proportions but few medical treatments exist. Anthropometric measures of obesity such as body mass index (BMI) are heritable and widely measured, and can be used to study the genetic basis of obesity. The Genetic Investigation of ANthropometric Traits (GIANT) consortium conducted a meta-analysis of genome-wide association results between BMI and ~2.8 million imputed or genotyped SNPs in over 126,000 adult men and women of European ancestry from 48 different cohorts. After using genomic control correction on individual studies and again after combining studies, we identified SNPs near 18 independent loci where the association with BMI exceeded genome wide significance ($p < 5 \times 10\text{-}8$); 6 of these are novel. Genes adjacent to the novel associated variants are important in the central nervous system, in retinoic acid biology, and in apoptosis (FAIM2, Fas apoptotic inhibitory molecule 2 ; MAP2K5, mitogen-activated protein kinase 5; GPRC5B, G-protein coupled receptor family C group 5 member B precursor or retinoic acid-induced gene 2 protein); some of the other genes have unknown functions. We also confirm 12 loci that we previously showed to be associated with anthropometric measures of obesity (FTO, TMEM18, MC4R, GNPDA2, NEGR1, ETV5, BDNF, LTZR2, TFAP2B, MTCH2, SH2B1, and KCTD15) and provide strong, albeit not genome wide significant replication for previously reported associations near PCSK1 ($p = 4.8 \times 10\text{-}5$) and SCG3 ($p = 2.8 \times 10\text{-}6$) and LMX1B ($p = 2.8 \times 10\text{-}6$). An additional 16 loci had meta-analysis p-values $< 1 \times 10\text{-}6$, a threshold at which only one false positive would be expected by chance. Many of these are likely to represent true associations with BMI. These loci include SNPs near NRXN3 ($p = 1.9 \times 10\text{-}7$), which has strong prior evidence for association, as well as some intriguing candidate genes (HNF4G, hepatocyte nuclear factor 4G; HTR1A, serotonin receptor 1A). We are assessing these potential associations in additional samples. Associated SNPs do not show heterogeneity across studies and affect weight more than height. Through the work of this consortium, we have greatly expanded the list of genetic loci that are associated with BMI. These new loci enhance our knowledge of underlying biology, open new avenues for follow-up functional and genetic research, and may provide guidance in developing or improving therapies or interventions for obesity.

## 92

**European Lactase Persistence Allele is Associated with Increase in Body Mass Index.** *J. Kettunen[1,2,3], K. Silander[2,3], O. Saarela[2], V. Anttila[1], J. Laitinen[4], A.-L. Hartikainen[5], A. Pouta[5,6], P. Lahermo[7], S. Männistö[2], A. Jula[2], J. Virtamo[2], V. Salomaa[2], T. Lehtimäki[8], O. Raitakari[9], G. Davey Smith[10], M.I. McCarthy[11,12], M.-R. Järvelin[13,14,15,16], M. Perola[2,3,17], L. Peltonen[1,2,3,17,18].* 1) Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) Dept of Chronic Disease Prevention, National Institute for Health and Welfare Helsinki, Finland; 3) FIMM, Institute for Molecular Medicine Finland, University of Helsinki, Finland; 4) Finnish National Institute of Occupational Health, Oulu, Finland; 5) Department of Clinical Sciences/ Obstetrics and Gynecology, University of Oulu, Finland; 6) National Public Health Institute and University of Oulu, Finland; 7) Finnish Genome Center, Institute for Molecular Medicine Finland, University of Helsinki, Finland; 8) Department of Clinical Chemistry, Tampere University Hospital and University of Tampere, Finland; 9) Department of Clinical Physiology, University of Turku and Turku University Central Hospital, Finland; 10) MRC Centre of Causal Analyses in Translational Epidemiology, University of Bristol, United Kingdom; 11) Oxford Centre for Diabetes, Endocrinology and Metabolism, Headington, Oxford, United Kingdom; 12) The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, United Kingdom; 13) Department of Epidemiology and Public Health, Imperial College London, United Kingdom; 14) Institute of Health Sciences, University of Oulu, Finland; 15) Department of Child and Adolescent Health, National Public Health Institute, Finland; 16) Biocenter Oulu, University of Oulu, Finland; 17) Department of Medical Genetics, University of Helsinki, Finland; 18) The Broad Institute of MIT and Harvard, Boston, MA , United States of America.

The global prevalence of obesity has increased significantly in the recent decades, mainly due to excess energy intake and increasingly sedentary lifestyle. Here, we test the association between obesity measured by body mass index (BMI) and one of the best known genetic variants showing evidence for strong selective pressure: the functional variant in the cis-regulatory element of the lactase gene. We tested this variant since it is presumed to provide nutritional advantage in specific physical and cultural environments. We genetically defined lactase persistence (LP) in 17 374 randomly selected Finnish population samples by genotyping the European LP variant (rs4988235). We performed a single test and found that the variant responsible for LP among Europeans was significantly associated with higher BMI ($n = 17\,374$, $p = 1.5*10^{-5}$). The difference between lactase persistent and non-persistent individuals was 0.4 kg/m$^2$. LP correlated with bioelectrical impedance measured fat mass but not with muscle mass or bone mass. Since this locus has been shown to be prone to population stratification we paid special attention to reveal any population substructure which might be responsible for the association signal. We had genome wide SNP data available for a population -wide subsamples of the genotyped individuals. We showed with principal components analysis in the genome wide SNP data that the association was not due to stratification. There are at least three explanations why this variant has not been detected in the published GWAS. Firstly, the variant is not on the currently used SNP arrays. Secondly, the studies have used only the additive model which has also reduced the power to detect the variants affecting in dominant model. Thirdly, the effect size of this variant may be inflated in Finland where the dairy product consumption is very common and there is a correlation with liquid dairy product consumption and the LP variant. Association with this variant and milk consumption has also been shown in other European countries, including Estonia and Austria. The variant explains a comparable amount of BMI variance and has a comparable effect size to that of *FTO* in our study sample. Although explanation at a functional level are a matter for future studies, we propose that the European variant behind lactase persistence, and perhaps other variants affecting the regulatory region of *LCT*, contributes to human obesity.

## 93

**Genome-wide analysis identifies alpha-2 catenin (CTNNA2) as a risk factor for diabetes complicating cystic fibrosis (CF).** *S.M. Blackman[1,2], S.E. Ritter[2], A. Webel[3], R. Pace[3], S. Norris[3], K.M. Naughton[2], M.L. Drumm[4], M.R. Knowles[3], G.R. Cutting[2].* 1) Division of Pediatric Endocrinology, Johns Hopkins University, Baltimore, MD; 2) Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 3) Cystic Fibrosis-Pulmonary Research and Treatment Center, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC; 4) Departments of Pediatrics and Genetics, Case Western Reserve University, Cleveland, OH.

Diabetes occurs in 25-50% of CF patients (a.k.a. CF related diabetes; CFRD) resulting in poorer lung function and increased mortality. Although the cause of CFRD is unknown, we previously showed by twin study that genetic modifiers are largely responsible for CFRD risk (heritability ~1.0; Blackman et al. 2009). Clustering of type 2 diabetes (T2D) and CFRD within families and sharing of a susceptibility gene (TCF7L2) suggested overlapping pathophysiology (Blackman et al., in press). As part of the North American CF Modifier Consortium GWAS, we have now tested single nucleotide polymorphisms (SNPs) genome-wide for association with CFRD. Genotypes for 1796 individuals (children with CF and available parents) in 559 families in the CF Twin and Sibling Study were obtained using the Illumina 610 Quad array. Stringent quality control criteria were met for 552,214 autosomal SNPs (96.5%). We compared transmissions in euglycemic and chronically diabetic CF patients using family-based association testing. Pre-test power, calculated by excluding offspring genotypes, was used to rank SNPs prior to association testing. Of the top 15% of SNPs ranked by power, association was most significant for CTNNA2 (alpha 2 catenin; $P=3 \times 10^{-6}$; involved in Wnt signaling), VANGL1 (vang-like 1; $P=3 \times 10^{-6}$; involved in Wnt signaling), ATRNL1 (attractin-like 1; $P=1 \times 10^{-5}$; reported to interact with the melanocortin-4 receptor), near PLEKHF2 ($P=1 \times 10^{-5}$), and STX18 (syntaxin 18; $P=8 \times 10^{-5}$). A second study was then performed using 167 CF patients with diabetes and 635 CF controls from the UNC/CW CF Genetic Modifiers Study. Of the above SNPs, the CTNNA2 SNP (same allele as above) showed association with diabetes (two-sided $P=6 \times 10^{-3}$). The combined P value ($P=2 \times 10^{-7}$) is statistically significant after Bonferroni correction for 82,835 tests in the two analyses. The risk ratio was estimated by Cox regression (ignoring family structure). After adjusting for previously identified risk factors for diabetes (steroid treatment and TCF7L2 genotype) the CTNNA2 SNP associated with diabetes incidence (hazard ratio=2.1 per allele, P=0.02). Involvement of CTNNA2 along with the previously identified TCF7L2 implicates a novel pathway of Wnt signaling in the pathogenesis of diabetes in CF. *Supported by CF Foundation CUTTIN06P0, KNOWLE00A0, and NIH DK076446 (S.M.B.), DK044003, HL068927 (G.R.C.), HL068890, DK066368, RR000046 (M.R.K.).*.

## 94

**Nine new genome-wide significant loci influence susceptibility to type-2 diabetes.** *B.F. Voight[1,2], L. Scott[3], R. Welch[4], V. Steinthorsdottir[5], U. Thorsteinsdottir[5], A.P. Morris[6], C. Dina[7], T.M. Frayling[8], E. Zeggini[6,9], Y. Aulchenko[10], R. Sladek[11], J.F. Wilson[12,13], P. Froguel[14,15], K. Stefansson[5], D. Altshuler[1,2,16,17], M. Boehnke[3], M.I. McCarthy[6,18] on behalf of the Diabetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium.* 1) Molecular Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA; 2) Center for Human Genetics Research, MGH, Boston MA; 3) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 4) Bioinformatics Program and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 5) deCODE Genetics, Reykjavik, Iceland; 6) Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 7) UMR-8090, Institute of Biology and Lille 2 University, Pasteur Institute, Lille, France; 8) Genetics of Complex Traits, Peninsula Medical School, Exeter, UK; 9) Wellcome Trust Sanger Institute, Hinxton, UK; 10) Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 11) McGill University and Genome Quebec Innovation Centre, Montreal, Canada; 12) Public Health Sciences, University of Edinburgh Medical School, Edinburgh, UK; 13) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Edinburgh, UK; 14) CNRS 8090-Institute of Biology, Pasteur Institute, Lille, France; 15) Section of Genomic Medicine, Hammersmith Hospital, Imperial College London, London, UK; 16) Department of Molecular Biology and Medicine, MGH, and Department of Genetics, Harvard Medical School, Boston, MA; 17) Diabetes Unit, MGH, Boston, MA; 18) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK.

A central goal of genetic mapping is to reveal etiological pathways and the underlying biology, e.g., ciliary dysfunction in Bardet-Biedl Syndrome [Ansley et al, 2003]. For type-2 diabetes (T2D), despite the success of several mapping studies, the biological pathways they perturb remain elusive. Here, we report the discovery of nine new susceptibility loci, bringing the total known common genetic risk factors to 29. Although these loci contain approximately 0.2% of all genes, we observe multiple gene pairs known to be coregulated or act in a shared pathway - specifically, genes involved in cell-cycle regulation, the coregulated genes *HMGA2* and *IGF2BP2*, and *NOTCH2-FURIN*. Our meta-analysis comprised genome-wide association data from 8,130 cases and 38,897 controls of European descent from eight studies. After removing 17 regions previously known, lead SNPs representing the 25 next strongest associations were taken into replication (48,468 cases, 99,020 controls). 12 loci exceeded a genome-wide statistical threshold (odds ratios from 1.06 to 1.13, p-values from $3.3 \times 10^{-8}$ to $2.3 \times 10^{-20}$). Three of these represent loci for which there is prior evidence of T2D-susceptibility effects (*MTNR1B*, *IRS1*, and *KCNQ1*). Two loci had near genome-wide significance in previous analysis (*BCL11A*, *HNF1A*). The remaining seven signals - mapping near to the genes *HMGA2*, *CENTD2*, *KLF14*, *FURIN*, *TP53INP1*, *ZBED3*, and *ZFAND6* - represent previously unreported T2D risk loci. Of the genes in these regions, several biological relationships are known. *HMGA2* regulates another T2D associated gene, *IGF2BP2* [Brants et al, 2004]. In human liposarcomas, both genes are strongly coexpressed [Cleynen et al, 2007], and the mouse knockout of *HMGA2* has defects in adipocyte differentiation [Anand and Chada, 2000]. This relationship, taken together with association at *PPAR*$\gamma$, posits the hypothesis that genes in this pathway mediate T2D susceptibility through adipocyte function. In a separate pathway, *NOTCH2* is known to be processed by furin-like proteases, and both the *NOTCH2* and *FURIN* gene regions contain SNPs associated with T2D. Finally, analysis of the 29 loci using pathway-based analytical methods (GRAIL, REACTOME, and PANTHER Gene-Ontology) implicates multiple genes involved in cell-cycle pathways. Our results illustrate that genes found by unbiased studies in T2D are, as the number of loci accumulates, beginning to generate biologically meaningful hypotheses about the etiology of disease.

## 95

**Novel genetic loci implicated in fasting glucose homeostasis and their impact on related metabolic traits.** *I. Barroso[1,20], C. Langenberg[2,20], I. Prokopenko[3,4,20], R. Saxena[5,6,20], N. Soranzo[1,7,20], A.U. Jackson[8], E. Wheeler[1], N.L. Glazer[9], N. Bouatia-Naji[10], L. McCulloch[4], A.L. Gloyn[4], R. Sladek[11,12], P. Froguel[10,13], R.M. Watanabe[14], J.B. Meigs[6,15], L. Groop[16], M. Boehnke[8], M.I. McCarthy[3,4,17], J.C. Florez[5,6,15,20], J. Dupuis[18, 19,20] on behalf of MAGIC.* 1) Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) MRC Epidemiology Unit, Cambridge, United Kingdom; 3) WTCHG, Oxford, United Kingdom; 4) OCDEM, Oxford, United Kingdom; 5) Broad Institute, Cambridge, MA, United States; 6) Massachusetts General Hospital, Boston, MA, United States; 7) King's College London, London, United Kingdom; 8) University of Michigan School of Public Health, Ann Arbor, MI, United States; 9) University of Washington, Seattle, WA, United States; 10) CNRS-UMR8090, Lille, France; 11) McGill University, Montreal, Canada; 12) Genome Quebec Innovation Centre, Montreal, Canada; 13) Imperial College London, Hammersmith Hospital, London, United Kingdom; 14) University of Southern California, Los Angeles, CA, United States; 15) Harvard Medical School, Boston, MA, United States; 16) Lund University, University Hospital Malmo, Malmo, Sweden; 17) Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK; 18) Boston University School of Public Health, Boston, MA, United States; 19) National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, United States; 20) equal contribution.

Impaired β-cell function and insulin resistance are key determinants of type 2 diabetes (T2D). To date, there are ~20 loci reproducibly associated with T2D with most of them implicated in reduced β-cell function and/or mass leading to concomitant decreases in insulin secretion. Genetic analyses of continuous glycaemic traits have also led to the discovery of loci influencing fasting glucose levels that appear to act primarily via effects on the glucose-sensing machinery of the β-cell, leading to defective insulin secretion. These approaches have additionally identified a novel locus, *MTNR1B*, whose glucose-raising allele also increases T2D risk. To identify additional glycaemic trait loci and investigate their metabolic impact, we performed meta-analyses of 21 genome-wide associations studies informative for fasting glucose (FG, N=46,263), fasting insulin, and indices of β-cell function (HOMA-B) and insulin resistance (HOMA-IR) (N=38,413). Follow-up of 25 loci in 61,219 independent samples discovered nine new genome-wide significant ($P<5x10^{-8}$) associations with FG (in or near *ADCY5*, *MADD*, *ADRA2A*, *CRY2, FADS1*, *GLIS3*, *SLC2A2*, *PROX1* and *FAM148B*) and a novel locus for fasting insulin and HOMA-IR (*IGF1*). Also associated with FG were established T2D loci *TCF7L2* and *SLC30A8*, and previously reported loci *GCK*, *GCKR*, *G6PC2*, *MTNR1B* and *DGKB/TMEM195*; *GCKR* also achieved genome-wide significant association with fasting insulin and HOMA-IR. Within the associated loci the most likely biological candidate genes influence signal transduction, development, glucose-sensing and circadian regulation. Furthermore, RT-PCR experiments across a panel of human tissues including pancreas, islets and flow-sorted β-cells demonstrate that the majority of these are highly expressed in the pancreatic islets and sorted β-cells. While this pattern of gene expression suggests that FG may be strongly influenced by genes expressed in β-cells, which is concordant with previous observations for T2D, the results of our well-powered meta-analysis demonstrates that T2D and QT traits related to blood glucose are influenced by distinct genetic variants.

## 96

**Variable relationships between loci influencing risk of type 2 diabetes and fasting glycemic traits revealed by large-scale genome wide association analyses.** *I. Prokopenko[1,2], B.F. Voight[3,4,5,6], R. Saxena[3,4,5,6] for the MAGIC and DIAGRAM investigators.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford OX3 7LJ, UK; 3) Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), Cambridge, MA 02142, USA; 4) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA; 5) Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA; 6) Department of Medicine, Harvard Medical School, Boston, MA 02115, USA.

Complex relationships can exist between disease traits and their related continuous endophenotypes. Recently, we have demonstrated involvement of the *MTNR1B* locus in both regulation of fasting glucose (FG) set point in healthy individuals and in influencing individual risk of novel type2 diabetes (T2D) locus. Gene discovery and replication efforts within the large-scale MAGIC (continuous glycemic traits) and DIAGRAM (T2D) consortia allowed us to explore these relationships further. Within MAGIC, genome-wide (GW) meta-analysis and follow-up including a total of up to 107,482 samples, nine new loci (*ADCY5, MADD, ADRA2A, CRY2, FADS1, PROX1, GLIS3, SLC2A2* and *FAM148B*) for FG and indices of β-cell function (HOMA-B), and a novel association with fasting insulin and HOMA-IR (*IGF1*) (all GW significant: $P<5x10^{-8}$) were identified. Among known T2D loci, only two (*TCF7L2, SLC30A8*) were GW significant for FG. Six other known T2D loci were associated with FG at $P<10^{-4}$ (*WFS1, CDKN2B, CDC123, IGF2BP2, KCNQ1, THADA*), while *CDKAL1, HHEX, PPARG, ADAMTS9* were only nominally ($P<0.05$) associated. In each case, the T2D-risk and FG-raising alleles were the same. No association with FG was seen within *NOTCH2, HNF1B, KCNJ11, TSPAN8, FTO* or *JAZF1* loci. *PPARG* and *FTO* were associated ($P<5x10^{-3}$) with increased FI and HOMA-IR. In the complementary analysis, we examined 10 novel and 4 known (*GCK, GCKR, DGKB, G6PC2*) fasting trait loci for their impact on T2D-risk in 23 case-control samples of European descent (~34,500 T2D cases; ~88,300 controls). These analyses demonstrated GW significant associations with T2D for *DGKB, ADCY5, PROX1*and *GCKR*. Variants in *GCK, ADRA2A, FADS1, GLIS3, SLC2A2* were associated (at $P<5x10^{-3}$) with increased risk of T2D; association at $P<0.05$ was seen for *G6PC2, CRY2* and *FAM148B*. No association was observed for 2 loci (*MADD, IGF1*). At all but *G6PC2*, the glucose-raising allele contributed to higher risk of T2D. This study extends examination of the overlap between common variant signals associated with T2D and fasting glycemic traits, closely-related phenotypes from both epidemiological and pathophysiological perspectives. Incomplete correspondence between loci influencing these traits reinforces the complex relationships that can exist between endophenotypes and their cognate disease entities.

## 97

**Comparing Sanger Whole Human Genome Sequencing to Targeted Genome Enrichment and Next Generation Sequencing.** *S. Levy[1], Y.H. Rogers[1], R.L. Strausberg[1], B. Boese[2], P. Bouffard[2], Z. Markovic[2], A. Goncalves[2], L. Li[2], Q. Zhao[2], T. Harkins[3].* 1) Human Genomic Medicine, J. Craig Venter Institute, Rockville, MD; 2) 454 Life Sciences, A Roche Company; 1 Commercial Street, Branford, CT, 06405; 3) Roche Applied Science 9115 Hague Rd, Indianapolis, IN 46250.

Genetic variation in the protein-coding compartment of the human genome harbors the consequences of altered protein function in human populations studies of disease. Understanding the contribution of such changes in the protein compartment toward disease phenotype would provide better targets for therapeutic interventions. The human exome represents the collection of exons that code for proteins and presents an attractive target in new experimental designs to quickly identify potentially diseases causing genetic variants. In attempt to assess the required sequence coverage necessary for routine detection of multiple human samples, we performed exome enrichment and sequencing on the HuRef DNA. This particular DNA sample possesses a sequenced and assembled genome on which a comprehensive set of DNA variants, (SNP's, indels and structural variants) has been detected. Thus, the HuRef sample provides an opportunity to evaluate the combination of sample enrichment via sequence capture, as performed using the Nimblgen Exome 2.1M array, and Roche 454 sequencing. DNA variants were identified employing data from 3 or more reads where at least two reads were in opposite seqeunce orientation and resulted in the detection of 53,150 variants consisting of SNP's and indels. Restricting our attention to the ~26.4 Mb of exons represented on the array, which consists of a region of total size ~34.2 Mb for capture, we derived the concordance of variant detected by Nimblegen Exome-Roche-454 compared to Sanger-based whole genome assembly. Our preliminary analysis shows that we found ~83% of the 10,385 variants detected by array capture followed by Roche 454 sequencing were corroborated by the Sanger sequencing data set. We will present further detailed analysis specifying the depth of coverage range for the target regions, the extent of coverage of exome variation with our data set in the HuRef sample and an accurate accounting of false detection rates. We expect that these analyses will allow us to define a strategy employing exome sequencing in large disease-specific cohorts.

## 98

**A Comprehensive and Cost Effective Approach for Large Scale Targeted Resequencing and Variant Discovery from the Human Genome.** *H. Ji[1,2], G. Natsoulis[1], J. Bell[2], H. Ordonez[2], K. Welch[2], M. Zhang[1], M. Jalali[1], I. Kela[1], R. Davis[2].* 1) Medicine / Oncology, Stanford University School of Medicine, Stanford, CA; 2) Stanford Genome Technology Center Palo Alto, CA.

Detection of clinically relevant germline rare variants or cancer somatic mutations from candidate genes and other genomic regions requires deep resequencing of large numbers of samples. As a general solution for the biomedical research community which will enable the large scale targeted resequencing of clinical populations, we have developed a novel strategy using sequence specific genomic circularization to target nearly any region of the human genome. A targeting oligonucleotide with flanking genomic homology sequences mediates the circularization of a genomic region-of-interest that has been linearized with a specific restriction enzyme site. With standard molecular biology lab requirements, this process of selective genomic circularization can be run in parallel such that hundreds if not potentially thousands of exons or any other region-of-interest can be targeted, amplified in a single reaction volume and subsequently resequenced on next generation DNA sequencer systems. A number of other methods capture or enrich specific targeted regions of the genome either through a solid-support mediated hybridization or in-solution. All of these approaches including ours demonstrate variability in i) successfully targeting or capture ii) distributions of targeted regions in terms of abundance and iii) sequencing with adequate fold coverage for genotyping. We have developed straightforward solutions for all of these issues. This includes normalization of the distribution of targeted regions, indexing of multiple samples and development of algorithms that enable straightforward variant discovery including insertion - deletions. We optimized our approach sequencing nuclear family trios. Furthermore, there is the issue of cost with commercially available capture approaches remaining high. Our system remains highly scalable and is cost effective such that large number of candidate genes and regions-of-interest can be targeted and resequenced at a fraction of the cost of commercially approaches. As a general solution for the human genetics community, we have recently completed an in-silico design which covers 17,000 human genes and has over 300,000 oligo designs which will be released in 2009.

## 99

**Genome sequencing coupled with cascade testing as a tool to identify rare disorders masquerading as common disease.** *C. Turner[1], F.M. Facio[1], P. Cherukuri[1,2], P. Cruz[1,2], N.F. Hansen[1,2], J.J. Johnston[1], R. Shamburek[3], J.C. Mullikin[1,2], E.D. Green[1,2], L.G. Biesecker[1], NISC Comparative Sequencing Program.* 1) National Human Genome Research Institute, Bethesda, MD; 2) NIH Intramural Sequencing Center, Bethesda, MD; 3) National Heart Lung and Blood Institute, Bethesda, MD.

Clinical research studies that utilize the power of large-scale medical sequencing (LSMS), allow the development of associated infrastructure to enable personalized medicine. As such, the ClinSeq project aims to diagnose and screen patients for disease with the initial focus on atherosclerosis and risk for myocardial infarction. A well-known disease, familial hypercholesterolemia (FH) is a disorder of LDL cholesterol metabolism caused by mutations in LDLR, APOB, or PCSK9. Identified FH-causing mutations in individuals have clinical implications for other family members. Untreated individuals with heterozygous FH have an increased risk of early myocardial infarction and death before the age of 60. This risk is substantially mitigated by aggressive treatment. The World Health Organization estimates the incidence of FH to be 0.2% (1 in 500) in the United States. However, it is estimated that only ~5% of affected individuals have been diagnosed and <10% of diagnosed FH patients are adequately treated. Genetic testing could help address this unmet need. Cascade testing, defined as the identification of an index case followed by subsequent identification of affected family members, has been proposed as the most effective means to identify individuals with FH. Here, we report the identification of eight individuals with mutations causing FH out of an initial cohort of 360 individuals screened by LSMS. In our cohort, the frequency of FH is 2%, which is over ten times the expected population frequency of the disorder and suggests that participants in our study and referring physicians recognize an unmet need for FH testing. Seven previously reported mutations were identified in LDLR (p.N687QfsX29, p.P685L, c.313+1G>A, p.G218VfsX47, p.Y188X, p.W87G, and a deletion of exons 17 and 18), and one in APOB (p.R3527Q). Review of the family history of these eight probands identified 112 relatives for cascade testing with cholesterol screening and possibly mutation analysis. We will present results of the cascade testing of these 112 individuals and treatment uptake. This approach to case finding would be equivalent to screening approximately 50,000 unrelated persons in a direct case finding strategy. Significant potential exists for LSMS to identify families affected with genetic disease, as well as improve disease prevention and treatment efficacy.

## 100

**Analysis Of The 1000-Genome Project Sequencing Data Reveals An Unexpectedly High Level Of Human Genetic Polymorphisms Derived From Mobile Elements.** *P. Liang.* Dept Biological Sci, Brock Univ, St. Catharines, Canada.

Mobile elements constitute approximately half of the human genome, and they impact the evolution and function of the genomes via a variety of mechanisms. With members of certain classes of mobile elements remain active, they contribute to an important part of genetic polymorphisms among humans via recent and *de novo* DNA transposition, as well as homologous recombination, attributing partly to the phenotype differences in humans, such as disease susceptibility. So far, several thousands of retrotransposon insertion polymorphisms (RIPs) have been documented, with the majority of those identified from computational comparative genomic analyses based on limited human genomic sequences. But an accurate estimation of the actual level of such types of polymorphisms existing in humans remains infeasible due to technical difficulties. In an effort to address this question, we utilized the very recently available new and rich human genome sequence data generated by The 1000 Genome Project Consortium. In this preliminary analysis, we focused on the data for a few individual genomes, for which deep sequencing data has been available, and used the pair end sequence reads for identifying mobile element insertions that are present in the reference genome but not in the new genomes. By comparing such the data with those documented in the database of Retrotransposon Insertion Polymorphisms in humans (dbRIP), we observed an unexpectedly large number of novel polymorphisms from Alu and L1 insertions, with the majority belonging to older subfamilies that are thought to be less active. Surprisingly, we also identified a large number of candidate cases for the movement of DNA transposons, which are considered to be completely dead in the human genomes. In addition, we attempted to assess the rate of *de novo* DNA transposition in one reproductive generation by comparing between the children and their parents. While these computationally identified DNA transposition polymorphisms required to be validated, our preliminary data does suggest: 1) personal genome data generated using the 2nd generation of sequencing technologies is of great utility for identifying novel transposition-derived genetic variations; 2) the degree of such type of variations may be much higher than what we currently appreciate. This research is in part supported by grants from the Canada Research Chair program, CFI, Ontario MRI and Brock University to PL and is facilitated by the use of SHARCNET HPC.

## 101

**A Genome Wide Method for Candidate Causative Mutation Discovery.** *M. Bainbridge[1,2], T. Albert[3], D. Muzny[1], D. Burgess[3], L. Nazareth[1], A. Hawes[1], R.A. Gibbs[1].* 1) Human Genome Sequencing Center, Baylor College Med, Houston, TX; 2) SCBMB, Baylor College Med, Houston, TX; 3) Roche NimbleGen, Inc. 504 South Rosa Road Madison, WI.

Solid-surface gDNA capture coupled with second generation high-throughput sequencing was used to screen the consensus coding sequence (CCDS) exons of two related individuals suffering from spino-cerebellar ataxia with unknown genetic cause. These two individuals are from a 25 member pedigree with 11 affected individuals. Ataxia is an autosomal dominant disease which can be caused by mutations in a wide range of genes and at non-conserved positions. Using both the 454 and SOLiD sequencing platforms, 55% and 92% of exonic bases were covered by at least 10 unique reads, respectively. We discover over 25,000 mutations in each individual. Approximately 85% of these mutations were in dbSNP, and 20% were non-synonymous. However, we show that by combining these data we are able to find high-quality mutations shared by both individuals even if the mutations are at low coverage in one or both samples. Combing the sequence data yields 200 shared, non-dbSNP mutations which either produce non-synonymous mutations or affect mRNA splicing. These mutations are being validated by Sanger sequencing in all members of the pedigree. This method of genome-wide analysis of close relatives illustrates a new general approach to disease gene mutation discovery. Examination of effected, distantly related probands allows facile elimination of non-causative mutations with no a priori information about likely disease-related genes, and does not rely on phenotype-based prediction methods to enrich for candidate mutations. This capture technique is more cost effective than whole genome sequencing by ensuring that the sequence coverage is directed towards regions of interest, allowing robust mutation discovery and genotyping. Unlike other methods, this technique is applicable to a small pedigree with as few as two affected individuals and directly identifies causative mutations.

## 102

**Whole Genome Sequencing for Identification of Causative Alleles in a Recessive Trait.** *R. Gibbs[1, 2], J. Reid[1], C. Gongaza-Jauregui[1], L. Nazareth[1], D. Deiros[1], M. Bainbridge[1], A. Sabo[1], M. Morgan[1], D. Wheeler[1], H. Dinh[1, 2], R. Irikat[5], W. Tom[5], D. Guo[5], C. Yan[5], C. Gehman[5], N. Fantin[5], F. Zhang[2], A. Beaudet[2], D. Muzny[1,2], J. Lupski[2,3,4].* 1) Hum Genome Seq Ctr, Baylor Col Med, Houston, TX; 2) Dept. Human Molecular Genetics, Baylor Col Med, Houston, TX; 3) Dept. of Pediatrics, Baylor Col of Med, Houston, Tx; 4) Texas Chilldrens Hospital, Houston, TX; 5) Life Technologies, Foster City, Ca.

Following the successful application of the Life Technologies SOLiD platform for sequencing genomes in the 1000 Genomes project, we applied SOLiD sequencing for the identification of disease alleles in one family with an apparent recessive Charcot-Marie-Tooth Syndrome. We generated deep (30 x) whole genome sequence (wgs) coverage of the genome of one proband and mapped the sequence reads to the reference genome. Variations between the sample and the reference were identified and the putative mutation lists were searched for likely candidates and tested for validation and transmission. In this family we identified a variant of likely functional significance at the SH3TC2 locus, that was clearly transmitted with the trait. A second putative allele has been identified at the PLEKHG5 locus and is currently being validated. This genetic study reveals the power of wgs approaches for solving individual cases of genetic disease allele identification. A similar approach has been applied to two other families with disorders with autosomal dominant inheritance. As costs for sequencing are being continually reduced, we predict generation of wgs to become a fundamental practice in personalized medicine.

## 103

**Sequencing the Unsequenceable: Applying Massively Parallel, Single-Molecule Sequencing to Badly Degraded DNAs.** *J. Thompson[1], D. Lipson[1], C. Hart[1], P. Kapranov[1], S. Letovsky[1], P. Milos[1], F. Ozsolak[1], T. Raz[1], J. Reifenberger[1], K. Steinmann[1], O. Loreille[2], M. Coble[2].* 1) Res & Development, Helicos Biosci, Cambridge, MA; 2) Armed Forces DNA Identification Laboratory, Rockville, MD.

Next generation DNA sequencing has revolutionized many aspects of genetics by providing huge amounts of data on many sample types. However, some sample classes have remained recalcitrant to examination because of their small size or high level of degradation. Because most next generation sequencing technologies require amplification and a particular size range of target molecules, DNAs not meeting those criteria cannot be handled. Single molecule sequencing does not suffer from those limitations as no amplification is required and degraded molecules can be used directly as templates. We have used single molecule sequencing for examining both DNA and RNA from FFPE samples that, in some cases, were too heavily modified and/or too short to be sequenced using amplification-based approaches. To test the extremes of poor sample quality and quantity, sequencing was attempted with a DNA sample extracted from Korean War-era remains from the National Memorial Cemetery of the Pacific that had been exposed to highly damaging conditions during embalming and which, even after extensive amplification, had not previously provided usable human sequence when analyzed using Sanger sequencing or another next generation technology. A simple sample preparation protocol with no amplification and utilizing only terminal transferase to tail the 3' end of the DNA was employed. A large quantity of both mitochondrial and nuclear sequence was obtained. Because such degraded samples cannot be genotyped using standard techniques, we developed a method, Sparse Allele Calling, in which the sequence information generated in a single channel of a HeliScope single-molecule sequencer is used at very low (~0.1X) coverage to determine whether DNA sequence is from the same individual or different individuals and, if different, the degree of relatedness of those individuals. After aligning a given DNA sequence to the reference human genome, positions of variation are compared between samples. While only one of two alleles can be detected at each position, and only a fraction of the variations are detected for each sample, the sheer number of available calls allow thousands of variations to be compared so the probability of matching with the same, or a related, individual becomes higher than with samples from non-related individuals. By comparing the number of identical allele calls and weighting by allele frequency, identity and degrees of relatedness and shared ancestry can be determined.

## 104

**Complete Genome Sequencing of the hunter-gatherers of the Kalahari Desert.** *S.C. Schuster[1], W. Miller[1], E.E. Eichler[2], E.R. Mardis[3], A. Siepel[4], R. Hardison[1], J.C. Mullikin[5], F. Chiaromonte[1], A. Ratan[1], B. Harris[1], B. Giardine[1], D.I. Drautz[1], L.P. Tomsho[1], L. McClellan[1], B.M. Giardine[1], C. Riemer[1], V. Taliwal[4], D.C. Peterson[7], T.T. Harkins[6], V.M. Hayes[7].* 1) Biochemistry and Molecular Bio, Penn State University, University Park, PA; 2) Department of Genome Sciences, University of Washington; 3) Genome Sequencing Center, Washington University in St. Louis; 4) Biological Statistics & Computational Biology, Cornell University; 5) National Human genome research Institute; 6) Roche Applied Sciences; 7) Children's Cancer Institute of Australia, Sydney, Australia.

The Kalahari Bushmen may be the last hunter-gathers of our times. Their evolutionary and cultural histories distinguish them from all other ethnic groups inside and outside of Africa. They are believed to be the oldest human population, and therefore represent a unique and ancient genetic heritage. In this study we investigate the extent of genetic variation of the Bushman genome in comparison to other publically available human genomes. For this undertaking we have performed whole-genome shotgun sequencing using Roche/454 Titanium chemistry of two complete genomes, together with exome sequencing of two matched and two additional samples. Our study investigates the SNP diversity, indel and repeat content, as well as copy number variation of this ethnic group against the human reference sequence. The genetic variations detected by sequencing are being validated in parallel using DNA SNP arrays. The study therefore aims at generating a high-quality version of a human genome that defines the outer boundaries of human diversity. The new data will aid the interpretation of other human genome sequences. As a large number of the discovered genetic variants are novel and currently not contained in dbSNP, this project will aid future studies on rare human alleles. Also, with current genome-wide association studies being largely limited to modern populations, disease associations have generally been mapped to broad genomic regions. Human diversity studies of the Bushmen will facilitate the narrowing of these regions. More important, we believe, will be the advantage to the Bushmen. It is hoped that availability of the first Southern African human genome sequences will help the development of drugs that benefit this ethnic group. Moreover, the data highlight the genetic richness of these largely forgotten people of Africa, emphasizing the importance of preserving this unparalleled archive of human cultural and genetic history, and spot-lighting the need to assist indigenous groups in Africa and elsewhere in their fight against home-land loss, disease and famine.

## 105

**Diversity survey of the chronic diabetic wound microbiome.** *E.A. Grice[1], L.J. Yockey[1], S. Conlan[1], N.I.S.C. Comparative Sequencing Program[2], K.W. Liechty[3], J.A. Segre[1].* 1) Genetics and Molecular Biology Branch, NHGRI, NIH, Bethesda, MD; 2) NIH Intramural Sequencing Center, NHGRI, NIH, Bethesda, MD; 3) Department of Surgery, Children's Hospital of Philadelphia, Philadelphia, PA.

As the rate of obesity increases worldwide, so does the incidence of type 2 diabetes and associated complications. One of the most common and costly complications of diabetes is chronic, non-healing wounds. Increased bacterial colonization and/or infection of diabetic wounds has a deleterious effect on healing and is a significant source of clinical complication. The precise relationship between microbes and impaired wound healing remains unclear despite general acknowledgement that the resident microflora is a source of complications. The main limitation to examining the role of bacteria in the pathogenesis of the diabetic wound healing impairment is the reliance on culture-based assays. Some bacterial species colonizing the diabetic wound have been identified in this manner, but these results are biased because it is estimated that only ~1% of bacteria are cultivable. We recently demonstrated the utility of sequencing the prokaryote-specific 16S small-subunit ribosomal RNA genes in order to profile the topographical and temporal bacterial diversity of normal human skin.

We hypothesize that unique microbial populations and/or shifts in microbial community structure contribute to impaired wound healing in diabetics and can be detected using less-biased genomic methods. Using an excisional wound model in genetically diabetic mice homozygous for the *Lepr$^{db}$* mutation (db/db) and non-diabetic heterozygous controls (db/+), we collected wound swabs over a time course of 28 days. 384 near full-length 16S genes were sequenced from each wound at each time point. Sequences were assigned to the prokaryotic taxonomy and community membership, structure, and phylogenetic lineages were compared. Using this method, we have identified a unique shift in the bacterial communities populating diabetic skin and wounds in the db/db mouse model. We are confirming that this selective shift also occurs in the human condition. These studies are a foundation for our long-term goal to elucidate the contribution of the diabetic wound microbiome to impaired wound healing and translate this into novel therapeutic approaches.

## 106

**The rate of mutation in male and female germlines.** *D.F. Conrad, S.J. Lindsay, M.E. Hurles, 1000 Genomes Project.* The Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom, CB10 1SA.

Mutation underlies all genetic studies, and yet the mutation rate in the human genome remains imprecisely estimated. Evolutionary studies have suggested that the base substitution rate in the male germline may be two to three-fold higher than in the female germline. These imprecise estimates necessarily rely on several evolutionary assumptions. The advent of high-depth genome-wide resequencing of related individuals promises to revolutionise our understanding of mutation. We have developed a novel statistical approach to assess each base in the genome as a potential de novo mutation in parent-offspring trios. We have applied this method to estimate mutation rates from data on two trios from the 1000 genomes project. By combining these results together with additional data that allow us to assign de novo mutations to haplotypes, we have estimated the rates of base substitution in male and female germlines in these trios, as well as the rate of somatic mutation in cell lines. We assessed the impact of sequence context on these mutation processes. For the first time, these results give us a precise estimate of the mutation rate in a single generation, and a direct estimate of the additional mutagenicity of the male germline.

## †107

**Copy number variation within *NME7* associated with neuroblastoma.** *S.J. Diskin[1], C. Hou[1,2], M. Laudenslager[1], M. Diamond[1], K. Bosse[1], M. Li[3], P.A. Kanetsky[3], K. Wang[2], J. Glessner[2], H. Zhang[2], C. Kim[2], A.I.F. Blakemore[4], M. Devoto[5,6], H. Li[3], E.F. Attiyeh[1,6], E.R. Rappaport[1,6], K.L. Nathanson[7,8], H. Hakonarson[2,5], J.M. Maris[1,6,8].* 1) Division of Oncology and Center for Childhood Cancer Research, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA; 4) Genomic Medicine, Imperial College London, London, UK; 5) Division of Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 6) Department of Pediatrics, University of Pennsylvania, Philadelphia, PA; 7) Department of Genetics, University of Pennsylvania, Philadelphia, PA; 8) Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA.

Neuroblastoma (NB) is a solid tumor of the developing sympathetic nervous system that most commonly affects young children and is often lethal. We previously showed that common SNPs at 6p22 and 2q35 (within *BARD1*) are associated with aggressive NB (*NEJM* 2008; *Nat Genet* 2009). We also demonstrated that a common copy number variation (CNV) at 1q21.1 predisposes to NB and disrupts a new member of the neuroblastoma breakpoint family (*NBPF23*) (*Nature* in press). Here, we performed a second CNV-based genome-wide association study (GWAS) in 1,441 patients and 4,160 control subjects of European ancestry from three case series after slightly relaxing the minimum number of SNPs required for CNV detection. We identified an under-represented deletion within *NME7* (non-metastatic cells 7) at 1q24.2 associated with NB in the discovery set ($P$=2.4x10$^{-8}$; OR= 0.51) that replicated in two independent case series (combined $P$=1.9x10$^{-12}$; OR=0.52). *NME7* is a member of the nucleoside diphosphate kinase (NDK) family, which consists of ten distinct members in humans. *NME1* has been shown to suppress metastasis in human melanoma and breast cancers. Conversely, *NME1* promotes metastasis in other cancers including NB, where oncogenic mutations have been identified. To study somatic alterations of *NME7* in NB, we assayed genome-wide DNA copy number in 599 primary NBs and genome-wide RNA copy number in a subset of 100 tumors. Relative gain of *NME7* was observed in 24% (141/599) of tumors, and *NME7* transcript levels were significantly higher in those tumors harboring somatic gain ($P$=0.007). Finally, as an initial effort to determine if the CNV at *NME7* is associated with neuroblastoma alone, or cancer in general, we evaluated our recent SNP-based GWAS of 277 Caucasian testicular germ cell tumor (TGCT) patients compared to 919 healthy Caucasian controls (*Nat Genet* 2009) and detected a significant association ($P$=1.0x10$^{-3}$, OR=1.74), suggesting that common variation at this locus may influence the initiation of multiple cancers. Replication efforts using an independent case series of TGCT patients are ongoing. We conclude that a common CNV within *NME7* at 1q24.2 is associated with susceptibility to NB, and perhaps testicular cancer, and propose that somatic alterations of *NME7* may also play an important role in tumorigenesis. Efforts are ongoing to understand the biological relevance of these findings and to assess this CNV within the context of our previously reported SNP and CNV associations.

## 108

**Integrative genomics identifies LMO1 as a neuroblastoma predisposition gene.** *K. Wang[1], H. Zhang[1], H. Cui[1], S.J. Diskin[2], C. Winter[2], C. Bosse[2], J.T. Glessner[1], C. Kim[1], L. Nguyen[3], M. Diamond[2], S.F. Grant[1,4], Y. Mosse[2], S. Hannenhalli[5], H. Li[3], M. Devoto[3,4], W. London[6], M. Capasso[7], N. Rahman[8], H. Hakonarson[1,4], J.M. Maris[2].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Oncology and Center for Childhood Cancer Research, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA; 4) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA; 5) Department of Genetics, University of Pennsylvania School of Medicine, Philadelphia, PA; 6) Department of Statistics, University of Florida and Children's Oncology Group, Gainesville, FL; 7) University Federico II, CEINGE - Biotecnologie Avanzate, Naples, Italy; 8) Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey, UK.

To identify genetic risk factors for neuroblastoma, an often lethal childhood cancer, we carried out a genome-wide association study (GWAS) on 1,627 patients and 3,254 control subjects of European ancestry, and uncovered LMO1 (LIM domain only 1) on 11p15.4 as a predisposition gene (rs110419, P=5.1x10-10, OR=0.75). The association signal was subsequently replicated in three additional GWAS from US, UK and Italy, respectively, totaling 933 patients and 2537 control subjects. Additionally, genome-wide copy number alteration analysis of tumor specimens indicated that 20% of primary tumors carry gains encompassing LMO1, making 11p15.4 one of the most duplicated regions in neuroblastoma tumors. LMO1 encodes a cysteine-rich transcriptional regulator, and all its paralogues (LMO2, LMO3 and LMO4) have been previously implicated in cancer. The SNP risk alleles, somatic copy number gains, and somatic copy-neutral loss-of-heterozygosity are independently associated with increased LMO1 expression in primary tumors, indicating that gain-of-function mutations may influence neuroblastoma tumorigenesis. Furthermore, somatic gain of LMO1 was significantly more common in tumors from patients with stage 4 disease, diploid tumors and advanced age, indicating that LMO1 gain could act as a predictor for clinical outcomes. Whole-genome transcriptional genomics further identified PTK7, NCBP2 and MLLT3 as potential downstream targets of LMO1, and their expression levels correlate with both LMO1 risk genotypes and copy number gains. Altogether, converging lines of genomic evidence confirm LMO1 as a neuroblastoma susceptibility gene, and illustrate the importance of applying integrative molecular and genomic approaches in identifying novel oncogenes.

† Trainee Award Finalist

## 109

**Stratification of Wilms tumor by genetic and epigenetic analysis.** *R.H. Scott[1], A. Murray[1], C. Turnbull[1], L. Baskcomb[1], K. Spanova[1], R. Al-Saadi[2], R. Williams[2], A. Kelsey[3], G.M. Vujanic[4], N.J. Sebire[5], P. Grundy[6], C.A. Stiller[7], K. Pritchard-Jones[2], J. Douglas[1], N. Rahman[1], The FACT Collaboration.* 1) Section Cancer Genetics, Institute of Cancer Research, Sutton, United Kingdom; 2) Section Paediatric Oncology, Institute of Cancer Research, Sutton, United Kingdom; 3) Department of Paediatric Pathology, Royal Manchester Children's Hospital, Manchester, United Kingdom; 4) Department of Histopathology, School of Medicine, Cardiff University, Cardiff, United Kingdom; 5) Department of Histopathology, Great Ormond Street Hospital, London, United Kingdom; 6) Pediatric Hematology, Stollery Children's Hospital, Edmonton, Canada; 7) Childhood Cancer Research Group, Department of Paediatrics, University of Oxford, Oxford, United Kingdom.

Background Somatic defects at five loci, *WT1*, *TP53*, *CTNNB1*, *WTX* and the imprinted 11p15 *H19*/*IGF2* locus, are involved in the pathogenesis of Wilms tumor, the commonest childhood kidney tumor. These loci have not been analysed in a single tumor series, restricting knowledge of their order, interactions and clinical associations. Methods We analysed 107 Wilms tumors at *WT1*, *TP53*, *CTNNB1* and *WTX* by sequencing and MLPA and at 11p15 by methylation-specific-MLPA. Results Epigenetic defects at 11p15 resulting in *H19* hypermethylation were present in 70%; sporadic tumors, caused by *H19* epimutation in 38%; and paternal uniparental disomy 11p15 (pUPD 11p15) in 32%. *WTX* mutations were present in 37%, *CTNNB1* in 16%, *WT1* in 15%; and *TP53* in 6%. Strong associations between loci were identified: between defects in *CTNNB1* and *WT1* (P<0.001); 11p15 and *WTX* (P=0.01); and pUPD 11p15 and *WT1* mutation (P=0.01). The data give insights into the order of events in tumorigenesis and allow stratification of tumors into molecular groups according to their status at *WT1* and 11p15. Group 1 tumors (~15% sporadic tumors) are defined by the presence of *WT1* mutations. They frequently undergo pUPD 11p15 and the majority have *CTNNB1* mutations. Group 2 tumors (~60%) are defined as those with an 11p15 defect but no *WT1* mutation. One third harbor *WTX* mutations. The 11p15 defect is either *H19* epimutation (Group 2A) or pUPD 11p15 (Group 2B). Group 3 tumors (~25%) are defined as those with neither *WT1* nor 11p15 defects. These tumors were also negative for mutations in *WTX*, *CTNNB1* and *TP53* (P<0.001), and are therefore typically 'poly-negative' at the five known Wilms tumor loci. Tumors with somatic *WT1* mutations (Group 1) occur at a younger age (P<0.001). There was a strong association between somatic *H19* epimutation (Group 2A) and bilateral tumors (P= 0.002), suggesting *H19* epimutation may be causing tissue-specific epigenetic predisposition to bilateral disease. Finally, we present a streamlined algorithm that allows molecular classification of tumors using two analyses: combined 11p15-*WTX* MLPA and *CTNNB1* sequencing. Conclusions The study provides a new understanding of the patterns and sequence of events in Wilms tumorigenesis and identifies somatic *H19* epimutation as an underlying defect in sporadic bilateral tumors. The streamlined analysis algorithm opens the way to molecular stratification of tumors in clinical trials and will likely facilitate further clinical and scientific advances.

## 110

**Family-based association study identifies *MDM2* as a modifier gene in retinoblastoma.** *L. Castéra[1], A. Sabbagh[2,3], C. Dehainault[1], D. Michaux[1], B. Patillon[1], E. Lamar[1], I. Aerts[4], L. Lumbroso-le Rouic[5], J. Couturier[1], D. Stoppa-Lyonnet[1,6,7], M. Gauthier-Villars[1], C. Houdayer[1,7].* 1) Service de Génétique Oncologique, Institut Curie, Paris, France; 2) UMR INSERM 745, Université Paris Descartes, France; 3) Service de Génétique Moléculaire, Hôpital Beaujon, Clichy, France; 4) Service d'Oncologie Pédiatrique, Institut Curie, Paris, France; 5) Service d'Ophtalmologie, Institut Curie, Paris, France; 6) INSERM U830, Pathologie Moléculaire des Cancers, Institut Curie, Paris, France; 7) Université Paris Descartes, Paris, France.

Variable tumor expressivity of the disease is occasionally described in retinoblastoma (Rb) families and the existence of genetic modifiers has been suggested. The existence of such factors could be investigated in the pRB or p53 pathways, both involved in Rb development. MDM2, which increases p53 and pRb proteasome mediated catabolism, is therefore a prominent candidate. The MDM2 309G allele on MDM2 promoter is known to enhance MDM2 expression and thereby accelerate tumor formation in Li-Fraumeni patients. To address MDM2 impact on Rb development, the genetic transmission of the MDM2 309T>G SNP (rs2279744) was studied in pedigrees of 70 Rb families (212 carriers of a germline mutation) and the marker genotype was then correlated with age at diagnosis and disease phenotype. This transmission was also studied in a cohort of 98 bilaterally-affected Rb sporadic predisposed patients and their parents. Family-Based Association analysis were performed and a significant association of the MDM2 309G allele with bilaterally- and unilaterally-affected patients in Rb family was found under a recessive model (Z = 3.305, exact P value = 0.001). This strong association with retinoblastoma development designates *MDM2* as a modifier gene in Rb and suggests that enhancement of pRB haploinsufficiency and/or resistance to P53-mediated apoptosis is critical to tumor formation.

## 111

**Germ cell mutations: genetic disease in offspring of cancer survivors in two countries.** *J.J. Mulvihill[1], J.D. Boice, Jr.[2,3], J.F. Winther[4], N. Malila[5], J.H. Olsen[4], P.M. Lahteenmaki[6], L.M. Madanat-Harjuoja[5], R. Sankila[5], M. Stovall[7], E.J. Tawn[8], GCCT Study Group.* 1) Dept Pediatrics & Genetics, OU Medical Ctr, Children's, Oklahoma City, OK; 2) International Epidemiology Institute, Rockville, MD; 3) Department of Medicine, Vaderbilt University, Nashville, TN; 4) Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark; 5) Institute for Statistical and Epidemiological Cancer Research, Finnish Cancer Registry, Helsinki, Finland; 6) Turku University Hospital, Turku, Finland; 7) Department of Radiation Physics, The University of Texas M.D. Anderson Cancer Center, Houston, TX; 8) University of Central Lancashire/Westlakes Research Institute, Cumbria, England.

Despite expectations, no evidence of germ cell mutation, seen as genetic disease and adverse pregnancy outcomes (APO), has been documented in children of survivors from atomic bombs in Japan nor in offspring of survivors of childhood and adolescent cancer. The Genetic Consequences of Cancer Treatment (GCCT) is assessing the risk of genetic disease in 23,889 children born to 14,519 survivors of cancer diagnosed up to the age of 35 years in Denmark and Finland through population-based record linkage. Comparison children include those born prior to cancer diagnosis (n=15,740) and 98,465 children of 45,037 siblings of cancer survivors. Genetic disease and APOs are being evaluated further with dose-response analyses over categories of radiation dose and administered chemotherapy. These findings in population-based studies in two countries are reassuring that the children of cancer survivors are not at high risk of genetic disease apart from the known genetics or familial predispositions.

| | Denmark | | Finland | |
|---|---|---|---|---|
| | Survivors (n=8,759) | Sibling controls (n= 19,393) | Survivors (n=5,760) | Sibling controls (n= 25,644) |
| | No.(%) of offspring with genetic disease | | No.(%) of offspring with genetic disease | |
| Type of genetic condition* | Survivors offspring (n= 13,894) | Sibling offspring (n= 39,455) | Survivors offspring (n= 9,995) | Sibling offspring (n= 59,010) |
| All cytogenetic abnormalities | 65 (0.47%) | 160 (0.41%) | 31 (0.31%) | 120 (0.20%) |
| Definite single-gene (Mendelian) disorders | 77 (0.55%) | 141 (0.36%) | NA | NA |
| Simple malformation | 87 (0.63%) | 321 (0.81%) | 93 (0.93%) | 407 (0.69%) |
| Stillbirths | 26 (0.19%) | 83 (0.21%) | 25 (0.25%) | 118 (0.20%) |
| Neonatal deaths <28 d | 39 (0.28%) | 107 (0.27%) | 21 (0.21%) | 80 (0.14%) |
| Non-hereditary Childhood cancer <15 y | 19 (0.14%) | 54 (0.14%) | 22 (0.22%) | 120 (0.20%) |

*A child with more than one outcome has been counted in the category of genetic condition with the highest genetic component (as listed)

Born after cancer diagnosis

Funded by R01CA104666 NIH/NCI.

## 112

**Parent of origin effects in colorectal cancer age at diagnosis.** *K.G. Rabe[1], G.M. Petersen[1], H. Chen[1], S. Gallinger[2], B. Bapat[2], J. Hopper[3], J. Young[4], M. Jenkins[3], A. Templeton[5], J. Potter[5], P. Newcomb[5], L. LeMarchand[6], R. Haile[7], J. Baron[8], J. Green[9], D. Seminara[10], M.R. Burgio[11], P.J. Limburg[1], S.N. Thibodeau[1], N.M. Lindor[1], Colon Cancer Family Registry.* 1) Mayo Clinic, Rochester, MN; 2) Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada; 3) University of Melbourne, Melbourne, Victoria, Australia; 4) Conjoint Gastroenterology Laboratory, Queensland Institute of Medical Research, Herston, Brisbane, Australia; 5) Fred Hutchinson Cancer Research Center, Seattle, WA; 6) University of Hawaii Cancer Research Center, Honolulu, HI; 7) Human Cancer Genetics Program, University of Southern California, Los Angeles, CA; 8) Dartmouth College of Medicine, Dartmouth, VT; 9) Medical Genetics Program, Memorial University, St. John's Newfoundland, Canada; 10) National Cancer Institute, National Institutes of Health, Bethesda, MD; 11) The Scientific Consulting Group, Inc., Gaithersburg, MD.

Background: Genomic imprinting refers to a parent-of-origin specific effect on gene expression. At least 1% of human genes are modulated in this manner. We sought evidence for genomic imprinting in colorectal cancer by studying the ages at diagnosis in offspring affected with colorectal cancer whose parents were also diagnosed with colorectal cancer. Methods: Affected parent: affected offspring pairs were identified in the Colon Cancer Family Registry (http://epi.grants.cancer.gov/CFR), ascertained from both population-based and clinic-based sources. Families with known Mendelian syndromes were excluded. Because of over recruitment of young-onset cases, these data are not appropriate to assess intergenerational anticipation. Results: 2061 parent-child pairs were included. Differences in colorectal cancer diagnosis mean age was compared between various offspring groups via t-tests; all p-values were two-sided. We found that the affected offspring of fathers were, on average, younger than offspring of affected mothers (55.8 vs 53.7 years; p=0.0003), but when divided into sons and daughters, this difference was seen to be driven entirely by younger age at diagnosis in daughters of affected fathers compared to sons (52.3 years vs 55.1 years; p=0.0004). A younger age at diagnosis in affected daughters of affected fathers was also observed in various subsets, including families that met Amsterdam Criteria and families with documented normal DNA mismatch repair in tumors. Conclusions: Imprinting effects are not expected to be affected by the sex of the offspring. Possible explanations for these unexpected findings include: (1) an imprinted gene on the pseudoautosomal region of the X chromosome; (2) an imprinted autosomal gene that affects a sex-specific pathway; or (3) an X-linked gene unmasked because of colonic tissue specific preferential inactivation of maternal X chromosome.

## 113

**Aspirin prevents cancer in Lynch Syndrome: results of long term follow-up in the CAPP2 trial.** *J. Burn[1], A.-M. Gerdes[1], J.-P. Mecklin[2], F. Macrae[3], G. Moeslein[4], S. Olschwang[5], M.-L. Bisgaard[6], R. Ramesar[7], D. Eccles[8], E.R. Maher[9], L. Bertario[10], A. Lindblom[11], G. Evans[12], J. Lubinski[13], P.J. Morrison[14], J.W.C. Ho[15], H.F.A. Vasen[16], H.T. Lynch[17], J.C. Mathers[18], D.T. Bishop[19], the CAPP2 consortium.* 1) Inst Human Gen, Newcastle Univ, Newcastle on Tyne, United Kingdom; 2) Jyvaskyla Central Hospital, Jyvaskyla, Finland; 3) Royal Melbourne Hospital, Melbourne, Australia; 4) St Josefs Hospital, Bochum-Linden, Germany; 5) Institut Paoli-Calmettes, Marseilles, France; 6) Danish Hereditary Non-Polyposis Colon Cancer Register, Hvidovre, Denmark; 7) University of Cape Town, Observatory, South Africa; 8) Princess Anne Hospital, Southampton,UK; 9) Medical and Molecular Genetics, University of Birmingham, UK; 10) Instituto Nazionale per lo Studio e la Cura dei Tumori, Milan; 11) Karolinska Instituten, Stockholm, Sweden; 12) St Mary's Hospital, Manchester,UK; 13) International Hereditary Cancer Centre, Szczecin,Poland; 14) Belfast City Hospital,Belfast,Northern Ireland,UK; 15) Queen Mary Hospital, Hong Kong, People's Republic of China; 16) Leiden University medical Centre, leiden, Netherlands; 17) Creighton University Medical Centre, Omaha, Nebraska,USA; 18) Human Nutrition Research centre, Institute of Ageing and Health, Newcastle University, Newcastle upon Tyne,UK; 19) Leeds Institute of Molecular Medicine, University of leeds, UK.

CAPP2 recruited 1009 eligible carriers of Lynch syndrome (HNPCC) to a randomised controlled trial of 600mg aspirin and/or 30 g Novelose (resistant starch) in 43 centres worldwide. After a mean of 29 months (range 7-74) there was no difference in colonic neoplasia between those on either active treatment or placebo (Burn et al NEJM 2008;359:2567-2578). All treatment was discontinued at exit colonoscopy. Randomisation category was not divulged to either the participants or the contributing clinicians. In 2008 all participants were advised of the negative result. All but one agreed to forego their right to be told their randomisation category and almost all confirmed their willingness to be followed. Based on 667 of the original cohort and commencing around 4 years from initial randomisation, the incidence of new cancers in the aspirin and placebo groups began to diverge. Up to 4 yrs from randomisation there were 14 colon cancers in the aspirin treated group and 16 in the placebo group despite regular colonoscopy. In the subsequent 3 years there were 3 in the aspirin group and 9 in the placebo group. There have been a total of 21 non-colorectal HNPCC related cancers with a relative risk of 0.38 in the aspirin group (p=0.05) The protective effect appears to persist for over 4 years after the episode of treatment and correlates with the duration of use on trial. The time profile of effect suggests a primary effect on pre-malignant changes, possibly via an impact on apoptosis in aberrant crypt stem cells analagous to the pro-apoptotic effect of salicylates in plants. Plans for a large scale international randomised dose finding study of aspirin in Lynch syndrome will be presented.

## 114

**A shared genetic origin for congenital disorders and testicular tumors revealed by activating mutations in FGFR3 and HRAS.** *A. Goriely[1], R.M.S. Hansen[1], I.B. Taylor[1], I.A. Olesen[2], G.K. Jacobsen[3], S.J. McGowan[4], S.P. Pfeifer[5], G.A.T. McVean[5], E. Rajpert-De Meyts[2], A.O.M. Wilkie[1]*. 1) Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK; 2) Dpt of Growth & Reproduction, Copenhagen University Hospital (Rigshospitalet), Copenhagen, DK; 3) Dpt of Pathology, Copenhagen University Hospital (Rigshospitalet), Copenhagen, DK; 4) Computational Biology Research Group, University of Oxford, Oxford, UK; 5) Dpt of Statistics, University of Oxford, Oxford, UK.

Genes mutated in congenital malformation syndromes are frequently implicated in oncogenesis, but the causative germline and somatic mutations occur in separate cells and at different times in an organism's life. Here we unify these processes for a subset of mutations arising in germ cells during spermatogenesis. These mutations occur spontaneously at high frequency (100-1000x above background rate) and nearly always originate from healthy fathers, who show an increase in age over the population mean (paternal age effect mutations). Examples are found in the genes FGFR2 (Apert, Crouzon and Pfeiffer syndromes), FGFR3 (achondroplasia and Muenke syndrome), HRAS (Costello syndrome), PTPN11 (Noonan syndrome) and RET (multiple endocrine neoplasia types 2A and 2B). We previously investigated the pathological basis of the Apert FGFR2 755C>G mutation using a quantitative assay. We found elevated, but unevenly distributed, mutation levels in the sperm of most men, leading us to propose that original mutational events are rare, but the mutations become progressively enriched because they confer a selective advantage leading to clonal expansion over time (Goriely et al, 2003 Science 301, 643). Here, we have extended this work by asking whether these paternal age effect mutations could contribute to the formation of tumors in the testis. We screened 30 spermatocytic seminomas for oncogenic mutations in 17 genes and identified 2 mutations in FGFR3 (both 1948A>G encoding K650E, which causes thanatophoric dysplasia in the germline) and 5 mutations in HRAS (3 encoding Q61R and 2 encoding Q61K). HRAS mutations are commonly found in Costello syndrome, but no Q61 mutation has been reported, probably due to fetal lethality. We used a novel application of massively parallel sequencing to show that the FGFR3 1948A>G mutation is present in the sperm of most men where it attains levels above 10-4 and accumulates with advancing age. The mutation spectrum observed at the K650 codon in sperm closely correlated with the distribution of germline disorders and with that of mutations found in bladder cancer. Most spermatocytic seminomas showed increased immunoreactivity for one or both of FGFR3 and HRAS. We propose that paternal age effect mutations activate a common "selfish" pathway promoting proliferation in the testis and leading in the next generation to diverse phenotypes ranging from fetal lethality, through congenital syndromes to complex disease predisposition, notably cancer.

## 115

**Structural bioinformatics mutation analysis reveals genotype-phenotype correlations and suggests molecular mechanisms in VHL disease.** *J.R. Forman[1,2], C.L. Worth[3,2], G.R.J. Bickerton[4,2], T.G. Eisen[5], T.L. Blundell[2]*. 1) Structural Bioinformatics Unit, Pasteur Institute, Paris, France; 2) Biocomputing Group, Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK; 3) Structural Bioinformatics Group, Leibniz-Institut fur Molekulare Pharmakologie, Campus Berlin-Buch, Robert-Roessle-Str. 10, 13125 Berlin, Germany; 4) Biological Chemistry and Drug DIscovery, College of Life Sciences, University of Dundee, Dundee DD1 5EH, UK; 5) Department of Oncology, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK.

**Background** Mutations in the VHL gene lead to von Hippel-Lindau (VHL) disease, a clinically heterogeneous cancer syndrome. The protein product, pVHL, is known to interact with elongin B, elongin C, and the HIF substrate. Mutations in pVHL are linked with retinal angioma, CNS hemangioblastoma, renal cell carcinoma, and pheochromocytoma. Our aims are to understand the effects of missense mutations on the pVHL protein, considering both its structure and interactions, and thereby understand the genotype-phenotype relationships observed in patients. **Methods** We have used software and database tools (SDM, Piccolo, Crescendo) to predict the changes in structural stability and interactions of pVHL upon pathogenic point mutations. By analyzing known and predicted interaction sites and predictions of thermodynamic stability change upon mutation, we generate new hypotheses regarding the molecular aetiology of renal cell carcinoma (RCC) and pheochromocytoma in VHL disease. With this information, we can understand and predict the phenotypes associated with missense mutations in the VHL gene. **Results** We find that the molecular causes of RCC and pheochromocytoma appear to be decoupled. RCC may arise through two distinct mechanisms: disruption of HIF interactions or binding at the elongin B interface. Pheochromocytoma is triggered by mutations which disrupt interactions at the elongin C binding site. **Conclusions** These findings have important implications for VHL disease and for non-familial RCC, since most cases of clear cell RCC are linked with VHL inactivation. Additionally, predicting effects of genetic variation is critical as genetic sequencing accelerates; the analytical strategy presented here can be used to elucidate other cases of genetic variation.

## 116

**Clinical Predictors and Algorithm for the Genetic Diagnosis of Pheochromocytoma Patients - A Cost Reduction Strategy.** *Z. Erlic[1], L. Rybicki[2,3], G. Opocher[4], A. Januszewicz[5], C. Eng[3,6], H.P.H. Neumann[1], European-American Pheochromocytoma Study Group.* 1) Nephrology-Preventive Medicine, Medical University Clinic, Freiburg i.Br., Germany; 2) Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH; 3) Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH; 4) Veneto Institute of Oncology and Department of Medical and Surgical Sciences, University of Padova, Italy; 5) Department of Hypertension, Institute of Cardiology, Warsaw, Poland; 6) Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH.

**Background:** Six pheochromocytoma susceptibility genes (*NF1*, *VHL*, *RET*, *SDHB*, *SDHC*, *SDHD*) causing distinct syndromes have been identified; ~1/3 of all patients presenting with pheochromocytoma carry a predisposing germline mutation. This mutation frequency mandates offering gene testing to all pheochromocytoma patients. When four major genes are analyzed (*VHL*, *RET*, *SDHB*, *SDHD*) in a clinical laboratory, costs are ~US$3000/patient. The aim of the study is to systematically obtain a robust algorithm to identify who should be genetically tested, and to determine the order in which genes should be tested.

**Methods:** DNA from 989 apparently non-syndromic patients presenting with symptomatic pheochromocytoma were scanned for germline mutations in the susceptibility genes. Clinical parameters were analyzed as potential predictors for finding mutations by multiple logistic regression, validated by bootstrapping. Cost reduction was calculated between prioritized gene testing compared to that for all genes.

**Results:** Of 989 apparently non-syndromic pheochromocytoma cases, 187 (19%) harbored germline mutations. Predictors for presence of mutation are age≤45, multiple pheochromocytoma, extra-adrenal location, and previous head and neck paraganglioma (HNP). If we used the presence of any one predictor as indicative of proceeding with gene testing, then 342 (34.6%) patients would be excluded, and only eight (0.8%) potential mutation carriers would be missed. We were also able to statistically model which gene(s) to begin testing and priority of genes to be tested given certain clinical features. E.g., for patients with prior history of HNP, the priority for testing would be *SDHD*>*SDHB*>*RET*>*VHL*. Using the clinical predictor algorithm to prioritize gene testing and order, a 44.4% cost reduction in diagnostic process can be achieved.

**Conclusion:** Clinical parameters can predict for mutation carriers and help prioritize gene testing to reduce costs in non-syndromic pheochromocytoma presentations.

## 117

**X-chromosome disorders: Identification of underlying mechanisms.** *P. Patsalis[1], G. Koumbaris[1], D. Rajan[2], T. Fitzgerald[2], S. Gribble[2], S. Clayton[2], H. Hatzisevastou[3], A. Kurg[4], S. Kitsiou-Tzeli[5], N. Scordis[6], Z. Kosmaidou[7], J. Vermeesch[8], I. Georgiou[9], N. Carter[2].* 1) The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus; 2) Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, United Kingdom; 3) Ippokratio Thessaloniki General Hospital. Thessaloniki, Greece; 4) Department Of Biotechnology IMCB, University of Tartu, Tartu, Estonia; 5) Agia Sophia Children's Hospital, Athens, Greece; 6) Department of Paediatrics, Makarios Hospital, Nicosia, Cyprus; 7) Department of Genetics, Alexandra Hospital, Athens, Greece; 8) University of Leuven, Leuven, Belgium; 9) University of Ioannina, Ioannina, Greece.

Non-allelic homologous recombination (NAHR), nonhomologous end joining (NHEJ), and recently Fork Stalling and Template Switching (FoSTeS) have been implicated as the main mechanisms for the creation of genomic disorders. In order to investigate the mechanisms responsible for the creation of X-chromosome disorders we analyzed 70 cases bearing cytogenetically visible X-chromosome abnormalities using whole genome tiling path BAC arrays and custom designed targeted ultra-high resolution oligo-arrays. Using whole genome tiling path BAC CGH, we were able to accurately map the breakpoints of 35 cases bearing isochromosomes of the long arm of chromosome X at a resolution of 150kb. Fifteen of these had breakpoints at the centromere and were considered monocentric. The remaining 20 were isodicentric and had breakpoints in proximal Xp in ChrX:51500000-58500000. This region of chromosome X is rich in segmental duplications and contains some of the largest and most homologous inverted repeats in the human genome. Based on the BAC-array CGH findings, we designed custom oligo-arrays which cover this region in ultra-high resolution (44K and 385K oligos in 7Mb region of interest) and feature enhanced coverage of segmental duplications. Screening the isodicentric cases with these ultra-high resolution arrays enabled us to identify previously undiscovered breakpoint complexity in 45% of the isodicentrics and demonstrate that they are formed by NAHR, facilitated by specific highly homologous inverted repeats. Twenty two percent of the isodicentrics were mapped within repetitive sequences and 33% have simple breakpoints that do not coincide with segmental duplications and are probably mediated by a nonhomologous recombination mechanism.

## 118

**New tools in the cost-efficient detection of pathogenic chromosomal aberrations in patients with MR/MCA by array-CGH.** *C.M.A. van Ravenswaaij-Arts, G.B. van der Vries, T. Dijkhuizen, Y. Swart, H. Alkema, H. Zorgdrager, B. Sikkema-Raddatz, R. Sinke, K. Kok.* Dept Genetics, UMC Groningen, Groningen, The Netherlands.

Array-based comparative genomic hybridization has become an indispensable tool in the hunt for submicroscopic chromosomal aberrations in patients with multiple congenital anomalies and/or mental retardation (MCA/MR). As anticipated with the increasing resolution of this technique the discovery rate of causally related aberrations has increased by 10%-15% compared to microscopic karyotyping. However, the discovery of neutral copy number variants has complicated the interpretation of the array results, especially since it has become clear that these variants may constitute well over 5 % of the genome. Thus, there is a need for procedures that efficiently distinguish pathogenic from neutral (benign) CNVs. An important step in this process is distinguishing inherited CNVs, that are more likely to be neutral, from the de novo CNVs, that are more likely to be pathogenic. In 2008, we have implemented an oligo-array platform (Agilent, Santa Clara, USA) in combination with a custom designed 2x105K array (L. Connell, Oxford) for the postnatal screening of patients with MR/MCA. For an efficient interpretation of the patient data, we have decided to perform trio analyses. On one array the patient is hybridized to a reference sample (a pool of either 40 males or 40 females). On the second array of the same slide, the parents of the patient are hybridized with opposite dyes. To facilitate the individual analysis of each parent, separate 2logR files for the parents are subsequently generated using a home made software package that creates export files that can be uploaded into several commercial data visualization and analysis platforms. By this procedure, about 85% of all aberrations detected by the patient could directly be traced back to either of the parents. This approach thus constitutes a cost-efficient and fast way to determine the de novo nature of the aberrations that are seen in the patient. The data obtained for 400 healthy parents has been used to generate a database of local CNVs. The UMCG database now contains 3200 entries. 190 CNVs occur at least twice in the database. This database has proven to be of high value in the analysis of individual patients for whom both parents were not available. Recently we have started using a self-made custom design 4x180K array. An adaptation of our software has enabled us to analyse 6 patients on one 4-plex slide, making our diagnostic tests even more cost-efficient.

## 119

**Uncovering complexity by microarray-based cytogenetics and FISH visualization: What have we been missing?** *L.G. Shaffer, B.A. Bejjani, B.A. Torchia, R. Schultz, T. Sahoo, A. Lamb, N. Neill, B.C. Ballif.* Signature Genomic Laboratories, Spokane, WA.

Since the advent of banding, classically trained cytogeneticists have rarely observed complex chromosome abnormalities because the limitations of banding permit the identification of only gross rearrangements. Recently, microarray-based cytogenetics has uncovered complexity seldom imagined at resolutions never before possible. We tested more than 35,000 patients and identified complex rearrangements involving deletions and duplications of the same chromosome arm and insertions at rates not predicted by chromosome banding (2.15% and 1% of abnormalities, respectively). Indispensable to this process is the use of FISH to visualize the rearrangements. Although microarrays identify changes in DNA copy number, they cannot reveal the specific rearrangements responsible for these changes or the mechanisms of their formation. Therefore, the information is incomplete and counseling is not accurate in the absence of a clear understanding of these mechanisms. For example, a DNA copy number gain by array analysis can result from a duplication, unbalanced translocation, marker chromosome or insertion. Each of these abnormalities has specific implications for recurrence risks. Our experience with the clarification of DNA copy changes in over 40 insertions and 160 complex rearrangements by FISH confirms the necessity of FISH visualization. In most cases, the insertions were "one-way" insertions of chromatin from one chromosome into another. However, FISH in one case with two apparently independent alterations by microarray analysis confirmed a deletion of one chromosome and revealed an insertion of another segment at the deletion site. In the cases for which parental samples were available, 58% of insertions were inherited from carrier mothers, 19% were inherited from carrier fathers, and 23% were de novo. Although the majority of insertions were unbalanced, many of which were causative of the child's abnormal phenotype, a recurrent insertion (Y;18) was identified as a benign variant in three families. Such complexity is unlikely to be revealed by chromosome banding or FISH alone, or by microarray analysis followed by MLPA or other molecular tests. Thus, FISH is essential to the proper visualization of the nature of the rearrangement and necessary for proper genetic counseling.

## 120

**Elucidation of Chromosome Structure: Lessons From Molecular Cytogenetic Technologies.** *S. Schwartz[1], R.D. Burnside[1], K. Deak[2], I. Gadi[1], V. Jaswaney[1], R. Pasion[1], H. Risheg[1], J.H. Teperberg[1], B. Williford[1], P.R. Papenhausen[1].* 1) Laboratory Corporation of America, Research Triangle, NC; 2) University of Chicago, Chicago, IL.

To better understand the underlying chromosome structure of aberrations, we have studied over 1600 chromosome abnormalities identified prospectively by SNP array analysis and approximately 200 abnormalities, identified initially by standard cytogenetic analysis, which were then studied by array and FISH analysis. These results show that there is much greater complexity of the abnormalities, with approximately 15% of the prospective group and 35% of the retrospective groups demonstrating increased alterations. The implications of these results are important and provides insight into the formation of abnormalities including: (1) while both groups show an increase in complexity, the frequency of the complexity in previously identified abnormalities is greater, providing more evidence for the need to study these with array analysis; (2) in both groups a higher than expected number of individuals have two or more abnormalities accounting for increased variability in phenotypes; (3) the frequency of duplications and deletions are similar, however a much smaller percentage of duplications appear to be facilitated by LCRs suggesting that these are not ascertained due to milder phenotypes; (4) the majority of duplications are direct tandem (~86%), not inverted tandem; (5) a greater than expected frequency (~40%) of retrospectively studied terminal deletions are not terminal and very few ascertained prospectively are terminal deletions; (6) although deletions and duplications facilitated by LCR are over-represented (~22%), the vast majority of abnormalities detected are unique; (7) a group of the abnormalities, while not facilitated by LCRs, have similar homology in their breakpoint regions; (8) new mechanisms responsible for abnormalities have been identified including discontinuous deletions and duplications, and duplication of chromosomal material from a non-adjacent region in the precise area where a deletion has occurred; (9) rings and marker chromosomes have been found to be formed by a large variety of underlying mechanisms: duplication of material; multiple different sized rings from the same chromosome in an individual; rings composed of material from two different chromosomes; and rings formed from non-continuous segments from one chromosome. These studies clearly show how a combination of array, FISH and cytogenetics has provided considerable knowledge of genetic imbalance, although the mechanism of underlying chromosome formation is elusive.

## 121

**Nonrecurrent genomic duplications of *EHMT1*, a histone methyltransferase, are associated with cognitive impairment and autism spectrum disorders: characterization of the 9q34.3 microduplication syndrome.** *S.A. Yatsenko[1], E.K. Brundage[1], A. Patel[1], S.W. Cheung[1], D.A. Scott[1], C.P. Schaaf[1], A.C. Chinault[1], J.R. Lupski[1,2,3].* 1) Molecular & Human Genetics, Baylor Col Medicine, Houston, TX; 2) Pediatrics, Baylor College of Medicine; 3) Texas Children Hospital, Houston, TX.

Constitutional submicroscopic deletions of the distal long arm of chromosome 9, encompassing the *EHMT1* (euchromatic histone methyltransferase 1) gene, or loss-of-function mutations in *EHMT1*, are associated with the 9q34.3 microdeletion syndrome. Our recent studies have revealed a high incidence of genomic rearrangements within the subtelomeric 9q34.3 region including terminal and interstitial microdeletions, duplications, triplications, derivative chromosomes and complex rearrangements. By array comparative genomic hybridization (aCGH), we identified 9q34.3 microduplications/triplications in nine patients with variable phenotype. The most common clinical features include psychomotor retardation, speech and developmental delay, behavioral problems, autism spectrum disorder, and mild dysmorphic features. The 9q34.3 microduplications range from 200 kb to 4 Mb in size with unique location of the breakpoints in each patient. Seven patients had intrachromosomal tandem rearrangements; one had a triplication with an inverted middle segment. One male patient had partial 9q34.3 trisomy due to a cryptic insertion of the 9q segment into the long arm of chromosome X. Comparison of the clinical features and molecular analyses demonstrate that increased dosage of *EHMT1* appears to be responsible for the neurodevelopmental impairment, speech delay, and autism spectrum disorders in these patients. These findings emphasize the potential significance of *EHMT1* gene dosage and consequently the degree of histone lysine methylation for brain development and functions. The identification and analysis of additional cases will help delineate other dosage sensitive genes within the 9q34.3 region responsible for other clinical features of the 9q34.3 microduplication syndrome.

## 122

**Molecular Analysis and Clinical Correlations in Ring Chromosome 20 Syndrome.** *N.B. Spinner[1,2], L.K. Conlin[1], B.D. Thiel[2], P.S. Munoz[1], W. Kramer[1], J.T. Glessner[3], H. Hakonarson[1,3], S.A Hosain[4].* 1) Dept of Pediatrics, Children's Hosp, and The University of Pennsylvania School of Medicine,Philadelphia, PA; 2) Dept of Pathology, The Children's Hospital of Philadelphia and The University of Pennsylvania School of Medicine; 3) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 4) Dept of Neurology, UMDNJ Robert Wood Johnson School of Medicine, New Brunswick, NJ.

Ring chromosome 20 syndrome is diagnosed by identification of a ring chromosome 20 on cytogenetic analysis, and characterized clinically by medically intractable epilepsy, behavioral problems and mild mental impairment. Unlike other chromosomal aberrations, dysmorphic features or major congenital anomalies are rarely reported. We analyzed the ring chromosome 20 from 18 patients (10 males and 8 females) by cytogenetic, molecular cytogenetic and molecular methods and correlated the molecular characteristics with clinical features. The patients formed two distinct groups. Thirteen of the 18 patients had a mosaic karyotype, (mos 46,XX(Y), r(20)). The percentages of cells with the ring varied with 4 to 55%. In these patients, FISH using subtelomeric probes for both the short and long arms of chromosome 20 did not reveal a deletion. Genomic analysis using a genome-wide SNP array platform confirmed there were no deletions along chromosome 20. The mosaic nature of these rings suggests a postzygotic origin and formation appears to be occurring by fusion of the telomeric regions with no loss of subtelomeric DNA. Five patients had a non-mosaic ring (46,XX(Y), r(20)). FISH analysis revealed that at least one of the subtelomeres was deleted in every patient. SNP array analysis revealed that all patients had deletions of the 20q subtelomere regions (55kb to 1.6Mb) and 2 patients also had 20p deletions (2 Mb and 322 kb). These ring chromosomes appear to form meiotically. We analyzed these two groups with respect to age of seizure onset and frequency of seizures. Our data indicates that the two groups varied significantly with respect to age of onset of seizures. The mosaic patients had an age of onset of 6.7 years, while the non-mosaic patients had an average age of onset less than 1 year. The non-mosaic (and deleted) patients also had additional findings including developmental delay, short stature, and dysmorphic features. The two groups did not differ with respect to average number of seizures per week. These studies indicate that the ring chromosome 20 syndrome is heterogeneous, with respect to both clinical and genomic features. We suggest that the mechanism of ring formation is different in these two groups. Further studies are aimed at understanding the cellular consequences of ring formation, the impact on gene expression and the mechanism of seizures, given that there is no loss of genomic content in the majority of patients.

## 123

**Identifying novel genetic conditions by screening for copy number changes in functionally significant genes.** *B.C. Ballif, B.A. Bejjani, L.G. Shaffer.* Signature Genomic Laboratories LLC, Spokane, WA.

Microarray technology has dramatically improved the cytogeneticist's ability to detect and precisely define DNA copy number gains and losses throughout the genome. However, the current trend toward increasingly higher density whole genome array screening often yields an abundance of copy number changes of unclear clinical significance which confound the interpretation of results in a diagnostic setting. Although the benefits of increased coverage are clear to a degree, one approach to whole genome copy number screening is to scrutinize the copy number status of functionally significant genes in the genome, such as transcription factors and genes in developmental pathways, as candidate loci for which abnormal dosage may be more likely to result in an abnormal phenotype than dosage differences for other genes in the genome. As proof of concept, we used this approach and included coverage over *DISP1*, a gene in the sonic hedgehog signaling pathway, and discovered the 1q41q42 microdeletion syndrome. Likewise, increased coverage over the *FOX* transcription factor gene cluster on 16q24.1q24.2 led to the recent identification of microdeletions resulting in alveolar capillary dysplasia and other malformations. In our diagnostic laboratory we have designed whole genome BAC and oligonucleotide microarrays with increased coverage over >500 functionally significant genes. Microarray analysis of >35,000 cases has identified more than 700 chromosome abnormalities which overlap these functionally significant genes and that do not overlap with known genetic conditions. Approximately 50% of these are < 2Mb in size and ≈30% are under 500 kb. Although 67% of these abnormalities are of unknown origin, at least 15% (110 cases) are de novo. Many of these may represent novel genetic disorders. The finding of small alterations over these clinically important genes supports a non-arbitrary cut off for microarray use and the utilization of databases and genetic content in the interpretation of genomic gains and losses. Overall, the inclusion of developmental pathway genes has identified several new microdeletion syndromes and supports the 'genotype-first' approach to uncovering novel genetic disorders.

## 124

**Acquired chromosome abnormalities in the lungs of patients with Pulmonary Arterial Hypertension: evidence for a two-hit hypothesis.** *M.A. Aldred[1], S.A. Comhair[1], M. Varella-Garcia[2], K. Asosingh[1], W. Xu[1], G.P. Noon[3], P.A. Thistlethwaite[4], R.M. Tuder[2], S.C. Erzurum[1], M.W. Geraci[2], C.D. Coldren[2].* 1) Cleveland Clinic, Cleveland, OH; 2) University of Colorado Denver, Aurora, CO; 3) Baylor College of Medicine, Houston, TX; 4) University of California, San Diego, CA.

Pulmonary arterial hypertension (PAH) is characterized by progressive vascular remodeling in which proliferation and migration of endothelial and smooth muscle cells leads to obliterative vascular lesions. This results in a sustained elevation in pulmonary artery pressure that can potentially lead to death due to right heart failure. Previous studies have indicated that the endothelial cell proliferation in PAH is quasi-neoplastic, with evidence of monoclonality and instability of short DNA microsatellite sequences. To assess whether there is larger scale genomic instability, we utilized SNP microarrays to perform genome-wide copy number analysis on pulmonary artery endothelial (PAEC) and smooth muscle cells isolated from the lungs of PAH patients. Mosaic chromosomal abnormalities were detected in five of nine PAEC cultures from PAH lungs and zero of four controls. Interphase fluorescent in situ hybridization analysis on tissue sections from explant lungs confirmed the presence of these abnormalities *in vivo* in two of three cases tested. One patient harbored a germline mutation of *BMPR2*, the primary genetic cause of PAH, and also a somatic loss of chromosome-13 in PAEC, which constitutes a second hit in the same pathway by deleting Smad-8. In two female cases with acquired mosaic loss of the X-chromosome, methylation analysis showed that the active X was deleted. Remarkably, one also showed completely skewed X-inactivation in the non-deleted cells, suggesting the PAEC population was clonal prior to the acquisition of the chromosome abnormality. Our data indicate a high frequency of genetically abnormal sub-clones within the lung vessels of patients with PAH and provide the first definitive evidence of a second genetic hit in a patient with a germline *BMPR2* mutation. We propose that these chromosome abnormalities may confer a growth advantage and thus contribute to the progression of PAH.

## 125

**Impact of miRNA in Osteosarcoma: an integrated analysis of genomic, expression and miRNA profiles.** *G. Maire[1], S. Chilton-MacNeill[2], P.S. Thorner[3], M. Zielenska[2,3], J.A. Squire[1].* 1) Pathology & Molecular Medicine, Queen's University, Kingston, ON, Canada; 2) Genetics and Genome Biology, Hospital for Sick Children, Toronto, ON, Canada; 3) Department of Pediatric Laboratory, Medicine and Pathobiology, The Hospital for Sick Children, Toronto, ON, Canada.

Osteosarcoma (OS) is an aggressive sarcoma of the bone that is characterized by complex and composite karyotype. The level of genomic complexity and heterogeneity of this tumor prevented the identification of recurrent simple chromosomal abnormalities or gene rearrangements, like in many other sarcomas. We have shown that some regions are more prone to harbor genomic imbalances and chromosomal breaks: 1p35-p36; 6p12-21; 8q24; 17p11-p12 and 19p13. However, due to this genomic instability, no prognosis biomarker has been identified in OS yet. In addition to the previously published genomic and expression profiles, we integrated a new layer of gene expression regulators: the MicroRNA (miRNA). It is thought that miRNA, a class of small non coding RNA, are involved in many biological processes, including oncogenesis. The expression profile of 723 human miRNA was established for a series of 7 OS. It showed that less than 5% of the miRNA (38) were differentially expressed compare to osteoblasts. Most of them were underexpressed (28): only 10 miRNA exhibited an overexpression. The gene copy profiles for the matching samples were integrated to the new miRNA profiles: only 9 miRNA showed a positive correlation between expression, and locus copy number. In OS, the gene copy number seems to play little role in the regulation of miRNA expression. The in-silico analysis also included the identification of miRNA target genes (Miranda, TargetScan, PicTar). Co-expression of miRNA and the target gene pair was established by integrating the available data set for OS expression profiles with the miRNA profiles. The gene copy number status for these target genes was also compared with their expected up or down-regulation (by the differentially expressed identified miRNA). By this mean we identified a list of genes regulated either by miRNA (GADD45A, WIF1...) or by other means such as gene copy number (FOS, PLK2...). This integrated in-silico approach of genomic/ expression/ miRNA profiling allowed us to identified: i) the miRNA OS signature; ii) the mechanism of miRNA regulation in OS; iii) sets of genes regulated by miRNA; iv) sets of genes regulated by gene copy number. This study represents a new step into the comprehension of the OS oncogenesis, and the identification of new therapy target and potential prognosis biomarkers.

## 126

**Detailed characterization of Fanconi Anemia and non-FA head and neck cancer using molecular cytogenetics and high resolution array CGH.** *H. Tonnies[1], S. Vahs[1], E. Klopocki[2], D. Nuvemann[3], A. Gerlach[4], I. Vater[1], J. Richter[1], J. Weimer[5], R. Siebert[1], A. Caliebe[1].* 1) Institute of Human Genetics, University Hospital Schleswig-Holstein, Kiel, Germany; 2) Institute of Medical Genetics, Charité Universitätsmedizin Berlin, Berlin, Germany; 3) Department of Oral and Maxillofacial Surgery, University Hospital Schleswig-Holstein, Christian-Albrechts University Kiel, Kiel, Germany; 4) Institute of Human Genetics, Charité Universitätsmedizin Berlin, Berlin, Germany; 5) Clinic of Obstetrics and Gynecology, Laboratory of Oncology, University Hospital Schleswig-Holstein, Christian-Albrechts University Kiel, Kiel, Germany.

Fanconi anaemia (FA) is a rare inherited chromosome instability disorder with congenital abnormalities, a high risk for acute myeloid leukemia (AML) and solid tumors. The most common life-threatening early event in Fanconi anemia (FA) patients is bone marrow (BM) failure, which typically develops during the first decade of life. In its course, many FA patients acquire clonal chromosomal aberrations in bone marrow (BM) cells. Recently, we were able to develop a sensitive single cell based FISH approach to detect these adverse clonal aberrations early in BM and peripheral blood cells of FA patients resulting in immediate bone marrow transplantation (BMT). Head and neck cancers are among the most common solid tumors to develop in FA patients (relative risk 700-fold). The frequency, age of onset, and clinical course of these tumors are different from non-FA patients. Additionally BMT enhances the risk to develop HNSCC in FA patients. Open questions are why FA patients preferentially and so early develop HNSCC and which role the FA/BRCA pathway genes play in the initiation of HNSCC. Up to now, molecular cytogenetic and array based data for FA-HNSCC are not available in the literature. We investigated three extremely rare FA-HNSCC cell lines published by van Zeeburg et al. (2007) and HNSCC of young non-FA patients using different molecular cytogenetic and array based techniques including SKY-FISH, locus-specific FISH, conventional CGH and BAC-, SNP- and high resolution oligo arrays (NimbleGen 385k, Agilent 244k). Additionally to the detection of genetic imbalances, we addressed the question of gene silencing by epigenetic changes. Our data show that most (large) chromosomal imbalances detectable in FA-HNSCCs are comparable to non-FA HNSCC (e.g. gain of 3q and loss of 8p). Additionally, we detected small hetero- and homozygous deletions of CSMD1 and SMAD4, of which the last gene is presumably affecting the TGF-beta-Smad signaling pathway. High-level amplifications next to breakpoints of gained and deleted chromosomal segments as e.g. in chromosomal bands 4p13 and 11q22.3 could be detected for the first time in FA-HNSCC. Hypermethylated regulatory sequences presumably leading to reduced gene expression or silencing could be detected for MGMT, CCND2, and p16 in FA-HNSCC. Main goal of our studies is to establish sensitive single cell approaches for the early detection of primary lesions in the tumorigenesis of HNSCC in Fanconi anemia patients.

## 127

**Rapid testing versus karyotyping in Down syndrome screening programs: Cost effectiveness and missed clinically significant chromosome abnormalities.** *J. Gekas[1,2], D.G. van den Berg[3], A. Durand[4], M. Vallée[2], H.I.J. Wildschut[3], E. Bujold[5], J.C. Forest[6], F. Rousseau[7], D. Reinharz[4].* 1) Centre de recherche du CHUQ, Service de Génétique Médicale, Unité de Diagnostic Prénatal, Université Laval, Québec City, Québec, Canada; 2) Laboratoire de Cytogénétique, Centre hospitalier universitaire de Québec (CHUQ), Québec City, Québec, Canada; 3) Department of Obstetrics and Gynecology, Erasmus University Medical Center, Rotterdam, the Netherlands; 4) Département de médecine sociale et préventive, Laboratoire de Simulations des Dépistages, Faculté de Médecine, Université Laval, Cité universitaire, Québec City, Québec, Canada; 5) Département d'Obstétrique-Gynécologie, Faculté de Médecine, Université Laval, Québec City, Québec, Canada; 6) Centre de recherche du CHUQ, Département de biologie médicale, Faculté de Médecine, Université Laval, Québec City, Québec, Canada; 7) CanGèneTest Research consortium, Centre de recherche du CHUQ, Département de biologie médicale, Faculté de Médecine, Université Laval, Québec City, Québec, Canada.

About 80% of antenatal karyotypes are generated by Down's syndrome screening programs (DSSP). Currently, there are 6 DSSP that respect US and Canadian guidelines. Reliable molecular methods (FISH and PCR) can detect common aneuploidies, are faster and less expensive than karyotyping. They are referred to as rapid aneuploidy diagnosis (RAD). In the UK, RAD is recommended as a stand alone approach in new DSSP, whereas the US recent guidelines recommend that RAD be followed up by karyotyping. A detailed cost effectiveness (CE) analysis of RAD in various DSSP is actually lacking. Moreover, there is a debate around the significance of chromosome abnormalities (CA) detected with karyotyping and not with RAD. Objectives: To compare the estimated CE of RAD vs karyotyping according to different DS screening options and to evaluate the clinically significant missed CA. Methods: We performed computer simulations to compare multiple screening options followed by FISH, PCR or karyotyping. This yielded 24 screening algorithms evaluated using a population of 110,948 registered pregnancies in the province of Quebec in year 2001. Measures: CE ratios for DS cases, amount of missed CA with background, low or high risk of an abnormal outcome and incremental cost effectiveness ratios (ICER). Results: All screening strategies were most cost effective when using PCR, while the detection rate for DS was the same. The most cost-effective strategy was contingent screening associated with PCR (CE ratio of $24,084 per DS). Using a full karyotype, the CE ratio increased to $27,898. The CE ratio of the least cost effective screening strategy (combined test using karyotyping) was $55,412. The contingent screening method associated with PCR missed per 100.000 pregnancies only 6 clinically significant CA of which only 1 was expected to confer a high risk of an abnormal outcome. Depending on which screening option used, the ICER to find the CA missed by RAD ranged from $59,034 to $128,376 per CA. These costs are much higher than the ones to detect DS cases. Conclusions: RAD, in particular PCR, are the most cost-effective diagnostic tools. While FISH misses the same number of clinically significant CA as PCR, it is associated with higher cost. Additional costs to detect clinically significant CA missed by PCR are substantial. Since the DSSP are mainly designed for DS detection, it may be relevant to question the additional costs of a full karyotype.

## 128

**Differential gene expression in fetuses with trisomy 18: a new approach to understanding pathophysiology.** *K. Koide[1], D. Slonim[2], K.L. Johnson[1], U. Tantravahi[3], J.M. Cowan[1], D.W. Bianchi[1].* 1) Pediatrics, Floating Hospital for Children at Tufts Medical Center, Boston, MA; 2) Computer Science, Tufts University, Medford, MA; 3) Pathology, Women and Infants' Hospital, Providence, RI.

**Objective.** As a novel means of identifying pathophysiologic changes in fetuses with trisomy 18 (T18), we characterized developmental gene expression using cell-free mRNA in amniotic fluid (AF) samples. **Methods.** We extracted cell-free fetal mRNA from 11 residual second trimester AF samples following clinically indicated testing. Five were from women carrying singleton female fetuses with 46, XX, +18, and 6 were carrying euploid female fetuses. mRNA extraction, cDNA synthesis, fragmentation, biotin labeling and hybridization were performed as previously described (Slonim and Koide et al., Proc Natl Acad Sci USA 2009; epub May 27). Initial analysis of euploid and T18 samples was done using the Affymetrix Gene Chip Microarray Suite 5.0, followed by comparative analysis using t-tests. We assessed statistical significance as p values <0.05 after Benjamini-Hochberg adjustment for multiple testing. Individual differentially-expressed genes were further examined by DAVID tools (Database for Annotation, Visualization, and Integrated Discovery) and Ingenuity® pathway analysis. **Results.** In the individual gene set 230 genes were up-regulated (64.7%) and 126 were down-regulated (35.4%) in the T18 fetuses (total=356). Only 7/356 genes were located on chromosome 18. DAVID analysis implicated the following functions as significantly different in the affected fetuses: ion and cation transport, cytoskeletal and cell adhesion proteins, zinc finger proteins, notch signaling, neurological development, and immunoglobulin-like proteins and t-cell receptors. Ingenuity analysis linked the differentially-regulated genes to endocrine function and lipid metabolism (up-regulated), respiratory disease (down-regulated), and organismal injury and abnormalities. **Conclusions.** Residual AF samples provide a source of mRNA for differential gene expression analysis in T18 fetuses. It has long been assumed by medical geneticists that proteins produced by the additional copy of chromosome 18 are almost exclusively responsible for the atypical development and function associated with this condition. It was unexpected that the molecular abnormalities we observed in T18 fetuses are predominantly produced by genes on the other chromosomes. This discovery-driven approach suggests new avenues to further understand the fetal pathophysiology of T18.

## 129

**Rapidly evolving genes as candidate genes for preterm birth.** *J. Plunkett[1,2], S. Doniger[1], I. Borecki[1], H. Puttonen[3], M. Hallman[4], R. Haataja[4], A. Luukkonen[4], K. Teramo[3], J. Fay[1], L. Muglia[2].* 1) Washington Univ, St. Louis, MO; 2) Vanderbilt Univ, Nashville, TN; 3) University of Helsinki, Helsinki, Finland; 4) University of Oulu, Oulu, Finland.

Despite the important influence of birth timing on infant health, the process of parturition remains largely uncharacterized. While some important physiological antecedents of labor have been identified in model organisms, such signals do not seem to precede human labor. High rates of divergence in human pregnancy-related genes compared to other primates support the notion that human gestation length has been shortened to accommodate a larger brain/body ratio and a narrower birth canal. This evidence suggests that the set of genes rapidly evolving on the human lineage includes genes that play important roles in regulating parturition and potentially influence preterm birth risk. We identified genes showing evidence of rapid evolution from 2 comparative genomic screens examining highly conserved elements and orthologous coding sequences from dog, cow, chimp, human, macaque, mouse, and rat. To minimize the number of tests and retain more power to detect small effects, we selected a subset of genes likely to be involved in parturition, based on expression and functional information, to use as candidate genes for preterm birth. We also examined duplicated genes expressed in pregnancy-related tissues and biological candidate genes showing evidence of rapid evolution for a total of 175 genes. We selected over 11,000 SNPs within the genes' regions, as available on the Affymetrix® Genome-Wide Human SNP Array 6.0, for analysis in 166 Finnish preterm and 131 control mothers. In genes with enrichment for suggestive evidence of SNP association ($p<0.01$), we genotyped the 41 SNPs augmented with 105 additional SNPs chosen to increase coverage of the relevant regions. 74 SNPs showing suggestive association ($p<0.1$) in this analysis were then genotyped in two independent cohorts of US White (147 preterm, 157 control) and US Hispanic (73 preterm, 292 control) mothers. One SNP, rs8110925, in the US Hispanics and one in the US Whites, rs11564620, were significant after correcting for 74 tests ($p<0.0007$), both located in *PLA2G4C*, encoding a phospholipase involved in prostaglandin synthesis. Another SNP in *PLA2G4C* rs1366442 showed association in all three populations, with pooled p-values of 0.01 and 0.02, in allelic and genotypic tests, respectively. These findings suggest that genes rapidly evolving along the human lineage may contribute to preterm birth risk.

## 130

**High-Throughput Discovery and Characterization of Fetal Protein Trafficking in the Blood of Pregnant Women.** *J.L. Maron[1], G. Alterovitz[2], M. Ramoni[2], K.L. Johnson[3], D.W. Bianchi[1, 3].* 1) Dept Pediatrics, Division of Newborn Medicine, Floating Hospital for Children at Tufts Medical Center, Boston, MA; 2) 2. Children's Hospital Informatics Program, Harvard-MIT Division of Health Sciences and Technology, Harvard Medical School, Boston, MA; 3) Dept Pediatrics, Division of Genetics, Floating Hospital for Childrne at Tufts Medicla Center, Boston, MA.

**Background:** Measurement of fetal proteins in maternal serum is part of standard prenatal screening. However, attempts to better understand the extent of feto-maternal protein trafficking and its clinical and biological significance have been hindered by the presence of abundant maternal proteins that limit the detection of fetal proteins in maternal serum samples. **Objective:** The objective of this study was to circumvent maternal protein interference by developing a computational predictive approach for a noninvasive and comprehensive, protein network analysis of the developing fetus using maternal whole blood. **Methods:** Previously identified fetal gene transcripts in maternal whole blood (n=157) (Maron et al. J Clin Invest 2007;117:3001-19) were translated into their corresponding proteins. Using a program called MASSOME, which automatically integrates a number of computational databases, we generated a comprehensive integrated predictive protein network of the developing fetus. To confirm the predicted computational model, we performed Western blot analyses of three predicted proteins on prospectively obtained maternal and umbilical cord whole blood samples. **Results:** Forty-six of the initial 157 fetal gene transcripts were classified into known protein networks, and 222 downstream proteins were predicted. Statistically significantly over-represented pathways were diverse and included T-cell biology, glucose regulation, neurodevelopment, and cancer biology. Western blot analyses were performed for three predicted proteins: VAV1, a proto-oncogene actively involved in hematopoiesis and T and B cell development; Plexin A1, involved in axon guidance, cardiac morphogenesis and the immune system; and Adenosine Deaminase (ADA), which when deficient results in a form of severe combined immunodeficiency disease. The computational predictive model was validated and confirmed by the presence of these specific downstream fetal proteins in the whole blood of pregnant women and their newborns, with absence (VAV1, Plexin A1) or reduced detection (ADA) of the protein in the maternal postpartum samples. **Conclusions:** This work demonstrates that extensive feto-maternal protein trafficking occurs during pregnancy, and can be predicted and subsequently verified to develop novel noninvasive biomarkers. This study raises important questions regarding the biological effects of fetal proteins on the pregnant woman.

Supported by NICHD R01 HD042053-06.

## 131

**NARP and MELAS mutations differentially impact mitochondrial DNA segregation throughout human embryofetal development.** *S. Monnot[1], N. Gigarel[2], N. Frydman[3], P. Burlet[2], S. Gobin[2], M. Sinico[4], M. Bonniere[2], M. Rio[1, 2], A. Rotig[1], R. Frydman[3], A. Benachi[5], A. Munnich[1, 2], J-P. Bonnefont[1, 2], J. Steffann[1, 2].* 1) INSERM Unit U781, Necker hospital, Paris, France; 2) Genetics Department, Necker hospital, Paris, France; 3) Reproductive Medecine Department, Beclere Hospital, Clamart, France; 4) Pathology Laboratory, CHIC, Creteil, France; 5) Obstetrics Department, Necker hospital, Paris, France.

Mitochondrial DNA (mtDNA) mutations cause a wide range of serious genetic diseases with maternal inheritance and a high transmission risk. Little is known about segregation of mutant mtDNA molecules during human *in utero* development. Taking advantage of our pregestational and prenatal programs, we measured the m.3243A>G (MELAS) and m.8993T>G (NARP) mutant loads in various types of single cells or tissues from 33 day3 (D3) - day5 (D5) embryos and 25 10-30 gestation-week (GW) fetuses, from 16 unrelated carrier females, using a semi-quantitative fluorescent PCR-restriction test.

Only 80% of embryos/fetuses inherited the maternal mutation. Maternal mutant loads, assessed in various somatic tissues, were not predictive for the embryofetal ones.

At D3-D5 embryonic stage, mutant loads *i)* did not vary among blastomeres of a given embryo, irrespective of the mutation; *ii)* apparently depended on the mutation type. NARP embryos (n=3) were wild-type (<2%) or mutant (>95 %) homoplasmic, in agreement with the bottleneck theory. MELAS embryos were mutation-free (<2%) or heteroplasmic (embryonic mutant loads ranging from 5% to 80%), suggestive of a random mtDNA segregation with a negative selection process against premature germ cells carrying a high mutant load during maternal oogenesis.

Over 8 GW, mutant loads *i)* had a bimodal distribution (<30 % or > 65 %) among NARP fetuses (n=13), suggestive of some selection event towards normal or mutant mtDNA during the first 8 GW, while it was homogeneously distributed among MELAS fetuses (0-80 %, n=12); *ii)* did not vary with time (8-30 GW); *iii)* were uniformly distributed across various tissues, contrasting with a marked variation at the single cell level, irrespective of the mutation type, thus supporting a random segregation of mutant mtDNA molecules during embryofetogenesis.

This inter-tissue stability of both MELAS and NARP heteroplasmy rates during embryofetal development (< 30GW) contrasted with a tissue-dependent variability specific for the MELAS mutation in the postnatal period. This suggests that some mutation-specific event interferes with the random mtDNA segregation after 30 GW or in the postnatal period.

## 132

**Adverse Effects of Trichothiodystrophy DNA Repair and Transcription Gene Abnormalities on Human Fetal Development.** *R. Moslehi[1,2], C. Signore[3], D. Tamura[2], J.L. Mills[3], J.J. DiGiovanna[2], M.A. Tucker[2], J. Troendle[3], A.M. Goldstein[2], K.H. Kraemer[2].* 1) Epidemiology and Biostatistics, School of Public Health, and Center for Excellence in Cancer Genomics, State University of New York (SUNY) at Albany, Rensselaer, NY; 2) National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, MD; 3) National Institute of Child Health and Human Development (NICHD), NIH, Bethesda, MD.

**Background** Effects of DNA repair and transcription genes in human prenatal life have never been studied. Trichothiodystrophy (TTD) is a rare (affected frequency of $10^{-6}$ in Europe) recessive disorder caused by mutations in genes involved in the nucleotide excision repair (NER) pathway and in transcription. Mutations in *XPD* (*ERCC2*), *XPB* (*ERCC3*), *TTDA* (*GFT2H5*) or *TTDN1* (*C7ORF11*) can cause TTD. Besides TTD, mutations in *XPD* can also lead to combined phenotypes of xeroderma pigmentosum (XP)/TTD and cerebro-oculo-facial syndrome (COFS)/TTD. **Methods** Based on clinical observations, we conducted a genetic epidemiologic study to investigate gestational outcomes associated with TTD. We compared pregnancies resulting in TTD, COFS/TTD and XP/TTD-affected offspring (N=24) with respect to abnormalities during their antenatal and neonatal periods to pregnancies resulting in their unaffected siblings (N=18), accounting for correlation, and to population reference values. **Results** Significantly higher incidence of several severe gestational complications was noted in all TTD and COFS/TTD-affected pregnancies, where the fetus had two mutations in a TTD gene. Small for gestational age (SGA<10th percentile) (RR=9.3, 95%CI:1.4-60.5,P=0.02), SGA<3rd percentile (RR=7.2, 95%CI:1.1-48.1,P=0.04), and neonatal intensive care unit (NICU) hospitalization (RR=6.4, 95%CI:1.4-29.5,P=0.02) occurred more frequently among TTD-affected neonates compared to their unaffected siblings. Compared to reference values, pregnancies that resulted in TTD-affected infants were significantly more likely to be complicated by hemolysis, elevated liver enzymes, and low platelets (HELLP) syndrome (RR=35.7, 95%CI:7.6-92.5,P=0.0002), elevated mid-trimester maternal serum human chorionic gonadotropin (hCG) levels (RR=14.3, 95%CI:7.0-16.6,P<0.0001), SGA<3rd percentile (RR=13.9, 95%CI:7.4-21.1,P<0.0001), preterm delivery (<32 weeks) (RR=12.0, 95%CI:4.9-21.6,P<0.0001), pre-eclampsia (RR=4.0, 95%CI:1.6-7.4,P=0.006), and decreased fetal movement (RR=3.3,95%CI:1.6-5.2,P=0.0018). Abnormal placental development may explain the constellation of complications in our study. **Conclusion** We hypothesize that TTD DNA repair and transcription genes play an important role in normal human placental and fetal development. Pregnancies resulting in XP/TTD, which is also caused by mutations in *XPD*, were not associated with gestational complications, indicating the relevance of the exact gene abnormality to the postulated mechanism.

## 133

**A Congenic Approach To The Identification Of Novel Sex Determining Genes.** *E. Vilain, V. Arboleda.* Dept Human Gen, Univ California Sch Med, Los Angeles, CA.

Disorders of sex development encompass a very large spectrum of phenotypes, from minor malformations of the genitalia (hypospadias, hypertrophy of the clitoris) to sexual ambiguity. Taken altogether, these anomalies have an estimated frequency of 0.5% to 1%. Moreover, sexual ambiguity has a major impact on quality of life. Recently, the debate about the management of intersex patients has intensified over issues of gender assignment and the indication for early genital surgery. Yet the scientific data on patient outcome have remained poor. The main obstacles to the optimal management of intersex patients have been a combination of lack of controlled outcome data and the lack of understanding of their pathophysiology, which prevents precise diagnostic categorization of patients. Despite much progress in the past 15 years, the molecular mechanisms underlying mammalian sex determination, are still far from understood, and the molecular basis of sex reversal in the majority of XY patientsabout 75%cannot yet be explained. Mouse models are invaluable to the understanding of human sex determination. C57BL/6J (B6) mice containing the Mus domesticus poschiavinus Y chromosome, YPOS, develop ovarian tissue, whereas testicular tissue develops in DBA/2J or 129S1/SvImJ (129) mice containing the YPOS chromosome. To identify genes involved in sex determination, we used a congenic strain approach to determine which chromosomal regions from 129Sl/SvImJ provide protection against sex reversal in XYPOS mice of the B6.129-YPOS strain. Genome scans using microsatellite and SNP markers identified a chromosome 11 region of 129 origin in B6.129-YPOS mice. To determine if this region influenced testis development in XYPOS mice, two strains of C57BL/6J-YPOS mice were produced and used in genetic experiments. XYPOS adults homozygous for the 129 region had a lower incidence of sex reversal than XYPOS adults homozygous for the B6 region. In addition, many homozygous 129 XYPOS fetuses developed normal-appearing testes, an occurrence never observed in XYPOS mice of the B6-YPOS strain. We conclude that a chromosome 11 locus derived from 129Sl/SvImJ protects against sex reversal in XYPOS mice. Further backcrossing of B6.129-YPOS with B6 mice combined with bioinformatics approaches based on sequence conservation allowed to considerably refine the critical region of chromosome 11 protecting against XY sex reversal and to identify novel candidate sex-determining genes.

## 134

***SPAG16* cDNA mutations associated with oligo- and oligoasthenozoospermia in infertile men.** *A.N. Yatsenko[1,2], P. Arias[1], R. Chen[1], L.J. Murthy[3], D.J. Lamb[3,4], M.M. Matzuk[1,2,4].* 1) Pathology; 2) Molecular Human Genetics; 3) Urology; 4) Molecular and Cell Biology, Baylor College of Medicine, Houston, TX.

Male infertility is a major health problem of multifactorial etiology. It is estimated that genetic factors account for 50% or more cases of male infertility. Oligozoospermia is a common semen abnormality condition associated with male infertility. However, genetic defects that cause oligozoospermia are unknown in the majority of patients. In present study, we investigated the role of *SPAG16* in severe oligoasthenozoospermia in humans. Previous reports show that chimeric mice with the disrupted protein, Spag16, were infertile due to reduced sperm count, motility defects, and the inability to fertilize oocytes. Interestingly, there are two transcript variants encoding different isoforms of the *Spag16* gene. SPAG16L and SPAG16S proteins have distinct spermatozoal cell locations and functions. The SPAG16L isoform is predominantly expressed in axonemal bridges, connecting two central microtubules of the structure “9+2”; therefore, it is associated with for sperm motility defects. The SPAG16S protein lacks N-terminal portion, contains contiguous C-terminal WD repeats, and is expressed primarily in spermatozoal nuclei. The protein interacts with chromosome/chromatin-binding protein MEIG1. Thus, the SPAG16S may regulate spermatozoal chromosomal structure and/or gene expression. We report a potential association between *SPAG16* mutations and isolated oligo- and oligoasthenozoospermia. Among 250 spermatozoal cDNA samples from oligozoospermic and oligoasthenozoospermic men, we identified 27 novel heterozygous missense and 2 splicing alterations in 29 patients (12%). None of these nucleotide changes are listed as known polymorphisms in the dbSNP database. The cDNA defects observed will be confirmed in the genomic DNA. Our preliminary results suggest that mutations in *SPAG16* may contribute to a large fraction of oligoasthenozoospermia in infertile men. This study was supported in part by the National Institutes of Health Infertility Center (P01HD36289) to MMM and DJL.

## †135

**Cross-over levels are established at the onset of prophase in mammalian males.** *B.S. Baier, P.A. Hunt, T.J. Hassold.* School of Molecular Bioscience, Washington State University, Pullman, WA.

Segregation of chromosomes during the first meiotic division relies heavily on crossovers (COs) established during prophase. Although COs are strictly regulated so that at least one occurs per chromosome arm and multiple COs on the same chromosome display interference, variation in CO levels is not uncommon. We have identified highly significant among-strain variation in overall rates of recombination in male mice. For example, three strains, CAST/EiJ, C3H/HeJ, and C57BL/6J, display "low", "medium", and "high" genome-wide CO rates, respectively, as measured by the number of MLH1 foci per cell. Since errors in segregation are associated with altered recombination, we have been interested in defining factors responsible for this variation; e.g., the timepoint at which the number of COs is established and the loci responsible for the variation. Our approach utilizes immunofluorescence to analyze the number and distribution of proteins that function at different stages in the meiotic recombination pathway: RAD51 and DMC1, strand invasion proteins acting shortly after double-strand break (DSB) formation, MSH4, part of the complex stabilizing double Holliday junctions, BLM, responsible for non-crossovers and, possibly, non-interfering COs, and MLH1, involved in CO resolution. We detected consistent strain-specific differences in the average number of foci at each stage of the recombination pathway. Specifically, the proportions of DSB sites (identified by RAD51) that acquired DMC1, MSH4, BLM, and MLH1foci remained constant among the strains, with C57BL/6J mice having the highest and CAST/EiJ mice the lowest average values for each protein. This indicates that differences in the number of DSBs are translated into differences in the number of COs, suggesting that the level of COs is established at, or before, DSB formation. Chromatin conformation is clearly important in this process, since analysis of synaptonemal complex (SC) length and DNA loop size also identified consistent strain specific differences; i.e., strains with higher MLH1 levels displayed longer SCs and, consequently, shorter DNA loops emanating from the chromosome axis. This indicates a relationship between recombination and chromatin compaction in which CO frequency increases with SC length and is inversely related to chromatin loop size, an association that may occur as DSBs form or earlier during the establishment of the chromosome axis.

## 136

**A French cohort of 15 162 children aged zero to four years and conceived by Assisted Reproductive Technology (ART).** *G. Viot[1,2], S. Epelboin[2], S. Follow up Association[3].* 1) Dept Genetics, Hosp Necker, Paris, France; 2) Med de la Reproduction, Hosp Cochin-Saint Vincent de Paul, Paris France; 3) Poissy, France.

AIMS : To estimate the risk of congenital malformations, particularities, mental retardation, obstetrical or prematurity complications of children conceived by in vitro fertilization or intracytoplasmic sperm injection. DESIGN : Prospective study from 2003 to 2007. SETTING : Thirty three centers all over the French country, registred for ART. PATIENTS : Fifteen thousand one hundred sixty-two infants conceived by IVF or ICSI. Children were recruited from birth and medical data were collected at birth and then at the age of 4, 9, 12, 24, 36 and 48 months. Questionnaires were filled in 1/3 of cases by the paediatrician and in 2/3 of them by the parents. We could conclude for all of them except 33. RESULTS : 473 (3.12%) and 286 children (1.89%) had experienced a major or minor congenital malformation respectively (5.01% vs 3% estimated in the general population). This higher rate observed in ART group was partly due to an excess of malformations in the urogenital system (uropathy : 0.72%, hypospadias : 0.37%, cryptorchidism : 0.39%), of cardiac malformations (0.95% vs 0.6%) and of hip dysplasia (1% vs 0,3%). Over the particularities observed in this cohort, a significant frequency of angioma was found, affecting 365 children (2.4% vs 0.5%). To our knowledge, this association had never been previously reported in ART-born children. Genetic disorders were found in 110 children. Over them, 6 children presented with Beckwith-Wiedemann syndrome (0.04% vs 0.007%) and 5 an unilateral retinoblastoma (0.03% vs 0.006%), both diseases influenced by epigenetics. No child was affected by Angelman syndrome in this cohort. Mental retardation was found in 78 children >37 weeks of amenorrhea (WA) (0.51%) and in 27 children < 37 WA (0,18%). Obstetrical and prematurity complications were found in 178 and 371 children respectively. CONCLUSION : In this large multicentric survey of ART-born children, 99.8% of the questionnaires were exploitable. We observed an excess of malformations in the urogenital and cardiac systems. Genetic diseases influenced by epigenetics appeared more frequent. Unexpectedly, the rate of angioma was significantly higher than in the general population. Long-term follow-up is planned to obtain more information concerning psychomotor and intellectual development of these children. New questionnaires will be next conducted on 5-years-old children.

† Trainee Award Finalist

## 137

**LBH589 induces up to 10-fold SMN protein levels by several independent mechanisms and is effective even in cells from SMA patients non-responsive to valproate.** *B. Wirth[1], M. Riessland[1], I. Hölker[1], R. Heller[1], J. Hauke[1], C. Tränkle[2], R. Coras[3], I. Blümcke[3], E. Hahnen[1], L. Garbes[1].* 1) Institute of Human Genetics, Institute of Genetics and Center for Molecular Medicine Cologne, Cologne, Germany; 2) Department of Pharmacology and Toxicology, Institute of Pharmacy, University of Bonn, Bonn, Germany; 3) Institute of Neuropathology, University of Erlangen, Erlangen, Germany.

Histone deacetylase inhibitors (HDACi) are potential candidates for therapeutic approaches in cancer and neurodegenerative diseases such as spinal muscular atrophy (SMA) - one of the most common autosomal recessive disorders and frequent cause of early childhood death. SMA is caused by homozygous absence of *SMN1*. Importantly, all SMA patients carry a nearly identical copy gene, *SMN2*, that, however, produces only minor levels of correctly spliced full-length transcripts and SMN protein. Since an increased number of *SMN2* copies strongly correlates with a milder SMA phenotype, activation or stabilization of SMN2 is considered as a therapeutic strategy. However, since clinical trials demonstrated effectiveness of the HDACi valproate (VPA) and phenylbutyrate only in <50% of patients, the identification of new drugs is of vital importance. Here we characterize a novel FDA-approved hydroxyamic acid LBH589 (Panobinostat), which is a HDACi already widely used in cancer clinical trials. LBH589 treatment of human SMA fibroblasts induced up to 10-fold elevated SMN levels, the highest ever reported, accompanied by a markedly increased number of gems. FL-*SMN2* levels were increased 2- to 3-fold by transcription activation via *SMN2* promoter H3K9-hyperacetylation and restoration of correct splicing via elevated hTRA2-β1 levels. Furthermore, LBH589 stabilizes SMN by reducing its ubiquitinylation as well as favouring incorporation into the SMN complex. Importantly, LBH589 markedly elevated SMN levels also in neurvous tissues such as cultured human neural stem cells. Furthermore, murine embryonic fibroblasts derived from transgenic mice carrying the human *SMN2* on a null *Smn* backround, presented a significant SMN elevation. Subcutaneous injection of LBH589 in these animals led to increased *SMN2* expression in neuronal tissues suggesting that LBH589 is able to pass the blood-brain-barrier. In general, cytotoxic effects were not detectable at *SMN2* activating concentrations. Finally, LBH589 induced SMN expression even in SMA fibroblasts inert to VPA. Hence, LBH589, which is active already at nanomolar doses in vitro, and induces SMN elevation by at least three different mechanisms is a highly promising candidate for SMA therapy.

## 138

**Hypophosphatasia: Treatment of Life-Threatening Disease Using Bone-Targeted Human Recombinant Tissue Non-Specific Alkaline Phosphatase.** *M.P. Whyte[1,2], C.R. Greenberg[3], T. Edgar[4], B.J. Van Sickle[5], M. Hamdan[6], N. Salman[6], M. Bober[7], W.H. McAlister[8], D. Wenkert[1], H. Landy[9].* 1) Research Center, Shriners Hospital, St Louis, MO; 2) Div of Bone and Mineral Diseases, Washington University School of Medicine, St Louis, MO; 3) Dept Human Genetics, University of Manitoba, Winnipeg, Manitoba, Canada; 4) Prevea Allouez Health Center, Green Bay, WI; 5) Pediatric Endocrinology Div, Vanderbilt University, Nashville, TN; 6) Neonatology, Tawam Hospital, Al Ain, United Arab Emirates; 7) Div Genetics-AIDHC, Nemours, Wilmington, DE; 8) Mallinckrodt Inst Radiology, Washington University School of Medicine, St Louis, MO; 9) Medical Affairs, Enobia Pharma, Montreal, Canada.

Hypophosphatasia (HPP) features low serum alkaline phosphatase (ALP) activity due to deactivating mutation(s) within the gene that encodes the "tissue nonspecific" isoenzyme of ALP (TNSALP). Consequently, inorganic pyrophosphate, a natural substrate for this ectoenzyme, accumulates extracellularly and blocks skeletal mineralization. HPP severity spans stillbirth from profound skeletal hypomineralization to osteomalacia presenting in adult life. There is no established medical treatment. ENB-0040 is a bone-targeted, human recombinant, TNSALP fusion protein that preserved skeletal mineralization and survival and prevented vitamin B6-responsive seizures and dental defects in a TNSALP knockout mouse model of infantile HPP (JBMR 23:777, '08). Patient trials began in the summer '08. In a phase 1, month-long, multi-center, open-label protocol, 6 adults received 1 IV infusion of 3 mg/kg ENB-0040 followed by weekly SC injections of 1 or 2 mg/kg (Endo Soc, abstract, in press). Here, we report findings from our 6-mo, open-label protocol involving 5 patients with life-threatening HPP (ages 0.5-36 mo at baseline), with up to 6 mo of ENB-0040 treatment. Previously, each had shown worsening skeletal disease and respiratory symptoms predicting a lethal outcome. At age 7 mo, pt 1 with infantile HPP received a single IV infusion of 2 mg/kg of ENB-0040 followed by 1 and then 3 mg/kg SC 3X/wk. During therapy, there was substantial remineralization of the skeleton, weaning from ventilatory support, and improved growth and motor development. After receiving only 3 wk of treatment, pt 2, also with infantile HPP, showed skeletal remineralization, and then improved ventilation. Both patients 1 & 2 have been released from hospital to continue ENB-0040. Patient 3 with perinatal HPP has also shown substantial radiographic and respiratory improvement, and corrected hypercalcemia soon after treatment began. There have been no drug-related serious adverse events, or development of anti-ENB-0040 antibodies. SC dosing has resulted in ENB-0040 activity in the therapeutic target range. Substantial skeletal remineralization and improved clinical status has been demonstrated in association with ENB-0040 bone-targeted enzyme replacement therapy in 3 severely affected infants with HPP in this short term study.

## 139

**CNS transplantation of purified human neural stem cells in Infantile and Late-Infantile Neuronal Ceroid Lipofuscinoses: Summary of the Phase I trial.** *R. Steiner[1], S. Huhn[2], T. Koch[1], A. Al-Uzri[1], D. Guillaume[1], T. Sutcliffe[2], H. Vogel[2], N. Selden[1].* 1) Oregon Health & Science University, Portland, OR; 2) Stem Cells, Inc., Palo Alto, CA.

Infantile and late-infantile neuronal ceroid lipofuscinosis (INCL and LINCL) are universally fatal neurodegenerative lysosomal storage disorders caused by palmitoyl protein thioesterase and tripeptidyl peptidase deficiency, respectively. An open-label dose-escalation Phase I trial of human central nervous system stem cells (HuCNS-SC®) was conducted in subjects with INCL and LINCL. This study represents the first US FDA authorized use of human neural stem cells for clinical testing. The primary goal of the trial was to evaluate the safety of HuCNS-SC, the immunosuppression regimen, and the surgical technique. The secondary goal was to evaluate preliminary efficacy. Six children, four boys and two girls ranging in age from 2 years 11 months to 9 years 4 months, underwent HuCNS-SC transplantation; two children had INCL and four had LINCL. The children were in the moderate to severe stage of disease. HuCNS-SC transplantation involved a single-stage procedure with delivery of the cells to multiple sites in both cerebral hemispheres and lateral ventricles, thus testing the safety of both parenchymal and ventricular administration. Immunosuppression with a calcineurin inhibitor was administered for twelve months after transplantation. An independent Data Monitoring Committee reviewed the conduct of the trial, adverse events, and outcome measures. Five subjects are surviving. One subject expired from disease progression eleven months after transplantation. The results of a brain-only autopsy in this subject revealed no signs of HuCNS-SC transplantation toxicity. The remaining five surviving subjects completed the assessments. The study results to be presented will include an analysis of adverse events (AEs), as well as the neurological, seizure frequency, EEG, radiological (CT and MRI), neuropsychological, functional, and quality-of-life measures that were assessed at frequent intervals post-transplant. The results of this Phase I study represent the first human safety data for CNS transplantation of purified and expanded neural stem cells. The dosing and neurosurgical approach represent a novel transplantation paradigm that differs significantly from historical CNS allograft investigations. This initial trial sets the stage for future clinical trials of HuCNS-SC in these and other neurodegenerative diseases.

## 140

**RNAi Therapy for FSHD.** *S. Garwick[1], L. Wallace[2], S.Q. Harper[1,2,3].* 1) Center for Gene Therapy, The Research Institute at Nationwide Children's Hospital, Columbus, OH; 2) MCDB Graduate Program, The Ohio State University, Columbus, OH; 3) Department of Pediatrics, The Ohio State University, Columbus, OH.

Over the last 15 years, muscular dystrophy gene therapy strategies have been primarily aimed at replacing defective or missing genes underlying recessive disorders, such as Duchenne muscular dystrophy (DMD). However, these gene replacement strategies are not feasible for treating dominant diseases. Instead, patients bearing dominant mutations would likely benefit from reduction or elimination of the abnormal allele. Until very recently, there was no feasible mechanism to reduce or eliminate disease genes, and molecular therapy development for dominant muscular dystrophies was largely unexplored. RNA interference (RNAi) has recently emerged as a powerful tool to suppress any gene of interest in a sequence specific manner. As such, RNAi is a leading candidate strategy to silence dominant disease genes, including those involved in muscular dystrophies such as facioscapulohumeral muscular dystrophy (FSHD). FSHD is a dominant genetic disorder caused by contraction of D4Z4 repeats on chromosome 4q35. Current evidence supports that FSHD is an epigenetic disorder. FSHD-associated deletions produce chromatin changes that may aberrantly up-regulate multiple myotoxic genes. Currently, two FSHD-candidate genes, FRG1 and DUX4, have been shown to cause dystrophy-related phenotypes upon over-expression *in vivo*, suggesting their potential involvement in FSHD pathogenesis. Thus, RNAi therapies targeting FRG1 or DUX4 may offer the first targeted treatment for FSHD. Currently, only mouse models over-expressing FRG1 exist. To establish proof-of-principle for FSHD targeted RNAi therapy, we developed therapeutic microRNAs to reduce toxic FRG1 expression in mice. These microRNAs (miFRG1) caused significant knock-down of FRG1 mRNA and protein *in vitro*. When delivered via AAV6 viral vectors to FRG1-high mouse muscle, miFRG1, but not control miGFP microRNAs, significantly reduced toxic FRG1 levels, and improved muscle mass and histopathology. Specifically, miFRG1-transduced muscles were normal in size, showed no fibrosis or fat deposition in muscle, and had no evidence of myofiber degeneration or regeneration. Our results provide the first proof-of-principle for RNAi therapy targeting an FSHD candidate gene, and support its feasibility for other dominant muscular dystrophies in general.

## 141

**A Phase 1 trial of recombinant human acid sphingomyelinase (rhASM) enzyme replacement therapy in adults with ASM deficiency (ASMD).** *M.M. McGovern[1,2], M.P. Wasserstein[1], B. Kirmse[1], L. Duvall[1], T. Schiano[1], B.L. Thurberg[3], S. Richards[3], G.F. Cox[3].* 1) Mount Sinai School of Medicine, NY, NY; 2) Stony Brook University School of Medicine, Stony Brook, NY; 3) Genzyme Corporation, Cambridge, MA.

Purpose: Preclinical studies of rhASM in ASM knockout mice have demonstrated efficacy with dose related reductions in sphingomyelin. At high starting doses (≥10 mg/kg) in ASM knockout mice, but not normal animals, cytokine elevations and shock were seen, which could be prevented by dose escalation. This Phase 1 trial evaluated the safety and PK of single doses of rhASM in nonneuronopathic ASMD (Niemann-Pick B). Methods: Single doses of 0.03, 0.1, 0.3, 0.6 and 1.0 mg/kg rhASM were infused sequentially by dose cohort. Dose cohorts 1 and 2 had 3 patients each, cohorts 3 and 4 had 2 patients each, and cohort 5 had 1 patient. Main entry criteria were age 18-65 yr, deficient ASM enzyme activity, spleen volume ≥2X normal, AST and ALT ≤250 IU/L, bilirubin ≤3.6 mg/dL, INR ≤1.5, DLco >30% predicted, platelets ≥60,000/μL, and no cirrhosis by liver biopsy or significant cardiac disease. Patients were monitored for 72 hr post-dose and returned for visits on Days 14 and 28. Results: Preliminary safety data are available from this 11-patient trial (mean age 30.1 yr, 6 males, 11 Caucasian, mean spleen volume 10.8X normal). No clinically significant changes were observed by cardiovascular monitoring or cytokine levels. Patients showed dose related rises in ceramide, bilirubin, C-reactive protein and other acute phase reactants that peaked at 24-48 hr post-dose and resolved by Days 3-14. Three patients in cohorts 3-5 experienced 11 mild to moderate clinical symptoms assessed as drug-related including fever (n= 2), pain (headache; myalgia; abdominal, leg, hip and back pain), fatigue, nausea (n=2), and vomiting. Onset of symptoms began 12 hr post-dose and resolved by 72 hr, except for hip pain that began after 72 hr. Although the number of patients was small clinical symptoms appeared to increase in severity with dose. Day 14 liver biopsies were unchanged from baseline except for 1 patient (cohort 4) with 2 new small inflammatory foci. Conclusion: This trial describes the first experience with enzyme replacement therapy in adult patients with ASMD. At bioactive doses in humans, rhASM did not cause significant cytokine elevations or cardiovascular changes. The major safety observations were dose-related hyperbilirubinemia and acute phase response. Several safety biomarkers were identified. Based on the hyperbilirubinemia findings the maximum tolerated starting dose of rhASM was 0.6 mg/kg. Within-patient dose escalation may be an option for higher repeat doses of rhASM.

## 142

**Phenylbutyrate Therapy for Maple Syrup Urine Disease.** *N. Brunetti-Pierri[1], B. Lanpher[1], A. Erez[1], R.M. Wynn[2], D.T. Chuang[2], C. Yu[3], H. Madhuri[3], M. Islam[4], S. Hutson[4], B. Lee[1,5].* 1) Dept Molec & Human Genetics, Baylor Col Medicine, Houston, TX; 2) Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX; 3) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 4) Human Nutrition, Foods and Exercise Virginia Tech, Blacksburg, VA; 5) Howard Hughes Institute, Baylor College of Medicine, Houston, TX.

Maple syrup urine disease (MSUD) is an inborn error of amino acid metabolism caused by deficiency of the mitochondrial branched-chain ketoacid dehydrogenase (BCKD) resulting in the accumulation of neurotoxic branched-chain amino acids (BCAA) and their corresponding branched-chain α-ketoacids (BCKA). Therapy with phenylbutyrate results in a selective reduction in BCAA despite adequate dietary protein intake in patients with urea cycle disorders. Based on this observation, we have investigated the effect of phenylbutyrate on plasma BCAA and their corresponding BCKA and we have found that they are both significantly reduced in control subjects and in patients with a late-onset form of MSUD. In both control and patient cells and in wild type mice, we find that the reduction in BCAA and their corresponding BCKA is due to an increase of the BCKD enzyme activity secondary to an increase of the active unphosphorylated form over the inactive phosphorylated form promoted by the phenylbutyrate. Phenylbutyrate enhanced unphosphorylated E1 (fully active enzyme) catalyzed decarboxylation of all three BCKA significantly. It not only increased the kcat but also increased the sensitivity of the enzyme to BCKA by lowering their Km values. Phenylbutyrate did not have an effect on phosphorylated E1 (inactive enzyme), but did prevent completely phosphorylation and inactivation of E1 in the presence of BCKD complex kinase. Phenylbutyrate enhanced overall BCKD complex activity (E1 + E2 + E3 activities) by the increases of 50 to 70% in kcat values for BCKA. Therefore, phenylbutyrate appears to activate decarboxylase activity and increase the activity state of BCKD complex by blocking kinase catalyzed inactivation of E1. In summary, we show that phenylbutyrate causes a reduction of neurotoxic BCAA and therefore, it may represent a valuable treatment for forms of MSUD with some levels of residual enzymatic activity.

## 143

**Brain-directed AAV5 gene therapy, in combination with copper, rescues a murine model of severe Menkes disease.** *S. Kaler[1], A. Donsante[1], J. Tang[1], D. Goldstein[2], C. Holmes[2], P. Sullivan[2], J. Centeno[3].* 1) Unit on Pediatric Genetics, Molecular Medicine Program, NICHD/NIH, Bethesda, MD; 2) Clinical Neurocardiology Section, NINDS/NIH, Bethesda, MD; 3) Division of Biophysical Toxicology, Department of Environmental and Infectious Disease Sciences, United States Armed Forces Institute of Pathology, Washington, DC.

Menkes disease (MD) is caused by defects in an X-linked copper transporter, ATP7A. Affected males exhibit low copper levels in blood and brain, abnormal neurotransmitter levels (due to decreased activity of dopamine-beta-hydroxylase, a copper enzyme), infantile neurodegeneration, dysmyelination, seizures, and death by age 3y. Treatment with subcutaneous injections of copper seems most effective in patients who are identified at birth and whose mutations allow some residual ATP7A activity *N Engl J Med* 2008 358:605-614, *Ann Neurol* 2009 65:108-113). For complete loss-of-function mutations, however, outcomes are suboptimal since copper injected peripherally does not traverse the blood-brain barrier efficiently without ATP7A. Thus, alternative treatment strategies are needed for MD patients with severe molecular defects. The mottled-brindled (*mo-br*) mouse on a C57BL/6J background provides a facsimile of severe MD in which to test novel therapies. This model is unresponsive to peripheral administration of copper (either ip or iv) and dies by 13d. As in MD, *mo-br* mice have reduced brain copper and abnormal brain neurotransmitter levels. We evaluated two brain-directed treatment approaches in this murine model: intracerebroventricular (ICV) administration of 50 ng of copper, and ICV delivery of an AAV5 vector harboring a reduced-size, functional version of human *ATP7A*. Both individual therapies extended lifespan slightly (to 14.3 and 15.6 days, respectively, $p<0.02$) and improved neurotransmitter levels. In conjunction with differences in brain copper levels, these results suggest the two treatments have different effects: ICV copper injections increasing copper presentation to neurons, and ICV gene therapy improving copper utilization. We hypothesized that combining the two treatments would yield a synergistic effect. Indeed, combination therapy dramatically shifts the Kaplan-Meier survival curve, with 5 of 16 (31%) combination-treated mutants surviving beyond 110 days of age ($p<0.0002$). Combination-treated *mo-br* mice weighed more than untreated *mo-br* mice (5.2 g vs. 4.2 g at 9 days of age, $p<0.02$) and less than wild type controls (11.9 g vs. 15.1 grams at 27 days of age, $p<0.01$). The combination-treated survivors appear healthy, active, and fertile. Preliminary data indicate normal EEG and rotorod performance. Our findings suggest that brain-directed therapies may be relevant for MD patients with severe ATP7A mutations unresponsive to conventional treatment.

## 144

**HDAC overexpression or HAT inhibition leads to CGG-specific repression of transcription and phenotypic rescue in a Drosophila model of Fragile X Tremor Ataxia Syndrome.** *P.K. Todd[1], J.P. Taylor[2], H.L. Paulson[1].* 1) Neurology, University of Michigan, Ann Arbor, MI; 2) Developmental Neurobiology, St. Jude Children's Research Hospital, Memphis, TN.

Fragile X Tremor Ataxia Syndrome (FXTAS) is a recently described inherited form of ataxia caused by a "pre-mutation" length (55-200 repeat) CGG expansion in the 5'UTR of the fragile X syndrome gene, FMR-1. FXTAS is thought to be an RNA-mediated neurodegenerative disorder and mRNA levels for FMR1 are markedly increased in FXTAS patients. Histone Deacetylases (HDACs) are known to play a role in transcriptional regulation at the FMR1 locus. We present evidence that overexpression of any of three different HDAC proteins from different classes, HDACs 3, 6, or 11, can rescue the neurodegenerative phenotype in a drosophila model of CGG repeat induced neurodegeneration. Phenotypic rescue is associated with transcriptional repression of the transgene, as measured by multiplex real time RT-PCR. Importantly, this transcriptional repression is not seen when the CGG repeat sequence is absent. Experiments in lymphoblasts derived from FXTAS patients demonstrate that FMR1 mRNA expression is elevated compared to controls. Treatment of FXTAS lymphoblasts with the Histone Acetyltransferase (HAT) inhibitors anacardic acid or garcinol results in a dose-dependent repression of expanded CGG repeat mRNA expression to near control lymphoblast levels. Taken together, these results suggest that modulation of HAT and HDAC activity at the FMR1 locus is a potential therapeutic target in patients with FXTAS. *This work was supported by a grant from the AAN foundation to P.K.T.

## 145

**The First Report of the MPS VI (Mucopolysaccharidosis VI, Maroteaux-Lamy syndrome)Clinical Surveillance Program (CSP).** *C. Hendriksz[1], V. Valayannopoulos[2], E. Teles[3], E. Miebach[4], P. Harmatz[5], G.M. Pastores[6], R. Steiner[7].* 1) Clinical department, Birmingham Children's Hospital, Birmingham, United Kingdom; 2) Centre de Référence des Maladies Métaboliques Hôpital Necker-Enfants Malades,Paris; 3) Hospital Sao Joao Dept Pediatria Alameda, Porto, Portugal; 4) Universitäts-Kinderklinik, Mainz, Germany; 5) Children's Hospital Oakland, Oakland, CA, USA; 6) New York University School of Medicine, New York, USA; 7) Oregon Health Science University, Oregon, Portland, USA.

Objectives: To evaluate clinical outcomes in the MPS VI CSP. Methods: The MPS VI CSP is a multinational, long-term observational program which tracks the clinical outcomes of patients with MPS VI. Clinical data collected from Sep 12, 2005 to Mar 20, 2009 on 101 patients were analyzed. Results: 101 patients were enrolled, including 95 patients on enzyme replacement therapy (ERT) with galsulfase (Naglazyme®). One patient had a successful hematopoietic stem cell transplant (HSCT); one patient had non-grafting HSCT and is now receiving Naglazyme. The median age of the enrolled subjects is 14 years (range 0 - 60). Sixteen patients are under the age of 5 and 34 are aged 18 or older. Of the 95 patients who received ERT, 86 received the recommended Naglazyme dose of 1 mg/kg given once a week, 5 patients received varying doses, and for the remaining 4 patients no dosing information was available. In 75 patients, enrollment into CSP occurred after ERT was started. Of the enrolled patients, 24 received infusions while in a previous clinical study. There have been 6 discontinuations from the CSP to date, 4 deaths, and 2 lost to follow up. A decrease in urinary glycosaminoglycans (uGAG) was seen after Naglazyme treatment in 47 evaluable patients (uGAG results at baseline and at 1 or more follow-up visits). An increase in both height and weight, improved endurance as measured by 12 min walk test and 3 min stair climb test was observed. No changes, indicative of improvement or deterioration, were noted on examination of the cardiac, ophthalmological and audiology data. Antibodies were detected in 34/37 evaluable patients (pre- and post- treatment serum antibody results). The mean time to sero-conversion was 5.6 months. No correlation between antibody level and changes in uGAG levels was observed. Based on medical history and MRI exams, 47 patients had cervical spinal stenosis or symptoms of spinal cord compression. A total of 44 serious adverse events were reported for 30 patients. Eleven patients experienced one or more Naglazyme infusion related reactions. Conclusions: This is the first report of the CSP, which indicates that treatment with Naglazyme appears to provide clinical benefit and is well-tolerated. Cervical spinal stenosis appears to be a frequently observed pathology. Further reports on untreated patients, those given ERT and HSCT will provide information of the natural history of MPS VI and the impact of ERT and/or HSCT on disease course.

## 146

**N-Carbamylglutamate accelerates ureagenesis in patients with hyperammonemia.** *N. Ah Mew[1], E. Daikin[2], I. Payan[2], I. Nissim[2], M. Yudkoff[2], M. Tuchman[1].* 1) Children's National Medical Center, 111 Michigan Ave NW, Washington DC, 20010; 2) Children's Hospital of Philadelphia, 34th St. & Civic Center Blvd. Philadelphia PA 19104.

The urea cycle is responsible for nitrogen regulation and balance in humans. Disruption of normal urea cycle function leads to hyperammonemia, neurological sequelae and can result in death. Current treatment for hyperammonemia is based on dietary protein restriction and ammonia-scavengers, but these interventions commonly fail to avert brain damage. In this study, we investigate the use of N-Carbamylglutamate (NCG), a new therapeutic agent for a subset of hyperammonemic conditions which are caused by a deficiency of N-acetylglutamate (NAG). N-acetylglutamate(NAG) is the allosteric activator for the carbamyl phosphate synthase I (CPSI) reaction. CPSI is the first and rate-limiting enzyme of the urea cycle. A shortage of NAG causes decreased flux through CPSI and is responsible for hyperammonemia in many conditions, including NAG synthase deficiency, propionic acidemia (PA) and methylmalonic acidemia (MMA). Though orally administered NAG is hydrolyzed in the gut, N-carbamylglutamate (NCG) is a stable analog of NAG and activates CPSI in vitro. We demonstrate through stable isotope studies that a 3-day trial of oral NCG improves urea cycle function and decreases plasma ammonia levels in patients with PA or MMA. Thus, NCG is a useful adjunct in the treatment of acute hyperammonemia of various metabolic etiologies that are due to a lack of NAG. We also demonstrate that in two subjects with NAG synthase deficiency, NCG can restore ureagenesis and is therefore curative for patients with this condition.

## 147

**De novo apparently balanced translocations in man are predominantly paternal in origin and associated with a significant increase in paternal age.** *P.A. Jacobs[1], J.K. Morris[2], J.C.P. Baptista[1], B.L. Ng[3], J.A. Crolla[1], N.S. Thomas[1].* 1) Wessex Regional Gen Lab, Salisbury District Hosp, Salisbury, United Kingdom; 2) Centre for Environmental and Preventive Medicine, Wolfson Institute of Preventive Medicine, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, United Kingdom; 3) The Wellcome Trust Sanger Institute, The Genome Research Campus, Hinxton, Cambs, CB10 1SA United Kingdom.

Balanced reciprocal translocations are the commonest class of human structural chromosome abnormality visible at the level of the light microscope. Using only a moderate level of chromosome banding for their detection they are seen in approximately 0.2% of unselected newborns. However, little is known about their mechanism of origin because of the difficulty of studying this in a balanced rearrangement. We report the parental origin of 27 de novo non-recurring balanced reciprocal translocations which were studied, either by flow sorting the derivative products of the translocation or by linkage, using molecular markers situated close to the points of breakage and exchange. We found 26 to be paternal and only 1 maternal in origin. Furthermore the 26 translocations of paternal origin all appeared to be unique events not mediated by sequence homology. In contrast, in the single translocation of maternal origin the points of breakage and exchange both occurred in defensin gene clusters suggesting an origin by non-allelic homologous recombination. The strong paternal bias in our series of translocations lead us to examine paternal age. We found the paternally derived translocations to be associated with a significantly increased paternal age ($p<0.008$), the odds of a de novo balanced reciprocal translocation more than doubling with every 10 year increase in paternal age. The association of increased paternal age with de novo balanced reciprocal translocations suggests that the majority may arise during one of the numerous spermatogonial mitotic cell divisions. In this respect they may be similar to classic mutations caused by single base pair substitutions, a proportion of which are also associated with increased paternal age.

## †148

**Specific binding motifs facilitate meiotic synapsis in mammalian males.** *M.E. Johnson[1], S.R. Shirley[1], G. Aksay[3], S.M. Schaeffer[4], V. Krishnan[5], A. Dhingra[4], C. VandeVoort[2], E.E. Eichler[3], T. Hassold[1].* 1) School of Molecular Biosciences and Center for Reproductive Biology, Washington State University, Pullman, WA; 2) California National Primate Research Center, University of California, Davis, CA 95616; 3) Department of Genome Sciences, University of Washington, Seattle, WA 98915; 4) Department of Horticulture, Washington State University, Pullman, WA 99164; 5) School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA 99164.

Despite the importance of synapsis and recombination of homologous chromosomes to normal meiotic progression, we know little about factors that influence this process. For example, it is not known whether specific genomic sequences influence the way in which chromosomes find one another. In an initial set of studies aimed at addressing this question, we have focused on identifying sequence signatures that determine binding sites of synaptonemal complex protein 3 (SYCP3) which, in association with other proteins, forms the scaffold that holds homologs together. To accomplish this goal, we made use of chromatin immunoprecipitation (ChIP) to investigate putative SYCP3 binding sites in testicular samples from male/female mice, male rhesus macaques and human male/females. Antibodies against SYCP3 were used in conventional pull-down experiments and the DNA that was isolated was amplified by whole genome amplification and sub-cloned into plasmids. Subsequently, clones were sequenced form mouse and macaque to validate the ChIP pull down and to obtain initial data on SYCP3-associated sequence signatures. Subsequently, ChIP DNA was sequenced on 454 GS FLX platform to generate Mbs of sequence data for in-depth identification of SYCP3-associated sequence signatures using custom computational scripts. Preliminary data suggest that in both male mouse and macaque SYCP3 binds to small interspersed elements (SINEs); B1 in mouse and AluY in macaque. These specific elements are the most actively retro-transposed SINEs in each species, suggesting a dual functional role for SYCP3; i.e., it may not only act as the anchoring point for the SC but also as a suppressor of retro-transposition expansion of SINEs. This suppressive role might play a role in primate evolution and as well in the genesis of human genomic disease and meiotic errors during male meiosis.

## 149

**Differences in cell cycle control contribute to aneuploidy in mammalian oocytes.** *S. Nagaoka[1,2], C.A. Hodges[3], T.J. Hassold[1], P.A. Hunt[1].* 1) School of Molecular Bioscience, Washington State University, Pullman, WA; 2) NIH Biotechnology Program, Washington State University, Pullman, WA; 3) Department of Pediatrics, Case Western Reserve University, Cleveland, OH.

Errors during human female meiosis are the leading cause of fetal demise and birth defects (e.g., Down syndrome). In mitotic cells, the spindle assembly checkpoint (SAC) plays a pivotal role in ensuring successful chromosome segregation by monitoring spindle formation and chromosome attachment. Although the high rate of errors in human females raises questions about the efficacy of this control mechanism, recent studies suggest that the SAC components are present and functional in the oocyte. In the current study, we used a mutation that severely reduces homologous recombination as a tool to study cell cycle control in the oocyte. MLH1 is a DNA mismatch repair protein that is involved in homologous recombination. The number of meiotic crossovers is severely reduced in Mlh1 mutant mice, and both males and females are sterile. We reported previously that the presence of numerous unpaired univalent chromosomes at MI causes meiotic arrest in oocytes from Mlh1 null females (Woods et al, 1999). We now report that this phenotype is background-dependent; transfer of the mutation to the C3H background results in wildtype rates of polar body extrusion, and these background differences are related to the ability of univalent chromosomes to form stable bipolar attachments. Analysis of meiotic progression in oocytes from C3H Mlh1 null females revealed a delay in anaphase onset and demonstrated a delay in both the formation of a stable spindle and in chromosome organization. Remarkably, normal metaphase with all chromosomes aligned at the equator was virtually never observed, and live cell imaging revealed anaphase onset even when one or more chromosome was located at or behind the spindle poles. Further, an analysis of kinetochore formation suggested that only a proportion of univalent chromosomes are able to separate sister kinetochores to allow bipolar attachment. Taken together, our data suggest that the oocyte can evade SAC-mediated arrest if a sufficient number of chromosomes form bipolar attachments. This is in marked contrast to mitotic cells, where a stable bipolar attachment of all chromosomes is necessary for anaphase onset to occur. Although the conclusion that the stringency of SAC-mediated control is reduced in the oocyte represents a major paradigm shift in cell biology, this finding provides a logical explanation for the high frequency of errors during female meiosis.

## 150

**Identification of Genomic Features Which May Play A Role in Chromosome Nondisjunction.** *T. Oliver[1], E. Feingold[2,3], S. Tinker[1], N. Hollis[1], A. Locke[1], S. Sherman[1].* 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, United States of America; 2) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America; 3) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America.

Trisomy 21, caused by nondisjunction (NDJ) of chromosome 21 in ~95% of cases, is the most frequent autosomal trisomy observed among infants. Our goal was to identify genomic features on 21q which may play a role in chromosome NDJ using two approaches: 1) examining human genome resources with bioinformatic tools and 2) densely mapping recombination breakpoints along chromosomes 21 involved in NDJ. In the first approach, the UCSC Genome Browser was used to examine the distribution of genomic features associated with 1) increased recombination and 2) altered chromatin conformation in normally segregating chromosomes. These include GC content, CpG content, nucleosome exclusion sequences, simple tandem repeats, RefSeq Genes and UniGene Clusters. From our previous work, we know that recombination is increased within medial 21q in normally segregating maternal chromosomes 21 compared with NDJ chromosomes. In cases of maternal meiosis I (MI) and meiosis II (MII) NDJ errors, there is excess telomeric and pericentromeric recombination, respectively. Thus, we compared the distribution of genomic features in medial 21q ( between 23.7 Mb and 30.1 Mb on 21q) to those within telomeric (i.e. the distal 2.5 Mb of 21q) and centromeric (i.e. the proximal 3.2 Mb of 21q) 21q. Our results suggest that the frequency of genomic features in medial and centromeric 21q appeared similar, whereas all genomic features examined were increased in telomeric 21q. This suggests that the concentration of genomic features associated with targeting recombination in the telomere may influence recombination-associated MI errors. For example, the presence of an excess of nucleosome exclusion sequences in this region suggests that it may be more accessible to recombination proteins. In contrast, the lack of recombination-associated genomic features within centromeric 21q suggests other risk factors. This is consistent with our finding that MII errors with a pericentromeric exchange are associated with increasing maternal age, which suggests the involvement of other meiotic processes. For our second approach, we have genotyped 1534 SNPs along chromosome 21 in over 900 maternal errors and will use these data to compare identified recombination breakpoints with our genomic findings to ultimately further our understanding of the role of genomic features in chromosome NDJ.

† Trainee Award Finalist

## 151

**The Effect of Translocation-Induced Nuclear Re-organization on Gene Expression.** *L. Harewood[1], F. Schütz[1,2,3], S. Boyle[4], P. Perry[4], M. Delorenzi[2,3], W. Bickmore[4], A. Reymond[1].* 1) Ctr for Integrative Genomics, UNIL Sorge, Lausanne 1015, Switzerland; 2) National Center of Competence in Research (NCCR) "Molecular Oncology"; 3) Swiss Institute of Bioinformatics (SIB), 1015 Lausanne, Switzerland; 4) MRC, Human Genetics Unit, Institute of Genetics and Molecular Medicine, Edinburgh EH4 2XU, Scotland, UK.

Chromosome organization within the nucleus is thought to have an impact on gene expression. To study the effect that balanced chromosomal rearrangements have on gene expression, we compared the transcriptomes of cell lines from control (cytogenetically normal) and t(11;22)(q23;q11) carrying individuals. This translocation between chromosomes 11 and 22 is the only recurrent constitutional non-Robertsonian translocation in humans. The number of differentially expressed transcripts between the translocated and control cohorts was significantly higher than that observed between control samples alone, suggesting that balanced rearrangements have a greater effect on gene expression than normal variation. Altered expression on the translocation chromosomes was limited to chromosome 11-mapping genes. Consistently, we showed that the nuclear position of the derivative chromosome 11, but not that of the derivative chromosome 22, was significantly altered compared to its normal counterpart, suggesting that expression changes of chromosome 11 genes are potentially due to their transposition into an anomalous chromatin environment. Our results are consistent with recent studies, which indicate that nuclear position plays a functional role in regulating the expression of some genes in mammalian cells. Rearrangements may also have implications on reproductive separation, as we show that reciprocal translocations not only provide partial isolation for speciation, but also result in significant changes in transcriptional regulation through alteration of the relative nuclear position of chromosome territories.

## 152

**Mechanisms of subtelomeric breakage and repair.** *K. Rudd, V. Periapoilan, K. Hermetz.* Dept Human Genetics, Emory University School of Medicine, Atlanta, GA.

Structural rearrangements at chromosome ends are a major cause of birth defects and intellectual disabilities, but only a few have been characterized at the molecular level. To uncover the mechanisms of subtelomeric double-strand break (DSB) and repair, we performed the first large-scale analysis of subtelomeric breakpoints via combined high-resolution array comparative genome hybridization (CGH), breakpoint cloning, and comprehensive sequence analysis. Using a custom oligonucleotide array, we mapped subtelomeric rearrangements in 39 patients corresponding to 49 DSBs at 20 chromosome ends; all array CGH findings were confirmed by fluorescence in situ hybridization (FISH). We analyzed 20 terminal deletions, 2 terminal duplications, 12 unbalanced translocations, and 5 deletions adjacent to inverted duplications, resolving breakpoints to a few hundred basepairs at each site. To isolate the repaired genomic structure and identify DSB sites, we are cloning and sequencing breakpoint junctions, with 13 junctions sequenced to date. In our analysis of translocation junctions, we find that the predominant mechanism of repair is non-homologous end joining, but we have also identified a recurrent translocation between highly identical translocating segments (95% identical across ~5 kb), implicating homologous recombination in at least some rearrangements. Several terminal deletion breakpoints have microhomology to the telomere repeat (TTAGGG)n, suggesting that complementarity to the RNA component of telomerase may be involved in the synthesis of a new telomere. Inverted duplications adjacent to terminal deletions are detected via array CGH as a terminal loss and an adjacent gain. In 3 cases we cloned and sequenced the breakpoint junctions, confirming that the duplicated segment is inverted and directly adjacent to the terminal deletion. We identified regions of inverted microhomology at the breakpoints, consistent with a hairpin/strand-invasion repair mechanism. Our large collection of defined subtelomeric breakpoints also allows us to identify sequence motifs that may promote genomic instability at subtelomeres; we find sequence motifs predicted to assemble secondary structures, like G-quadruplex DNA and inverted repeats. This detailed analysis of subtelomeric breakage and repair will unravel the mechanisms of structural rearrangement at these and other loci that give rise to human copy number variation.

## 153

**Evolutionary History and Associated Chromosomal Abnormalities of a Segmental Duplicated Superstructure found on Chromosome 17.** *D. Chen[1, 2], T. Miettinen[2], V. Leppä[2], O. Choi[3], A. Palotie[2, 4, 5, 6], L. Peltonen[2, 5, 6], J. Saarela[2].* 1) Univ California, Irvine, Irvine, CA; 2) Institute for Molecular Medicine Finland FIMM, Univ of Helsinki, Helsinki; Finland; 3) Department of Human Genetics, Univ California, Los Angeles, Los Angeles, CA; 4) Department of Clinical Chemistry, University of Helsinki, Helsinki; 5) Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA; 6) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK.

Human chromosome 17 is enriched for segmental duplications, structural variations & chromosomal abnormality. We have previously identified a segmental duplicated superstructure on the 17q arm. The structure consists of 13 discrete genomic domains significantly enriched for homologous sequence fragments and is found to associate with several retrotransposons. Two transposable elements, AK125814 and AK125932 are found to retropose over 30 times were mapped exclusively within the superstructure. Our present study is to delineate evolutionary history of duplicated structure and to analyze contribution of duplicated structure to 17q chromosomal abnormality. To delineate evolutionary history of duplication structure on chr17, we analyze genome maps of 10 additional mammals (chimpanzee, orangutan, rhesus monkey, dog, cow, horse, mouse, rat, opossum and platypus). Only genome maps of primate show evidence of complex duplication superstructure. Like in human genome, footprints of 2 retrotransposon homologues are tightly associated with corresponding segmental structures in primate. The association of the 2 retrotransposons to corresponding duplicated structure across evolutionary time suggests potential roles for retrotransposons serving as initial templates for accelerated chromosomal evolution. To analyze contribution of duplicated structure to structural abnormality on 17q, pairwise sequence comparisons are conducted between all possible duplicated domain pairs. Over 300 sequence fragments pairs are found to be highly homologous (>90%)and over 500bps in size. These sequence pairs serve as potential templates for non-allelic homologous recombination (NAHR) resulting in chromosomal structural changes. Through literature searches, we find several well characterized chromosomal abnormalities matched with specific homologous sequence pairs. These include deletion of NF1 gene contributing to neurofibromatosis, deletion of exon1 & 2 of BRCA1 resulting in breast cancer and deletion of 17q21.31 loci resulting in microdeletion 17q21.31 syndrome. In summary, segmental duplicated superstructure on 17q appears to be primate specific and associated with 2 expressed retrotranspons. The active expression of 2 retrotransposons suggests evolution of duplicated superstructure to be an active, on going, process. Lastly, associations of several diseases with duplicated superstructure suggests potential role of structure in generation of chromosomal abnormality.

## 154

**ALT-immortalized human cells are critically dependent on Fanconi Anemia Protein D2 limiting of BLM-dependent telomeric recombination and amplification.** *H.A. Root[1,2], A. Larsen[1,3], D. Bazett-Jones[1,3], M.S. Meyn[1,2,4].* 1) Program in Genetics & Genome Biol, Hospital for Sick Children, Toronto, ON, Canada; 2) Dept of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 3) Dept of Biochemistry, University of Toronto, Toronto, ON, Canada; 4) Dept of Paediatrics, University of Toronto, Toronto, ON, Canada.

Fanconi anemia (FA) is an inherited disorder characterized by bone marrow failure, cancer predisposition and congenital abnormalities. FA proteins have been implicated in homologous recombination (HR), a process involved in Alternative Lengthening of Telomeres (ALT) pathways. A core complex of eight FA proteins is required to promote FANCD2 monoubiquitnation and foci formation. FANCD2 foci form spontaneously during S-phase and in response to multiple types of DNA damage. FANCD2 colocalizes and interacts with other HR proteins, however the precise role of FANCD2 in HR has remained elusive. Here we report data suggesting that FANCD2 plays a critical role in human ALT cells by constraining the activity of BLM, the Bloom syndrome helicase, so as to limit BLM-dependent telomeric recombination and amplification events.

FANCD2 localizes to telomeric foci and PML bodies in ALT, but not in telomerase-positive cells. FANCD2 almost always localizes to telomeric foci containing BLM, and telomeric localization of FANCD2 is BLM dependent. FANCD2 depletion using multiple different siRNAs and ALT lines yields a consistent phenotype: ALT-specific telomere dysfunction characterized by increases in telomeric DNA synthesis, entanglements, recombination events, and association with DNA damage response proteins. Amplified telomeric DNA in FANCD2-depleted cells appears to be primarily extrachromosomal, and accumulates both outside of and within PML bodies. We have previously reported that overexpression of BLM causes a similar phenotype of rapid, large scale ALT-specific amplification of telomeric DNA. siRNA depletion of FANCD2 does not cause overexpression of BLM, however co-depletion of BLM with FANCD2 completely suppresses the telomere phenotypes caused by FANCD2 knockdown, suggesting that the FANCD2-depletion phenotype requires functional BLM. We are now examining the structure of telomeric DNA in FANCD2-depleted cells to gain insight into the mechanisms through which FANCD2 normally regulates ALT. Our working hypothesis is that FANCD2 restrains BLM-dependent recombination and amplification of telomeric DNA in ALT cells by limiting the production of ssDNA gapped regions in telomeric DNA, and/or by affecting the stability of recombinational intermediary structures.

## 155

**Beyond counting copies: direct visualisation of CNVs on stretched DNA fibers. Application to the diagnostics of facioscapulohumeral dystrophy (FSHD).** *P. Walrafen[1], K. Nguyen[2,3], A. Vannier[1], E. Renard[1], C. Vovan[2], C. Chaix[1], R. Bernard[2,3], A. Bensimon[1], N. Lévy[2,3].* 1) Diagnostics Division, Genomic Vision, Paris, France; 2) Département de Génétique Médicale, Hôpital d'enfants la TImone, AP-HM, Marseille, France; 3) Inserm UMR_S910 "Génétique Médicale et Génomique Fonctionnelle", Université de la Méditerranée, Faculté de Médecine, Marseille, France.

The importance of CNVs was widely overlooked until technological developments allowed to efficiently address them, leading to a new picture where they play a greater role than SNPs in genetic diversity. Most of these techniques, with the notable exception of FISH, provide a bulk quantification of copy numbers, averaged over a great number of cells. This inherently limits sensitivity to rare events in a context of heterogeneous cells, e.g. mosaicism or tumor cells, and no information on the physical organization of multiple copies is provided. FISH does provide useful physical cartography and single-cell analysis, but with low resolution (hundreds of kbs). Yet, as is obvious for balanced translocations introducing fusion genes, the physical organization of sequences is critical.

Physical cartography by Molecular Combing (MC) relies on the visualization of fluorescent probes hybridized on stretched DNA fibers. The constant stretching factor and long molecule length allow direct reading of the physical organization of huge loci (several Mbs) at high resolution (~1 kb), while high DNA density and linearity of fibers allow for automatic scanning of hundreds of genome.

The first diagnostic application of this technology is facio-scapulo-humeral dystrophy (FSHD), one of the most common muscular dystrophies, with autosomal dominant transmission. FSHD is associated with the contraction of a repeat array in the subtelomeric region of chromosome 4q (4q35). Because of a non-pathogenic variant of 4q and of the presence of a homologous array on 10q26, it is critical that the copy number of the repeat sequence, be counted within its chromosomal context. Routine testing involves southern blotting, but results are unreliable and it is a painstaking technique. We have developed an MC test for FSHD and transferred the technology in a routine diagnostics lab. Initial results of the ongoing clinical validation not only show a perfect correlation with previous diagnostics, but also allowed to identify unreported haplotypes and complex rearrangements. Automation has reduced hands-on time to allow routine use of the test. MC, which is readily adaptable to other pathologies, is thus suitable for routine diagnostics or research investigation of CNVs with added physical organization data.

**156**

**Assessing copy number variation in 1,184 individuals from 11 world populations using multiple SNP/CNV array platforms.** *J. Nemesh[1], C. Barnes[2], JM. Korn[1,3,4,5,6], M. Hurles[2], SA. McCarroll[3].* 1) Broad Institute of MIT and Harvard, Cambridge, MA; 2) Wellcome Trust Sanger Institute, Cambridge, UK; 3) Harvard Medical School Department of Genetics, Boston, MA; 4) Graduate Program in Biophysics, Harvard University, Boston, MA; 5) The Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA; 6) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA.

Association studies for copy number variation (CNV) rely on accurate maps of the CNVs segregating in human populations - their genomic locations, allele frequencies, and relationships to other markers. To create a reference data resource of particular accuracy and utility, we used two SNP/CNV array platforms (the Affymetrix SNP 6.0 and Illumina 1M arrays) to analyze CNVs in a pan-ethnic population sample of 1,184 individuals from 11 world populations (the “HapMap3 phase 3” samples).

We reconciled and integrated genome-wide analyses of CNV from two different arrays to develop enabling data sets that support CNV analysis on both platforms. Since each array ascertains copy number at different positions across the genome, combining these data sets has many advantages, including experimental confirmation of CNVs by a distinct array-based approach and increased resolution of CNV breakpoints. We merged probe intensity data from the two arrays, and then used and evaluated multiple CNV discovery algorithms (Birdsuite, QuantiSNP) to nominate CNV regions. Building upon earlier work, we developed new clustering and correlation-based algorithms to resolve the locations of segregating, common CNVs within these regions. We further developed algorithms for genotyping these CNV segments -- assigning a measurement of integer copy number in each sample. The two array platforms gave strongly consistent profiles of CNV in each sample, though their joint utilization frequently yielded genotype data more accurate than that from either platform on its own. The resulting data resource contains high-quality data on 851 CNVs with an allele frequency >= 1%, including estimates of integer copy number for each CNV in 1,184 individuals. 94.9% of the CNVs analyzed were diallelic and thus amenable to standard SNP based analysis. The 812 diallelic CNVs had a mean call rate of 99.2%, a mean Mendel error rate of 0.1%, and no region had a HWE p-value <0.001. For 44 multiallelic CNVs, Fisher’s *h* measure of heritability showed copy number to be highly heritable (*h* not significantly different from 1.0)

We describe several applications of this resource, including: (1) population-genetic analysis of CNV in these populations; (2) more-powerful CNV analysis in GWAS using Illumina or Affymetrix arrays; and (3) extension of GWAS data to CNVs via imputation.

## 157

**A universal SNP panel for individual identification.** *K.K. Kidd[1], J.R. Kidd[1], E. Straka[1], W.C. Speed[1], R. Fang[2], F. Hyland[2], M.R. Furtado[2], A.J. Pakstis[1].* 1) Dept Genetics, Sch Med, Yale U, New Haven, CT; 2) Genetic Systems Division, Applied Biosystems, Foster City, CA.

After testing more than 500 potential candidate SNPs, we have assembled a panel of 92 SNPs useful for individual identification (II) and have made the list available at the laboratory website (http://info.med.yale.edu/genetics/kkidd/92snpJan2009.pdf) and the individual population allele frequencies available in ALFRED (http://alfred.med.yale.edu). All of these SNPs have average heterozygosities of >0.4 and Fst values <0.06 on the 44 population samples we studied (average 56 individuals per group) representing the major regions of the world. The 92 SNP list includes 45 essentially unlinked SNPs distributed across the human autosomes. These 45 SNPs constitute an excellent panel for individual identification providing match probabilities in the range of $10^{-15}$ to $10^{-18}$ in the 44 population samples tested, values comparable to those for the 13 CODIS STR markers. The unlinked status of these 45 SNP also makes them useful for situations involving close biological relationships. All 92 SNPs have been reliably typed by TaqMan; how best to multiplex specific subsets to use for different identification tasks will likely depend on the application. No meaningful departures from Hardy-Weinberg ratios were seen for any SNP in any population. Linkage disequilibrium (LD) values (r2) for all pairwise combinations of SNPs in all populations showed only chance levels of LD for almost all pairs (including all 45 unlinked SNPs) (median r2= 0.012; mean r2=0.029). Only seven pairs involving 12 SNPs show strong LD in most populations; they are closely linked. Separating out 6 of those SNPs as alternatives leaves 86 SNPs with no significant pairwise LD. Using all 86 IISNPs for individual identification gives match probabilities ranging from $10^{-31}$ to $10^{-35}$ for the 44 population samples. While it is certainly desirable to confirm the utility of this IISNP panel in other populations than those we have been able to study, we expect that this marker set will have essentially the same characteristics useful for individual identification in other large human populations. Supported by U.S. NIJ grants 2004-DN-BX-K025 and 2007-DN-BX-K197.

## 158

**ANALYSIS OF THE GENETIC VARIATION FROM FIVE HAPMAP INDIVIDUALS OBTAINED BY WHOLE-GENOME, SECOND-GENERATION SEQUENCING RECAPITULATES SIGNATURES OF RECENT NATURAL SELECTION AND DEMOGRAPHIC HISTORY.** *F.M. De La Vega[1], F.C.L. Hyland[1], S. McLaughlin[2], A.R. MacBride[3], K. Bryc[4], A. Auton[4], E.F. Tsung[2], H. Peckham[2], C. Scafe[1], C. Lee[2], G. Costa[2], M. Barker[1], M.G. Reese[3], A. Clark[4], C.D. Bustamante[4], K. McKernan[2].* 1) Genomic Systems R&D, Life Technologies, Foster City, CA; 2) Genomic Systems R&D, Life Technologies, Beverly, MA; 3) Omicia, Inc. Emerville, CA; 4) Cornell University, Ithaca, NY.

Ultra-high throughput sequencing is becoming a cost-effective method for the analysis of human genomes to discover genetic variation that could have implications in health and disease. We analyzed the SNPs and structural variants from the genomes of five diverse individuals of the HapMap panels; an African-American, a CEPH European, a Mexican, and two Yoruba individuals. Whole-genome sequencing was performed with the Applied Biosystems SOLiD System using mate-pair libraries. We identified over 3.5 million SNPs per individual genome: ~80% are present in dbSNP and the remaining are either novel or personal SNPs. SNPs are under-represented in exons as compared to introns/intergenic regions. Of the coding SNPs, 54% are silent, 45% are missense, and 0.6% are nonsense. We categorized the functions of genes using the Panther ontology, and annotated the damaging potential of non-synonymous SNPs (nsSNPs) using predictions from PolyPhen. About 20% of nsSNPs in this sample are predicted to be damaging. There are fewer damaging SNPs in homozygote than heterozygote state, consistent with the role of purifying selection, this reduction being statistically significant as compared with benign SNP zygosity. We identified nsSNP alleles previously associated with human disease (OMIN database), and found very few in homozygous state and none of highly penetrant Mendelian diseases. We report the first individual genomes of admix individuals - this data is likely to be important for future genome-wide association studies in these populations. Admixture change points are derived from SNP allele data leveraging previous genotyping information of the parents of these samples. Recent studies of genetic variation of functional significance in individual genomes have so far mostly focused on SNPs. However, it is becoming clear that structural variation can have functional implication in gene integrity and function. We studied the impact of large indels on gene integrity, by looking for breakpoints regions that overlapped with gene regions. Of the disrupted genes we identified, ~15% are contained in a curated collection of 3,600 human disease genes; the functional or disease impact of these events is currently unknown. Our results suggest that much more genetic variation remains to be uncovered in human populations, in particular structural, which must be considered to obtain a complete picture of their functional impact in individual genome sequences.

## 159

**Complex genetic History of human populations in East Africa.** *J. Hirbo[1,2], A. Ranciaro[2], S. Omar[3], M. Ibrahim[4], S. Tishkoff[2].* 1) Dept. of Biology, University of Maryland, College Park, Maryland 20742; 2) Department of Genetics and Biology, University of Pennsylvania, 428 Clinical Research Building 415 Curie Boulevard; 3) Kenya Medical Research Institute, Centre for Biotechnology Research and Development (CBRD) Nairobi, Kenya; 4) Dept. of Molecular Biology, Institute of Endemic Diseases, University of Khartoum, Khartoum, Sudan.

Results from disparate fields indicate that anatomically modern Homo sapiens (AMHS) originated in Africa ~200 thousand years ago (kya), and that East Africa is the likely source of migration of modern humans out of Africa within the past 100,000 years. However, Africa, especially east Africa, has not been well studied for human genetic diversity compared to non-African populations, probably due to the fact that DNA samples from many regions of Africa are currently not available. The goal of this study was therefore to sample 1500 individuals from previously unstudied east African populations in order to characterize their mitochondrial DNA (mtDNA) and Y chromosome genetic variation. These data were then compared to independently collected data of the same populations from ~1327 nuclear markers (848 microsatellites and 479 insertion/deletion polymorphisms). The data was used to gain insight into patterns of genetic diversity, to construct past relationships of East African populations to each other and to other African populations, to clarify historical demographic events such as population expansion, contraction, and migration that these populations might have experienced. Population phylogenetic analysis, two dimensional plots, regression analysis and mantel tests of genetic distance versus geographic and linguistic distance, STRUCTURE analysis and AMOVA analysis all showed that there is significant correlation between genetic and geographic/linguistic distances among East African populations. The genetic correlation with geography is stronger than it is with linguistics. Overall, the correlations between genetic versus geographic/ linguistic variation is stronger for autosomal and paternal lineage data than for maternal lineages. Paternal and maternal lineage distributions seem to cluster geographically and for some lineages, linguistically. Two major migration events, the migration of Bantu-speaking populations from Central/West Africa across sub-Saharan Africa and the migration of pastoralist populations from Sudan and Ethiopia within the past 5,000 years has had a major influence on extant genetic patterns in East Africa.

## 160

**Explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa.** *M. DeGiorgio[1], M. Jakobsson[2], N.A. Rosenberg[1,3].* 1) Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Evolutionary Biology, Uppsala University, Uppsala, Sweden; 3) Department of Human Genetics, University of Michigan, Ann Arbor, MI.

Studies of worldwide human variation have discovered three trends in summary statistics as a function of increasing geographic distance from East Africa: a decrease in heterozygosity, an increase in linkage disequilibrium (LD), and a decrease in the slope of the ancestral allele frequency spectrum. Forward simulations of unlinked loci have shown that the decline in heterozygosity can be described by a serial sampling model, in which populations migrate outward from Africa through a process in which each of a series of populations is formed from a subset of the previous population in the outward expansion. However, this model has not yet been used to investigate patterns of LD or the allele frequency spectrum. We have extended this model by developing a retrospective coalescent-based serial sampling framework that incorporates linked loci. We find that our new model not only recovers the observed decline in heterozygosity with distance from Africa, it also produces the patterns observed in LD and the ancestral allele frequency spectrum. We generalize our model to include migration between neighboring populations and admixture between modern and archaic humans, describing the effects that these phenomena have on the predictions of the model. By developing a simpler coalescent-based population divergence model inspired by our serial sampling model, we illustrate that the three trends observed with increasing distance from Africa can be explained by the cumulative effect of genetic drift as humans colonized the world.

## 161

**Genetic Variation Based on SNPs, CNVs and RoHs in European Populations.** *A. Metspalu[1,2,3], M. Nelis[1,2,3], R. Mägi[3], P. Palta[3], F. Zimprich[4], D. Toncheva[5], L. Peltonen[7], M. Lathrop[8], T. Meitinger[10], B. Melegh[11], D. Toniolo[12], P. Gasparini[13], J. Klovins[14], V. Kucinskas[15], J. Lubinski[16], S. Limborska[17], X. Estivill[18], S. Antonarakis[19], C. van Duijn[20], M. Macek[6], M. Krawczak 9, M. Remm 2,3 T. Esko 1,2,3.* 1) Estonian Genome Project, Univ Tartu/EBC, Tartu, Estonia; 2) Estonian Biocentre, Tartu, Estonia; 3) Institute of Molecular and Cell Biology, University of Tartu, Estonia; 4) Department of Clinical Neurology, Medical University of Vienna, Vienna, Austria; 5) Department of Medical Genetics, Medical University of Sofia, Sofia, Bulgaria; 6) Department of Biology and Medical Genetics, Cystic Fibrosis Centre, University Hospital Motol and 2nd School of Medicine, Charles University Prague, Prague, Czech Republic; 7) Wellcome Trust Sanger Institute, Cambridge, UK and the Institute of Molecular Medicine, Biomedicum Helsinki, Finland; 8) Commissariat à l'Energie Atomique, Institut Genomique, Centre National de Génotypage, Evry, France; Fondation Jean Dausset-CEPH, Paris, France; 9) PopGen Biobank, University Hospital Schleswig-Holstein, Campus Kiel, Germany; 10) Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 11) Department of Medical Genetics and Child Development, University of Pécs, Pécs, Hungary; 12) Division of Genetics and Cell Biology, San Raffaele Research Institute, Milano, Italy; 13) Medical Genetics, Department of Reproductive Sciences and Development, IRCCS-Burlo Garofolo, University of Trieste, Italy; 14) Latvian Biomedical Research and Study Center, Riga, Latvia; 15) Department of Human and Medical Genetics, Vilnius University, Vilnius, Lithuania; 16) International Hereditary Cancer Center, Pomeranian Medical University, Szczecin, Poland; 17) Institute of Molecular Genetics, Russian Academy of Science, Moscow, Russia; 18) Center for Genomic Regulation (CRG-UPF) and CIBERESP, Barcelona, Spain; 19) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 20) Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands.

Using principal component (PC) analysis, and other methods we studied the SNPs, CNVs and RoHs based genetic variation of more than 3,800 individuals from 17 European (21 cohorts) populations using Illumina 318K/370CNV genotype data. Those analyzes revealed that the genetic structure based on SNP alleles correlates closely with European geography. The first PC highlights a Northwest to Southeast gradient of genetic similarity whereas the second PC differentiates the Finnish population from both the Baltic region and Central Europe. The resulting genetic map differs from that obtained in previous studies in that it forms a triangular structure with a) Finland, b) the Baltic region, Poland and Western Russia, and c) Italy as its vertexes, and with d) Central- and Western Europe in its centre. Inter- and intra- population genetic differences were quantified by the inflation factor lambda ($\lambda$) (range 1.00 to 5.34) and by the number of markers exhibiting significant allele frequency differences in pair-wise population comparisons. When the PC analysis was confined to the 1,019 Estonian individuals (0.1% of the Estonian population), a fine geographical structure emerged that correlated with the individual counties within 50 km radius. Regarding to the CNVs we used 13 European populations (genotyped with Human370CNV). Here the PC analysis was not as clear as in the SNP case, but all 5 first PC components reflect partially the SNP results. Finally, human genome harbors the long streches of homozygoze SNP's: Runs of Homozygosity (RoH). We have analyzed RoH patterns in about 3,800 individuals from 17 populations (21 cohorts). Our analyzes reveal huge variety of RoH alleles in every single population and covering the whole genome, although the mean number and lenght varies. This phenomenon is possibly due to shared ancestral alleles and such blocks can be useful for studying common diseases and for reconstructing ancestral pedigrees. The transmission of RoH alleles was checked in 50 Estonian and HapMap3 trios. In addition to SNP's, CNV's and RoH's can also be used as additional genetic markers in GWAS studies for disease gene mapping. Together with previously published data, our results allow the creation of a comprehensive, European genetic map of the structural variations that will greatly facilitate inter-population genetic studies.

## 162

**ABRAHAM'S CHILDREN IN THE GENOME ERA: MAJOR JEWISH DIASPORA POPULATIONS COMPRISE DISTINCT GENETIC CLUSTERS WITH SHARED MIDDLE EASTERN ANCESTRY.** *L. Hao[1,6], G. Atzmon[2], C. Velez[1], A. Pearlman[1], B. Riley[1], C. Campbell[1], M. Sasson[1], P. Palamara[3], J. Shan[4], D. Reynolds[4], E. Friedman[5], B. Morrow[4], C. Oddoux[1], I. Pe'er[3], E. Burns[2], H. Ostrer[1].* 1) Department of Pediatrics, Human Genetics Program, NYU Langone Medical Center, New York, NY; 2) Department of Medicine, Albert Einstein College of Medicine, Bronx, NY; 3) Department of Computer Science, Columbia University, New York, NY; 4) Department of Genetics, Albert Einstein College of Medicine, Bronx, NY; 5) Chaim Sheba Medical Center, Tel Hashomer, Israel; 6) Center for Genome Informatics, New Jersey Medical School, University of Medicine and Dentistry of New Jersey, Newark, NJ.

Despite residence all over the world, Jewish populations have maintained continuous genetic, cultural, and religious tradition over 4,000 years. The unique ethnic makeup and social practices provide an invaluable opportunity to understand their genetic origins and migrations and to elucidate the genetic basis of complex disorders. To generate a comprehensive HapMap of ethnically diverse, healthy Jewish populations, we used the Affymetrix array 6.0 to genotype 381 samples recruited from 7 Jewish communities with different geographic origins: Eastern European Ashkenazim; Italian, Greek and Turkish Sephardim; Iranian, Iraqi, and Syrian Mizrahim (Middle Easterners). Here, we present population structure results from compiled datasets after merging with the *Human Genome Diversity Project* and the *Population Reference Sample* studies, which consisted of 146 non-Jewish Middle Easterners (Druze, Bedouin and Palestinian), 30 northern Africans (Mozabite from Algeria), 1547 Europeans, and 653 individuals from other African, Asian, Latin American, and Oceanian populations. Both principal component analyses and multi-dimensional scaling analysis of pairwise *Fst* distance show that Jewish populations form a cluster clearly distinct from all major continental populations. The results also reveal a finer population substructure in which each of 7 Jewish populations studied here form distinctive clusters - in each instance within group *Fst* was smaller than between group, although some groups (Iranian, Iraqi) demonstrated greater within group diversity and even sub-clusters, based on village of origin. By pairwise *Fst* analysis, the Jewish groups are closest to Southern Europeans (i.e. Tuscan Italians) and to Druze, Bedouins, Palestinians. Interestingly, the distance to the closest Southern European population follows the order from proximal to distal: Ashkenazi, Sephardic, Syrian, Iraqi, and Iranian, which reflects historical admixture with local communities. STRUCTURE results show that the Jewish Diaspora groups all demonstrated Middle Eastern ancestry, but varied significantly in the extent of European admixture. There is almost no European ancestry in Iranian and Iraqi Jews, whereas Syrian, Sephardic, and Ashkenazi Jews have European admixture ranging from 30%~60%. Analysis of identity-by-descent provides further insight on recent and distinct history of such populations. These results demonstrate the shared and distinctive genetic heritage of Jewish Diaspora groups.

## 163

**Scanning for Selective Sweeps via the Differentiation of Populations.** *H. Chen[1,2], N. Patterson[2], D. Reich[1,2].* 1) Dept Genetics, Havard Med Sch, Boston, MA; 2) Broad Institute of MIT and Harvard, Cambridge MA.

Selective sweeps can increase the genetic differentiation among populations and cause the departure of allele frequency spectra from neutrality. In this study, we present a novel likelihood method for detecting selective sweeps that involves jointly modeling the multi-locus allele frequency differentiation between two populations at a local region. We use Brownian motion to model genetic drift under neutrality, and a deterministic model to approximate the effect of a selective sweep on SNPs around the vicinity. We test the method with extensive simulated data, and demonstrate that the method provides higher power than previous reported approaches to detect ancient fixed selective sweeps. The method can accurately locate the causal mutant position. Our technique uses the allele frequency differentiation between populations, and so is robust to ascertainment bias in SNP discovery. We apply this method to whole-genome data comparing pairs of human populations including European (CEU), Han Chinese (CHB), Japanese (JPT) and Yoruba (YRI) from the HapMap Phase II Project, and Northern and Southern European Populations, and we control the false discovery rate to address the challenge of multiple hypothesis testing. Our analysis identifies a list of genes as candidate targets of selection, including well known selected loci (EDAR, SLC24A5 and LCT etc). It also highlights a substantial number of new regions for further investigation.

## †164

**Toward a more Uniform Sampling of Human Genetic Diversity: A Survey of Worldwide Populations by High-density Genotyping.** *J. Xing[1], W.S. Watkins[1], Y. Zhang[1], D.J. Witherspoon[1], C.D. Huff[1], T.S. Simonson[1], R.B. Weiss[1], S.R. Woodward[2], B.J. Mowry[3], L.B. Jorde[1].* 1) Dept Human Gen, Univ Utah, Salt Lake City, UT. 84112, USA; 2) Sorenson Molecular Genealogy Foundation, Salt Lake City, UT. 84115, USA; 3) Queensland Centre for Mental Health Research, The Park-Centre for Mental Health, Brisbane, Australia.

High-throughput genotype data are useful for making inferences about human evolutionary history on a fine scale. However, populations sampled to date are distributed unevenly across the globe, and some areas (*e.g.* western Asia, the Middle East) have rarely been sampled in large-scale studies. To assess human genetic variation more evenly, we sampled 323 individuals from 14 worldwide populations which are not covered by previous studies, including populations from western Asia, southern Asia, western Africa, Polynesia, and America. These samples were genotyped for more than 900,000 SNPs using the Affymetrix 6.0 microarray. To study the effect of the additional populations on human genetic diversity, we combined this new dataset with the HapMap III data (~800 samples from 9 populations) and an earlier dataset generated in our lab (~ 300 samples from 20 populations), resulting in a final dataset of more than 1,300 individuals. Principal components analyses and frappe analyses revealed that individuals from each population can usually be distinguished based on high-density SNP genotypes, and genetic distances are in general correlated with geographic distances among populations. Importantly, we found a substantial decrease in the degree of genetic differentiation (*Fst* = 10.3%) when populations are sampled more evenly than, for example, the populations from the HapMap II project (YRI, CHB/JPT, CEU, *Fst* = 15.2%). More detailed analyses of Eurasian populations suggest complex genetic structure in populations from Central and South Asia, including samples from Kyrgyzstan, Mongolia, Nepal, Pakistan and India. Central Asian samples have a smaller average haplotype block size and a lower average pairwise LD (measured by D' and r2) compared with the European and East/Southeast Asian samples. These results suggest high rates of gene flow in Central Asia, consistent with the position of this region as a cross-road between East and West. Our results provide insight in the patterns of human genetic variation in previously under-sampled regions. In addition, genotypes generated in this study from intermediate populations can serve as a resource for future studies of human genetic variation.

## 165

**Modeling population structure using sparse factor analysis.** *B.E. Engelhardt[1], M. Stephens[2,3].* 1) Department of Computer Science, University of Chicago, Chicago, IL; 2) Department of Statistics, University of Chicago, Chicago, IL; 3) Department of Human Genetics, University of Chicago, Chicago, IL.

Understanding the genetic structure of natural populations is a common problem in population genetics, and has important application for avoiding spurious associations due to population stratification in genetic association studies. Three different types of tools are used: distance-based clustering methods (e.g., neighbor-joining trees); model-based clustering methods (e.g., the software packages Structure or BAPS); and dimension-reduction techniques such as principal components analysis (PCA). PCA has the advantage that it can be easily applied to large numbers of SNPs: when applied to genome-wide SNP data, PCA can elucidate fine-scale structure, to the level of distinguishing among individuals with ancestors from different European countries. However, PCA is not robust to the sampling scheme used, so results can change significantly based on sampling different numbers of individuals from different populations. Nor are the principal components always interpretable as a single population, so admixed individuals may be difficult to identify. We present a new method for analyzing population structure, which combines the computational advantages of PCA with the advantages of interpretation and robustness to sampling scheme of the model-based approaches. The method is a model-based dimension-reduction technique called sparse factor analysis, which represents K different populations as a set of K factors, and each individual's genotype is a sparse linear combination of those factors. Individuals who have ancestry in only one population typically have membership in only one factor, whereas admixed individuals have membership in multiple factors. The model shares representational depth with Structure, but is potentially more easily applicable to large data sets. Being model-based, sparse factor analysis is more robust than PCA to sampling scheme. We illustrate the potential of this approach by applying it to subsets of the HDGP-CEPH human genome diversity cell line panel. When applied to genotypes of the 210 unrelated HapMap phase II individuals, our model correctly identifies each of the three populations (African, Asian, and European), each with a single factor. We illustrate how the results differ from PCA in that each individual has appreciable membership in only one factor, as opposed to multiple PCs. We show the potential for our method to deal with admixture by applying it to the complete set of 1137 individuals from 52 different populations.

## 166

**Population Genetic Inference from Low Coverage 1000 Genome Data.** *A. Auton[1], R. Hernandez[2], M. Przeworski[2], G. McVean[1,3], The 1000 Genomes Project.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Department of Human Genetics, University of Chicago, Chicago, Illinois 60637; 3) Department of Statistics, University of Oxford, Oxford, UK.

The rise of short-read DNA sequencing promises to provide exciting new insights into genetic variation. In the 1000 Genomes Project, a large number of individuals have been sequenced at relatively low coverage (2-6X). This design is excellent for discovery of novel genetic variants in the population, but can be problematic for many population genetic studies as it is difficult to obtain accurate genotypes. This is due to the probability of observing both alleles of a genotype at a given position being a function of coverage. Inaccuracy in the genotype calls can cause traditional estimates of diversity to be biased. For this reason, we have developed a probabilistic method for estimating levels of population genetic diversity directly from short-read alignments. Our method integrates over the uncertainty in the data at each site in the sample, and hence does not require genotypes to have been called prior to the estimation of diversity. In this framework, we are able to estimate measures of polymorphism such as SNP density, Pi and sample allele frequencies in an unbiased fashion. We have applied our method to the whole genome and show variation in diversity at multiple scales. We have also investigated spatial patterns of diversity around a number of genomic features including genes, repeat elements, and recombination hotspots.

† Trainee Award Finalist

## 167

**Genomic Profiling of Human LINE-1 Variation.** *A.D. Ewing, H.H. Kazazian Jr.* Department of Genetics, University of Pennsylvania, Philadelphia, PA.

Human-specific LINE-1 elements (L1Hs) are the only active family of LINE-1 retrotransposons in the human genome. Throughout evolutionary history there has been a succession of L1 families which through their copy-and-paste replication mechanism are directly responsible for ~17% of genomic sequence. The active L1Hs family is still inserting new copies into the genome as evidenced by ~250 polymorphic insertion sites catalogued in several studies and at least 18 known cases of a de novo L1 insertion contributing to human diseases. These known polymorphic and private insertions are likely to represent only a small fraction of the total amount of genomic variation induced by L1 activity. To address this possibility, we have developed a targeted resequencing assay whereby the locations of Human-specific LINE-1 elements can be obtained for an individual genome. Following amplification of the 3' flanking sequences via a PCR-based approach, amplicon ends are sequenced on the Illumina Genome Analyzer platform yielding several million reads per individual. These reads are analyzed via a computational pipeline whereby clustered alignment peaks indicate the presence of human-specific L1 insertions that are then validated by PCR. This technique is applicable to any similar targeted resequencing project involving other transposable elements or repetitive genome features. Our results indicate that in excess of 100 L1 insertions are present in any given individual that are not found in the reference human genome sequence. Additionally, we are able to utilize the data available from the 1000 Genomes Project to validate insertion sites from individuals whose genomes have been resequenced. Studies using this technique to directly measure the rate of retrotransposition in humans and examine L1 allele frequencies at the population level are underway.

## 168

**LINE-1 variation in human genomes.** *C. Beck[1], P. Collier[3], C. Macfarlane[3], M. Malig[4], J.M. Kidd[4], E.E. Eichler[4], J.V. Moran[1,2].* 1) Departments of Human Genetics, and Internal Medicine, University of Michigan Medical School, Ann Arbor MI; 2) Howard Hughes Medical Institute; 3) Department of Genetics, University of Leicester, Leicester, UK; 4) Department of Genome Sciences and Howard Hughes Medical Institute, University of Washington, Seattle, WA.

The average human genome contains ~80 to 100 LINE-1s that are capable of moving via the copy-and-paste mechanism known as retrotransposition. Here, we have used a fosmid paired-end strategy to characterize full-length LINE-1 insertions that are present in 6 geographically diverse individuals, but are absent from the human genome reference sequence (HGR). The majority (62/69) of these LINE-1s are novel with respect to a database of known retrotransposon insertion polymorphisms (dbRIP). The 69 elements then were evaluated for hallmarks of retrotransposition as well as their ability to mobilize in a cell culture retrotransposition assay. The majority (37/69) of full-length LINE-1s found in this study are highly active in the retrotransposition assay, increasing the number of known active elements by ~4-fold. The allele frequency and population distribution of a subset (26/69) of these elements has also been determined. L1s first are typed on a panel of 129 unrelated Northern European CEPH individuals. If the element is absent from these individuals, it subsequently is typed in 72 Zimbabweans. Finally, if absent from both genotyping panels, the LINE-1 is typed on the Human Genome Diversity Panel (HGDP) comprising ~1050 individuals from ~51 worldwide populations. This scheme has identified 4 LINE-1s that are absent from the CEPH and Zimbabwean panels. We determined that 2 of these elements are African specific and 2 of them are absent from the H952 subset of the HGDP, signifying that they are population restricted or perhaps private alleles. Thus, these data suggest there are likely many more L1 source loci in extant humans than previously expected, and that LINE-1 continues to generate inter-individual variation.

## 169

**Genome-wide transposon mapping reveals extensive polymorphism in L1 and its role in diseases.** *C. Huang[1,2,5], K. Burns[3,4], S. Lu[5], T. Niranjan[5], S. Wheelan[4,6], C. Schwartz[7], D. Valle[5], T. Wang[5], H. Ji[6], J. Boeke[1,2,4,5].* 1) HiT Center, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Molecular Biology & Genetics, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD; 4) Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, MD; 5) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 6) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 7) JC Self Research Institute of Human Genetics, Greenwood Genetic Center, Greenwood, SC.

Transposons are mobile genetic elements found in genomes across virtually all life form on earth. They play an important role in shaping the modern human genome. This is an ongoing process as a subset of them has remained active and moves around within the genome. We developed a technique, termed Transposon Insertion site Profiling Chip (TIP-Chip), with high density genomic tiling microarrays, that allows us to comprehensively locate transposons in human genomes. This involves extensive computational analysis of the genome sequence to achieve whole-genome coverage and stringently defined conditions to selectively amplify all the unique transposon flanking sequence from only the polymorphic transposons, but not from older, fixed families. In parallel, we have implemented an algorithm to automatically detect candidate insertions from the array data. L1 retrotransposons comprise ~17% of the human genome and possess the ability to significantly alter neighboring gene expression levels and mRNA structure. Using this technology, We have successfully profiled the two transcriptionally active L1 subfamilies (L1(Ta), L1 pre-(Ta)) and a human endogenous retrovirus family (HERV-K) in a number of individuals, and have lead to the discovery of many novel transposon insertion points. Our result showed extensive polymorphism in L1 between different individuals. This unexpected degree of variation in insertion points, for L1 in particular, suggests that their contribution to genetic diversity, heritable disease and oncogenesis are likely underestimated. To further investigate this, we profiled a group of male, presumptively X-linked mental retardation (XMR) patients and normal individuals. The TIP-Chips yielded many previously uncharacterized candidate insertions some of which are unique to those patients and are not found in 400 male controls with normal cognitive function. A number of insertions are located in genes known to cause mental retardation when mutated, suggesting that transposon insertions might explain a significant fraction of MR cases. Functional studies of each candidate insertion are currently being conducted. Preliminary results are promising and showed potential alteration of gene expression at the insertion loci in some cases. We expect this technique to be applied to study many other disease processes and expand our understanding to the intimate relationship between transposons, genome evolution and human phenotypic diversity.

## 170

**Outcome of array CGH analysis for 255 subjects with intellectual disability (ID) and search for candidate genes using bioinformatics.** *Y. Qiao[1,2], C. Harvard[1], C. Tyson[3], X. Liu[4], C. Fawcett[3], J. Holden[4], S. Lewis[2], E. Rajcan-Separovic[1].* 1) Dept Pathology (Cytogenetics), Child and Family Research Inst, UBC, Vancouver, BC, Canada; 2) Dept Medical Genetics, Child and Family Research Inst, UBC, Vancouver, BC, Canada; 3) Cytogenetics Lab, Royal Columbian Hospital, New Westminster, BC, Canada; 4) Dept Psychiatry, Queen's University, Kingston, Ontario, Canada.

We used high resolution oligo (Agilent, Nimblegen) and 1 Mb BAC (Spectral Genomics) array CGH to screen for copy number variants (CNVs) in 255 subjects with idiopathic intellectual disability (ID). Forty-seven unique CNVs (not reported in the Database of Genomic Variants, http://projects.tcag.ca/variation/) were identified in 43 probands. All unique CNVs were FISH-confirmed; 20 were de novo and 27 were familial. Nine de novo and 3 familial CNVs overlapped with known genetic microdeletion/duplication syndromes. The CNV size ranged from 680kb to 9.7Mb for de novo (average 3.2Mb) and 148 kb to 1.7 Mb for familial CNVs (average 0.6Mb). The de novo CNVs contained 1 to 97 genes/CNV, while the familial CNVs contained 0 to 21 genes. To facilitate identification of potential candidate genes for ID within the de novo and familial CNVs, we used 5 computational candidate gene prioritization softwares (Endeavor, GeneWanderer, ToppGene, PosMed and Suspects) and 6 different sets of training genes, known to be associated with ID (selected from Ensembl, OMIM, Decipher, Suspects, GenTrepid as well as obtained by manual literature searching). In a pilot experiment, we used all the web prioritization tools and training gene lists to create an *overall* top 5 candidate ID gene list for eight CNVs and to compare it to the gene priority lists obtained with each of the five tools and six training sets individually for the same CNVs. The *overall* top 5 ID candidate gene list for each of the CNVs was most comparable to the priority list obtained with Endeavour, regardless of the training set. Based on this finding, the gene prioritization for the remaining CNVs was continued using Endeavour and the Ensembl list of training genes. Using pathway website tools (e.g. WebGestalt), we found that the top 5 prioritized ID genes from all de novo CNVs frequently contribute to gene ontology categories such as development, regulation of biological processes and transcription regulation and show recurrent participation in Mitogen Activated Protein Kinases (MAPK) pathways and neurite/axon outgrowth. In contrast, no gene ontology category or pathway was enriched for the prioritized ID candidate genes from the unique familial CNVs. The benefits of using gene prioritization and bioinformatics tools for identifying causative genes or pathways contributing to the neuropathophysiology of ID from de novo and familial CNVs will be discussed.

## 171

**Using a mouse Copy Number Variation (CNV) -engineering system to study neurobehavioral phenotypes in two genomic disorders SMS and PTLS.** *W. Gu[1], M. Heney[1], W. Bi[1,2], J. Yan[1], C. Spencer[1], R. Paylor[1], J.R. Lupski[1,3,4].* 1) Molecular and Human Genetics, Baylor College Med, Houston, TX; 2) Medical Genetics Laboratories, Baylor College Med, Houston, TX; 3) Pediatraics, Baylor College Med, Houston, TX; 4) Texas Children's Hospital, Houston, TX.

In the post-genomic era, more and more CNVs have been identified that are responsible for Mendelian and complex traits and the susceptibility to these traits. How gene CNVs mediate these traits is, however, barely understood. Smith-Magenis syndrome (SMS) and Potocki-Lupski syndrome (PTLS) are two prototypical genomic disorders caused by reciprocal deletions (SMS) and duplications (PTLS), or sometimes point mutation in the major dosage-sensitive gene RAI1 (SMS); both disorders manifest a broad spectrum of phenotypes. We have established a unique CNV-engineering system that includes mouse models mimicking a diversity of genetic conditions found in patients including deletion of different sizes, duplication and point mutations in RAI1. By combining these alleles, we can obtain CNV of the SMS/PTLS region or of the single gene Rai1 ranging from 0 to 4 copies. The different sizes of deletions also allow us to observe the effect of flanking regions on the CNV manifestation, perhaps through alterations of chromatin structure. In this study, we address the neurologic and behavioral phenotypes of SMS and PTLS including mental retardation, circadian rhythm distortion, self-injury and autism. We explored the learning and memory with Conditioned Fear and Morris Water Maze assays, their pain sensitivity with hotplate and tail-flicking assays, social novelty by tube tests and the circadian rhythms by recording their activity in the dark-dark conditions after being established in light-dark cycles. We found that a substantial portion of the human phenotypes can be recapitulated in our mouse models. In the effort to utilize these models to study the disease pathways of SMS and PTLS, we measured the long-term potential (LTP) of some of the strains to search for the physiological basis of the cognitive deficiency caused by CNV. Expression profiling experiments are being performed to search for the molecular basis for the same phenotypes. A wealth of CNVs have been identified in recent years, our mouse models and these experiments is a unique pilot study to begin to systematically investigate the physiological and pathological pathways downstream of these CNVs and may provide insights into how CNVs can perturb neuronal networks and elicit cognitive phenotypes.

## 172

**Integrated Genome-wide Datasets Identify Neuronal Processes Affected in Autism.** *H. Kilpinen[1,2,3], K. Rehnström[1,2,4], E. Jakkula[1,2], D. Greco[5], S. Ripatti[1], T. Varilo[1,2,4], L. von Wendt[6], I. Hovatta[3], L. Peltonen[1,2,7,8].* 1) Institute for Molecular Medicine Finland (FIMM); Helsinki, Finland; 2) Unit of Public Health Genomics, National Institute for Health and Welfare; Helsinki, Finland; 3) Research Program of Molecular Neurology, University of Helsinki; Helsinki, Finland; 4) Department of Medical Genetics, University of Helsinki; Helsinki, Finland; 5) Institute of Biotechnology, University of Helsinki; Helsinki, Finland; 6) Unit of Child Neurology, Hospital for Children and Adolescents; Helsinki, Finland; 7) The Broad Institute; Cambridge, USA; 8) Wellcome Trust Sanger Institute; Hinxton, Cambridge, UK.

Despite recent progress in genome-wide approaches in mapping of complex traits, the genetic risk factors for autism spectrum disorders (ASD) have remained largely unsolved. In this study, we have attempted to fully utilize the possibilities of genomics and the special genetic features of the Finnish founder population in the search for rare, potentially high impact autism susceptibility variants enriched in this isolate. We carried out a genome-wide association (GWA) scan in an internal population isolate of Finland ($n_{cases}$=51), and replicated all suggestive findings in our nationwide autism dataset ($n_{cases}$=76). In addition, we analyzed global gene expression profiles of a subset of the autism patients ($n_{cases}$=10), and combined this data with the GWA data through pathway and transcription factor analysis. Finally, we utilized publicly available autism GWA and gene expression datasets (AGRE and GSE6575) to further assess the general significance of our findings. Based on the GWA results, the genetic risk factors of autism show locus heterogeneity even within genealogically connected individuals from a population isolate. However, when GWA data was combined with transcript profile data through pathway analysis, we were able to pinpoint several relevant biological processes, such as nervous system development and neuronal cell adhesion in the etiology of autism. In addition, we identified two transcription factors, NR3C1 and HIF1A, which putatively function together and regulate a significant number (~18%) of the p=0.01 level genes within all datasets in this study. The target genes of this regulatory module overlap significantly with the genes identified in the pathway analysis, and are involved in small GTPase mediated signal transduction and neuronal cell adhesion. In agreement with previous studies, the results of this study suggest that the genetic risk of autism is unlikely to be explained by common variants but rather by rare family-specific variants. We have successfully applied a systems-wide approach to a complex phenotype, as exemplified by the identification of functionally relevant biological processes, of which neuronal cell adhesion has repeatedly been implicated in the etiology of ASDs.

## 173

**Identifying Head And Neck Squamous Cell Carcinoma Signaling Pathways From Loss-Of-Heterozygosity/Allelic Imbalance Regions.** *G. Bebek[1,3], M.S. Orloff[1,2], C. Eng[1,2,3,4].* 1) Genomic Medicine Institute, Cleveland Clinic, Cleveland, OH; 2) Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH; 3) Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH; 4) Department of Genetics, Case Western Reserve University, Cleveland, OH.

Head and neck squamous cell carcinoma (HNSCC) arises from a series of genetic alterations in the squamous epithelium of the tissue. Similarly, nonmalignant stroma surrounding the tumor epithelium is known to acquire genomic alterations and contribute to cancer progression. In an earlier genome-wide loss-of-heterozygosity/allelic imbalance (LOH/AI) screening of laser capture microdissected epithelium and stroma of HNSCC lesions, 70 markers were observed to have significantly different LOH/AI frequency ($p<0.026$). Assuming that genes in proximity of identified LOH/AI regions are correlated in the tumorigenic phenotype, we expect to see differential mRNA expression levels and changes in the interaction patterns among relevant genes. Therefore we sought to data mine related publicly available biological data to facilitate identification of pathway segments and understanding the role of tumor microenvironment (stroma) in HNSCC. First, we collected functional annotations of known pathway proteins and mapped them to their functional annotations to capture their underlying characteristics. Association rule (AR) mining was then used to discover patterns from these known pathways. Next, we created an extended and filtered protein-protein interaction network (PPIN) by assigning reliability scores to each interaction using microarray expression levels, PPIN topology, and protein subcellular localization data. We inferred possible PPIs by examining protein family relationships derived from multiple species. After determining the target proteins, we utilized the acquired ARs to search for likely signaling pathway segments on the PPIN. We counted ARs on each path connecting our target proteins and identified significant paths that may help understand cancer signaling. In this study, we verified the involvement of T-Cell receptor signaling pathways in HNSCC as well as associated oncogenes such as Wwox (testis, prostate, and ovary cancer), Plcb1 (breast and colon cancer), and Dag1 (cell-cell interactions between cancer and noncancer cells) with HNSCC. We also identified novel genes that may play a role in HNSCC development, e.g. Cav3, known to play a germline role in muscular dystrophy. Hence, the proposed framework reduces the search space and facilitates focus on proteins and genes that most likely are active in HNSCC only. Pathways and networks are built efficiently, utilizing widely available high-throughput data and providing powerful discovery tools for research.

## 174

**Integrating gene networks and clinical phenotypes using graphical models.** *J. Chu, S.T. Weiss, R. Lazarus, V.J. Carey, B.A. Raby.* Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA.

Though all clinical and cellular phenotypes likely arise through complex interactions of multiple genes acting in concert as defined networks and pathways, modeling these interactions for the identification of disease-susceptibility genes remains challenging. Here we describe a novel approach to identify interaction among genes underlying dichotomous disease phenotypes based on Graphical Gaussian Models (GGMs). Using microarray expression data, we construct a GGM gene network for each disease state for which a posterior probability of interaction is estimated for each pair of genes. We then compare the posterior probabilities for each gene pair across the two disease states, expressed as an analytic posterior odds-ratio (postOR) for each pair. Though most gene pairs have similar probabilities regardless of disease state (and thus have postORs approaching 1), gene pairs with extreme postORs represent differentially interacting gene pairs and can be used to identify network components most relevant to disease status. We demonstrate the method by comparing gene networks developed using expression data from peripheral blood CD4+ lymphocytes in 288 asthma patients with two different severity levels. Of around 4 million gene pairs, significant differences of gene-gene interaction patterns between disease states were noted for 7,003 (0.14%). We also identified several key genes with extensive differential connectivity ("hubs") such as GSTM3, NPSR1 and ARG1 which have been previously implicated in asthma pathobiology. The results are based on the effects on the whole network and could not be found with regular linear regression or pairwise-association tests. The method has been shown to be computationally efficient and work well with small sample sizes, and can also be extended to other quantitative traits and environmental variables. More importantly, results from the application of graphical models can significantly enhance understanding of the connection among diseases and genetic variations, demonstrating the value of this methodology.`.

## 175

**Bioinformatic and Experimental Characterization of Microsatellite Repeats in Transcribed Regions of the Human and Chimpanzee Genomes.** *D.E. Dickel[1], M-C. King[1, 2].* 1) Dept of Genome Sciences, Univ Washington, Seattle, WA; 2) Dept of Medicine (Medical Genetics), Univ Washington, Seattle, WA.

Microsatellites are tandem repeats of short sequence motifs (generally 1-6 basepairs) found throughout the human genome. Many microsatellites occur in human genes, where they can influence gene transcription and the functions of encoded proteins. This, combined with their high mutation rate, makes microsatellite repeats good candidates for studies of human diseases, human diversity, and recent human evolution. More than 30 microsatellites are known to expand to cause human disease, but these are only a small subset of all polymorphic microsatellites in human genes. In an effort to more completely characterize microsatellite repeats in the human genome, we used a bioinformatics approach to identify microsatellite repeats that occur in genes and are likely to be polymorphic. We used the UCSC Genome Browser Table Browser function to identify all tri-, tetra-, and penta-nucleotide repeats in Build 36.1 of the human genome that overlap any mRNA and are at least 10 units long with 100 percent purity or at least 15 units long with 90 percent purity. This search yielded 314 microsatellites. Annotating the microsatellites based on the Browser's RefSeq track indicated 61 in coding regions, 74 in 5-prime or 3-prime untranslated regions, 99 in introns of RefSeq genes and overlapping uncharacterized transcripts, and 80 of other genomic contexts. We have undertaken an analogous identification process for chimpanzee microsatellites that overlap known genes. We are genotyping the microsatellites in phenotypically normal humans of various ancestries and in chimpanzees. Of the 250 human microsatellites genotyped to date (235 autosomal or pseudoautosomal and 15 on the X chromosome), 249 are polymorphic in humans. We are working to compare human and chimpanzee microsatellite profiles with respect to species of ascertainment, number and size of alleles, heterozygosity, repeat type, and genomic context. Characterizing microsatellites that impact genes in phenotypically normal individuals can help to inform future projects that examine these repeats in disease populations.

## 176

**Mapping of the pseudoautosomal boundary in the domestic dog and comparison with mouse and human sex chromosomes.** *J.D. Degenhardt[1], A. Auton[2], A.R. Boyko[1], T. Spady[3], P. Quignon[3], B. von Holdt[4], A.G. Clark[5], R. Wayne[4], E.A. Ostrander[3], C.D. Bustamante[1].* 1) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 2) Wellcome Trust Centre for Human Genetics, Oxford University, Oxford UK; 3) National Human Genome Research Institute, National Institute of Health, Bethesda, MD. 20892; 4) Department of Ecology and Evolutionary Biology, University of California Los Angeles, CA, 90095; 5) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, 14853.

The mammalian pseudoautosomal region (PAR) is a short segment of homology on the X and Y-chromosomes. This shared region allows for pairing, and therefore proper segregation, of the sex chomosomes during male meiosis, and represents the sole region of X/Y recombination. To date this region has been mapped in only a small number of taxonomically divergent mammalian groups (including human/chimp, mouse, horse and cat). Comparative analysis of these groups has shown that there has been an independent reduction in the size of the PAR compared to the ancestral eutherian PAR in each of these groups. Here we present a novel method for the mapping of the PAR boundary using a comparison of male and female intensity data and heterozygosity from Affymetrix SNP chips and map the location of this boundary in the domestic dog and wolf. We show that the position of the PAR boundary is consistent between domestic dog and wolves and also between the domestic dog and the domestic cat suggesting that the position of the PAR boundary has been relatively stable over the last ~42 million years of carnivore evolution. However, we also find evidence of lower recombination near the PAR boundary in the domestic dog suggesting that the exact location of the boundary may show some instability. We examine the gene content of canine PAR and compare this to the human and mouse sex chromosomes. We find a high degree of similarity in the genic content of the canine PAR and human PAR and adjacent X and Y chromosomal regions but no similarity with the mouse PAR. Based on this finding we suggest a more extensive restructuring of the mouse sex chromosomes has occurred compared to other mammals. With the addition of the canine PAR we able to use a comparative genomics approach to reconstruct the evolutionary history of structural changes in the human PAR. We find two regions representing likely translocations to the human Y-chromosome, which may have disrupted the X/Y homology and lead to the shortening of the human PAR.

## 177

**Simultaneous genotype calling and haplotype phase inference improves genotype accuracy and reduces false positive associations for genome-wide association studies.** *B. Browning[1], Z. Yu[2].* 1) Department of Statistics, The University of Auckland, Auckland, New Zealand; 2) Department of Statistics, University of California Irvine, Irvine, California.

We present a new method that performs simultaneous genotype calling and haplotype phase inference. Our method employs the computationally efficient BEAGLE haplotype frequency model and can be applied to large-scale studies with millions of markers and thousands of samples. We compare our method to state of the art genotype calling methods using genotype data called with the GenCall, Illuminus, Chiamo, and Birdseed genotype-calling algorithms from the Illumina 550K and 1M arrays and Affymetrix 500K and 6.0 arrays. For Affymetrix data, our method reduces discordance rates with high-quality Illumina genotypes by a factor of 3 or more. For Illumina data, our method improves genotype accuracy and reduces missing data by an order of magnitude.

We have re-called genotype data for the Wellcome Trust Case Control Consortium Bipolar Disease study (Nature 2007;447:661-78). More than 90% of the known false positive association signals caused by genotyping artefacts in the original study are eliminated when using genotype calls from our new method. The phased haplotypes produced by our genotype calling method also eliminate a similar proportion of the false positives association signals that occur in a BEAGLE haplotypic analyses of these data (Hum Genet 2008;123:273-80). Our new genotype calling methods are freely available from www.stat.auckland.ac.nz/~bbrowning/beagle/beagle.html.

## 178

**Genome-wide Association Analysis with Stochastic Block Lasso.** *S. Kim, E.P. Xing.* Computer Sci, Carnegie Mellon Univ, Pittsburgh, PA.

A genome-wide association study involves examining a large number of single-nucleotide polymorphisms (SNPs) to identify SNPs that are significantly associated with the given phenotype, while trying to reduce the false positive rate. Researchers have proposed haplotype-based approaches that take advantage of linkage disequilibrium structure by considering a group of nearby markers for association, rather than a traditional single-marker test that assumes independence of markers. While these haplotype-based approaches showed promising results, their main disadvantage is that they used a fixed-sized window as a haplotype, even if each haplotype block has a different size, depending on where recombination events have occurred. We propose a new approach called a stochastic block lasso for association mapping that exploits prior knowledge on linkage disequilibrium structure in the genome such as recombination rate and distance between adjacent SNPs in order to identify blocks of SNPs for association. In a region with a high recombination rate, the SNPs previously linked in ancestor chromosomes are likely to be decoupled in the descendant chromosomes, whereas a segment of tightly linked SNPs is preserved in the absence of recombination during inheritance. In addition, two SNPs that are far apart in distance on the chromosome are likely to be decoupled over time by a recombination event. The SNPs that are tightly linked with each other are likely to be jointly relevant or irrelevant to the phenotype. Following a typical linear regression framework between the genotypes as covariates and the phenotype as output, our proposed model employs a sparsity-enforcing Laplacian prior for the regression coefficients, augmented by a 1st-order Markov process along the sequence of SNPs that incorporates the prior information on the linkage disequilibrium structure to probabilistically determine the boundaries of the haplotype with association. The Markov-chain prior models the structural dependencies between a pair of adjacent SNPs, and allows us to look for association SNPs in a coupled manner, combining strength from multiple nearby SNPs. We describe a sampling-based algorithm for estimating the regression coefficients. We demonstrate the performance of our method on simulated and mouse data for marker identification under whole-genome association.

## 179

**Combining gene pathways via a Markov random field model in genome-wide association studies.** *M. Chen[1], J. Cho[2,3], H. Zhao[1,3].* 1) Yale Center for Statistical Genomics and Proteomics, Yale University, New Haven, CT; 2) Digestive Diseases, Yale University School of Medicine, New Haven, CT; 3) Genetics, Yale University School of Public Health, New Haven, CT.

Genome-wide association studies (GWAS) examine a large number of markers across the genome to identify associations between genetic variants and disease. Most published studies examine only single markers, which may be less informative than considering multiple markers jointly because genes interact with each other to affect disease risk. In addition, much knowledge has been accumulated in the literature on biological pathways and interactions. It is conceivable that appropriate incorporation of different data sources will likely improve the chance of making genuine discoveries. A number of methods have been developed recently to prioritize genes using prior biological knowledge. However, very few existing pathway-based methods consider pathway topology, that is all the genes in the same pathway are treated as an exchangeable set. However, how genes are related in a pathway can be informative for GWAS data analysis. As a motivating example, we have analyzed a Jewish Crohn's disease data set. A total of 4,892 genes were mapped to over 400 pathways, and each gene's association with disease was measured by the p-value. A permutation test suggests that neighboring genes tend to have similar association evidence. To formally model the disease association dependence among neighboring genes in a pathway, we propose a Markov Random Field (MRF) model to incorporate pathway topology in GWAS data analysis. The conditional distribution of our MRF model takes on a simple logistic regression form. We propose an iterated conditional modes algorithm for statistical inference of each gene's association with disease. Samples from the posterior distribution can be effectively drawn via a Gibbs sampler,and the inference can be based on commonly used criteria such as maximum a posteriori probability, maximum marginal probability, and maximum conditional probability. Simulation studies show that our proposed framework is more effective to identify genes associated with disease. We will also illustrate the power of our approach through its applications to real data examples.

## 180

**Genome Wide Association Study of Age-Related Macular Degeneration Identifies TIMP3 and HDL-associated alleles as new susceptibility loci.** *W. Chen[1], D. Stambolian[2], A.O. Edwards[3], K.E. Branham[4], M. Othman[4], J. Jakobsdottir[5], N. Tosakulwong[3], M.A. Pericak-Vance[6], P.A. Campochiaro[7], M.L. Klein[8], D.E. Weeks[5, 9], D. Zack[7], E.Y. Chew[10], J.R. Heckenlively[4], P.J. Francis[8], N. Katsanis[11], J.L. Haines[12], M.B. Gorin[13], G.R. Abecasis[1], A. Swaroop[4,14,15], CAPT Research Group, AMD GWAS Consortium.* 1) Department of Biostatistics, Univ Michigan, Ann Arbor, MI; 2) Department of Ophthalmology and Human Genetics, University of Pennsylvania, Philadelphia, PA; 3) Department of Ophthalmology, Mayo Clinic, Rochester, MN; 4) Department of Ophthalmology and Visual Sciences, University of Michigan, Ann Arbor, MI; 5) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh; 6) Miami Institute for Human Genomics, University of Miami Miller School of Medicine, Miami FL; 7) Department of Ophthalmology, Wilmer Eye Institute, Johns Hopkins Hospital School of Medicine, Baltimore, MD; 8) Macular Degeneration Center, Casey Eye Institute, Oregon Health & Science University, Portland, OR; 9) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 10) Division of Epidemiology and Clinical Applications, National Eye Institute, Bethesda, MD; 11) McKusick-Nathans Institute of Genetic Medicine, Department of Ophthalmology, Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD; 12) Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN; 13) Department of Ophthalmology, David Geffen School of Medicine, University of Californian Los Angeles, and Jules Stein Eye Institute, Los Angeles, CA; 14) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 15) Neurobiology-Neurodegeneration & Repair Laboratory (N-NRL), National Eye Institute, Bethesda, MD.

Age-related macular degeneration (AMD) is the most common cause of blindness in the elderly. We performed a genome-wide association study (GWAS) to evaluate the contribution of common genetic variants to disease susceptibility. We genotyped 370,404 markers in 2,157 unrelated Caucasian AMD cases and 1,150 unrelated matched Caucasian controls using the Illumina Human CNV370-Duo BeadChip. In addition, we used information generated by the HapMap consortium to impute genotypes at ~2 million additional common SNPs. We identified several genetic variations already reported to be associated with AMD. In particular, we observe strong evidence of association with SNPs at CFH ($p < 10^{-74}$), C2/CFB ($p < 10^{-21}$), C3 ($p < 10^{-9}$) and CFI ($p < 10^{-7}$), emphasizing the role of the complement pathway on disease susceptibility. We also confirm the strong contribution of ARMS2 ($p < 10^{-59}$), suggesting a potential role of mitochondrial dysfunction in AMD pathogenesis. In addition to providing strong support for these established susceptibility loci, our scan and follow-up data also provides evidence for new susceptibility loci suggesting that genetic variants near TIMP3 (a locus previously associated with early onset macular disorders) and near several HDL-associated loci contribute to the risk of age-related macular degeneration. Multi-locus analysis, including new and previously identified loci, shows 329/331 individuals with the highest risk genotypes in our sample are cases, 84% with advanced AMD. These results provide new insights into the biology of macular degeneration and should eventually lead to better disease management, prevention and treatment.

## 181

**Testing for pleiotropy in quantitative traits using data from genome-wide association studies.** *D.L. Nicolae[1], E.O. Kistner[2], N.J. Cox[1].* 1) Dept Medicine, Univ Chicago, Chicago, IL; 2) Department of Biostatistics, Univ Chicago, Chicago, IL.

Genome-wide association studies are often conducted on collections of patients in which multiple correlated quantitative traits are recorded. Examples include variables collected as risk factors for the metabolic syndrome (HDL, LDL, tryglicerides, BMI, blood pressure), and variables studied in osteoporosis (bone mineral density, quantitative ultrasound, bone geometry). The correlation among traits will lead to overlap in association signals even for false positive associations, thus confounding the ability to assess pleiotropy. We propose here a new approach for testing for pleiotropy in the presence of marginal association signal using a two-stage strategy: in the first stage, we test the association of SNPs with the quantitative traits, and then in the second stage we test for pleiotropic effects. Our approach will lead to test statistics that are asymptotically independent of the marginal association statistic, and thus evaluation of statistical significance in the second step is straightforward. We applied this method to data from the Genetics of Kidney in Diabetes (GoKinD) data set that has phenotypic information from 1648 unrelated Caucasian individuals with type 1 diabetes. Each study participant in GoKinD was genotyped using the Affymetrix SNP Array 5.0. The variables we focus on are the creatinine and cystatin C levels. We found a significant association between cystatin C levels and the SNP rs6048920 (p= 5.48e-07) located on chromosome 20 near the CST9 gene. Our test for pleiotropic effects shows also a significant association (p= 0.007) suggesting that this variant affects the measured levels of creatinine independently of the association between the two quantitative traits.

## 182

**Efficient Two-Step Testing of Gene-Gene interactions in Genomewide Association Studies.** *J.P. Lewinger, C.E. Murcray, W.J. Gauderman.* Dept Preventive Medicine, Univ Southern California, Los Angeles, CA.

Genomewide association studies (GWAS) have successfully uncovered new genes for complex traits. Most of these genes, however, have only modest effects and explain only a small proportion of the variation in the trait. Understanding the combined effects of genes and their interactions will likely be a key to further understanding of complex traits. Despite the potential importance of gene-gene (GxG) interactions, most investigators conducting GWAS only perform tests of GxG interactions on the subset of SNPs that already demonstrate a direct association with a trait. However, it is likely that many genes involved in a GxG interaction may only exhibit modest main effects that will not be detected in a main-effect screen of the genome, and thus will never make it onto a short list for subsequent targeted GxG analysis. We propose an efficient two-step analysis of genomewide association data aimed at identifying genes involved in GxG interactions that does not rely on detecting main effects. The first step of our procedure screens all GxG interactions by testing each pair of SNPs for association in the pooled sample of cases and controls. In the second step, the SNP pairs that pass the screening step are tested with the usual multiplicative test for GxG interaction. Because the screening and testing steps are independent, only a multiple testing correction for the number of the SNP pairs that pass the screening step is required to preserve the overall type I error. By contrast, exhaustive '1-step' testing of all pairwise interactions using a standard test requires a correction by an extremely large number of tests (e.g. with 1 million SNPs there are approximately 500 billion SNP pairs). We explored the performance of our 2-step method relative to the 1-step method for a wide range of scenarios using an analytic framework to compute power. We found that the 2-step test approach is always more powerful than the 1-step approach and dramatically so for scenarios of moderate intrinsic power (e.g. for an interaction OR=3 with no marginal effects and a MAF=11% at both genes the power of the 2-step approach to detect a GxG interaction with 3,000 cases/3,000 controls is 83% vs. 50% for the 1-step approach). The two-step procedure is complementary to a screen for marginal genetic effects, and thus it has the potential to uncover new genetic signals that have not previously been identified in a main effects GWAS.

## 183

**Detecting Epistatic Interactions Using Directed Acyclic Graphs in Genome-wide Association Studies.** *Y. Guan[1,2], M. Stephens[1,2].* 1) Human Genetics, University of Chicago, Chicago, IL; 2) Statistics, University of Chicago, Chicago, IL.

We present a novel statistical strategy to search for epistatic interactions among SNPs in a genome-wide association study. Our strategy contain three steps. The first step is to perform whole genome Bayesian multiple-SNP analysis, under an additive model with sparse and shrinkage priors (Guan and Stephens, Bayesian multi-SNP analysis with shrinkage priors, in preparation), to select a few hundreds candidate SNPs (S1). Informative priors on SNPs such as close to candidate genes and relevant biological pathways can be incorporated into this SNP selection procedure. Next, we infer directed acyclic graphs (DAG) on SNPs S1 and the phenotype (either case/control or quantitative), where a DAG represents a joint distribution between SNPs and phenotype. A recently developed statistical method (Guan and Stephens, Sampling directed acyclic graphs with fast mixing MCMC, in preparation) is capable of sampling DAGs of a large amount of nodes with sparse priors on the in-degree of each node. The result of DAG sampling are summarized by the posterior inclusion probability of directed edges. Those SNPs that have edges pointing to the phenotype node will be selected (S2). Finally, epistatic interactions among SNPs S2 can be identified by fitting a full interaction model of SNPs S2 with the phenotype after controlling for the marginal effects of SNPs S1. We will demonstrate the effectiveness of this strategy for detecting epistatic interactions via both simulation studies and its application on real dataset.

## 184

**Genome-wide Gene and Pathway Analysis.** *G. Peng[1], L. Luo[2], Y. Zhu[1], H. Dong[1,2], C. Amos[3], M. Xiong[1].* 1) Laboratory of Theoretical Systems Biology and Center for Evolutionary Biology, School of Life Science and Institute for Biomedical Sciences, Fudan University, Shanghai 200433, China; 2) Human Genetics Center, University of Texas School of Public Health, Houston, TX 77225; 3) Department of Epidemiology, University of Texas, M. D. Anderson Cancer Center, Houston, TX 77030.

Current GWAS have primarily focused on testing association of a single SNP at a time. To only test for association of a single SNP has limited utility and is insufficient to dissect the complex genetic structure of many common diseases. To meet conceptual and technical challenge raised by GWAS, we propose a general framework for gene and pathway-based GWAS and develop three statistics for testing association of genes and pathways with disease: linear combination test (LCT), quadratic test (QT) and decorrelation test (DT) that take correlations among SNPs within a gene or genes within a pathway into account. We calculate type 1 error of the developed statistics by large simulations. As a proof of concept, the proposed statistics have been applied to GWAS of rheumatoid arthritis in the Wellcome Trust Case Control Consortium (WTCCC) and the North American Rheumatoid Arthritis Consortium (NARAC) studies. Our preliminary results show that the proposed new paradigm for GWAS not only can identify the genes which have large genetic effects and can be found by single SNP association analysis, but also can detect new genes in which each single SNP conferred small disease risk, but their joint actions can be implicated in the development of diseases. The new paradigm can allow us to form the core of pathway definition of complex diseases and to unravel the functional bases of an association finding.

## 185

**Using prior information attained from the literature to improve ranking in genome-wide association studies.** *M. Johansson[1], Y. Li[2], J. Wakefield[3,4], M. Greenwood[2], T. Heitz[2], I. Roberts[2], H. Cunningham[2], P. Brennan[1], A. Roberts[2], J. Mckay[1].* 1) Genetic Epidemiology Group, International Agency for Research on Cancer (IARC), Lyon, Rhone-alpes, France; 2) Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello, Sheffield, S1 4DP., UNITED KINGDOM; 3) University of Washington, Department of Biostatistics, Box 357232, Seattle, WA 98195-7232, USA; 4) Department of Statistics, Box 354322, Seattle, WA 98195-4322, USA.

Advances in high-throughput genotyping have made it technically possible to analyze hundreds of thousands of single nucleotide polymorphisms (SNPs) across the whole genome. Using this technology it is now feasible to conduct genome-wide association studies (GWAS) aiming to investigate the majority of common genetic variation and relate it to some phenotypic differences, often to risk of some disease. Whilst the price of GWAS assays are decreasing rapidly, conducting a GWAS is still a very expensive exercise, typically requiring genotyping several thousands of subjects at several hundreds of euros per sample in order to gain sufficient statistical power to distinguish the true association signals from the background noise. Recognizing that a large proportion of GWAS findings reside near potential candidate genes for many of the investigated phenotypes, we here explore means to incorporate prior information attained from the literature to improve ranking in GWAS. We use this information to assign a crude prior probability of association for each SNP. The prior probabilities are thereafter integrated with the association result from the GWAS and the SNPs are re-ranked according to Baysian false-discovery probability (BFDP). We show that this methodology improves the ranking substantially for many known susceptibility loci with examples from studies on lung cancer and cancer of the upper aero digestive tract (UADT). We have implemented this methodology in a web application where a user can specify a list of keywords and receive priors for all SNPs of interest. These priors can thereafter be used to rank the SNPs according to the BFDP.

## 186

**BR-squared: a Practical Solution to the Winner's Curse in Genome-Wide Scans.** *L. Sun[1,2], S.B. Bull[3,1], L. Faye[1], A. Dimitromanolakis[1], D. Waggott[3], A.D. Paterson[4,1], The DCCT/EDIC Research Group.* 1) Dalla Lana School of Public Health, University of Toronto, Toronto, ON Canada; 2) Department of Statistics, University of Toronto, Toronto, ON Canada; 3) Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON Canada; 4) Genetics and Genomic Biology, Hospital for Sick Children, Toronto, ON Canada.

In genome-wide scans, the most significant variants detected in the original discovery study tend to have inflated effect size estimates due to the Winner's Curse phenomenon. The Winner's Curse has recently gained much attention in Genome-Wide Association Studies (GWAS), because it has been recognized as one of the major contributing factors to the failures of many attempted replication studies. For example, five Nature Genetic publications in the first three months of 2009 acknowledged the effect of the Winner's Curse in their discovery samples (e.g. Nair et al., 2009). However, none made statistical adjustments to the naïve estimates. We extend our previous work (Sun and Bull, 2005) developed in the context of genome-wide linkage analysis to provide Bias-Reduced estimates via Bootstrap Re-sampling (BR-squared) for GWAS without collecting additional data. In contrast to the likelihood-based approaches (e.g. Zollner and Pritchard, 2007; Ghosh et al., 2008; Zhong and Prentice, 2008), the proposed method adjusts for the effects of selection due to both stringent genome-wide thresholds and maximization of the association statistics over the genome. In addition, our method explicitly accounts for the effect of allele frequency because the expected bias is inversely related to power of the association test. We have implemented the bootstrap method in a user-friendly and efficient software package, BR-squared, tailored for GWAS of both binary diseases and quantitative traits. We applied our method to a number of existing GWAS including the studies of seven common diseases by WTCCC (2007) and Psoriasis by Nair et al. (2009), and an on-going study of complications of type 1 diabetes in the Diabetes Control and Complications Trial (DCCT) samples. We observed over 50% reduction in effect estimates for many associated SNPs. For DCCT, the estimated percentage of variation in HbA1c explained by the most significant SNP was reduced from ~5% to ~0.05%, and the latter estimate was confirmed by an independent dataset. As a result, the sample size required for a successful replication study (alpha=0.05, power= 80%) is increased from ~200 to ~15,000. Thus, it is crucial to adjust for the Winner's Curse when interpreting GWAS findings and planning for replication studies.

## 187

**Accelerated neurodegeneration in a polyglutamine *C.elegans* model by RNAi silencing of an ubiquitin-like gene mutated in a late-onset recessive cerebellar ataxia.** *J. Demers-Lamarche[1], I. Thiffault[1], M. Tetreault[1], S. Bel Hadj[1], M. Srour[1], M.J. Dicaire[1], J. Mathieu[2], J.P. Bouchard[3], J. Ricardson[4], J. Woulfe[5], J. Lessage[6], B. Brais[1,2], A. Parker[1].* 1) Center of Excellence in Neuromics of Université de Montréal, CRCHUM Hopital Notre-Dame, Montreal, Quebec, Canada; 2) Clinique des maladies neuromusculaires, Carrefour de la Santé de Jonquière, Saguenay, Quebec, Canada; 3) Service de neurologie, Hopital de l'Enfant-Jesus, Universite Laval, Quebec city, Quebec, Canada; 4) Neuropathology, Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada; 5) Department of Pathology and Laboratory medicine, Faculty of Medicine, Ottawa, Ontario, Canada; 6) Department of Radiology, CHUM-Notre-Dame, Montreal, Quebec, Canada.

Late-onset ataxias have largely been considered as either milder forms of dominant ataxias or sporadic diseases seemingly not caused by genetic factors. The unique population structure of the aging Quebec population, with its numerous large elderly living sib ships, provides an exceptional setting to explore the genetic bases of neurodegeneration. We have recruited a large French-Canadian cohort of late-onset cerebellar ataxia (LOCA) cases. SNP genome-wide scan uncovered linkage to a region not previously associated with ataxia or any other neurodegenerative disorders (multipoints LOD score 5.18). Sequencing of genes in the candidate region uncovered 7 mutations in a ubiquitin-like gene. We investigated the phenotypes obtained with RNA interference of the ubiquitin-like homologue in *C. elegans*. We observed that RNAi silencing in both wild-type (N2) and neurosensitive (rrf-3) strains reduced lifespan. The identification of recessive mutations in this gene suggested that abnormal protein folding also plays a role in LOCA disease so we proposed that the RNAi gene homologue should affect proteotoxicity. We tested this hypothesis with transgenic worms expressing a polyglutamine-tract of 40 residues fused to YFP. The RNAi accelerated the mean onset of paralysis of these animals compare to control and the appearance of Q40::YFP aggregation puncta. Our results suggested that down regulating the expression led to an acceleration of the degenerative phenotype that correlates with an increase in polyglutamine aggregation. Our observation agree well with a growing body of knowledge on different neurodegenerative disorders that abnormal protein folding leading to neuronal protein accumulation is a common pathogenic mechanism.

## 188

**MEF2C haploinsufficiency caused either by microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations.** *D. Bonneau[1], N. Le Meur[2], M. Holder-Espinasse[3], A. Goldenberg[2], P. Amati-Bonneau[1], H. Journel[4], JP. Kerckaert[5], P. Saugier-Veber[2], V. David[6], C. Dubourg[6], T. Frébourg[2], J. Andrieux[7].* 1) Département de Génétique et INSERM U694, CHU d'Angers, France; 2) Service de Génétique et INSERM U614, CHU de Rouen, France; 3) Service de Génétique, CHU de Lille, France; 4) Service de Génétique, CH de Vannes, France; 5) Plateforme de Génomique Fonctionnelle, Université de Lille II, France; 6) Laboratoire de Génétique Moléculaire CHU de Rennes, France; 7) Laboratoire de Génétique Médicale, CHU de Lille, France.

Array-CGH has remarkably improved the ability to detect cryptic unbalanced rearrangements in patients presenting with mental retardation, congenital anomalies or neuropsychiatric disorders and has led to the identification of several genes involved in monogenic disorders. Using whole genome oligonucleotide array-CGH, we detected de novo 5q14.3 microdeletions ranging from 216 kb to 8.8 Mb in 5 unrelated patients showing phenotypic similarities. All patients had early and severe developmental delay, hypotonia and absent speech. Stereotypic movements were present in 3/5 patients. Different types of epilepsy were observed in 3/5 children, from well controlled generalized seizures to early refractory tonicoclonic or myoclonic epilepsy. All patients displayed MRI abnormalities including corpus callosum agenesis (2/5), short corpus callosum (1/5), abnormal gyration (1/5), fronto-parietal atrophy and enlarged pericerebral spaces (1/5), enlarged lateral ventricle (2/5) or enlarged fourth ventricle (2/5). The minimal common deleted region of these 5q14 microdeletions encompassed only MEF2C, known to act in brain as a neurogenesis effector. In a patient presenting a similar phenotype, we subsequently identified a de novo MEF2C nonsense mutation. The causal role of MEF2C alteration in the phenotype observed is in agreement with the biological function of the MEF2C protein and the murine models of MEF2C inactivation. Indeed, MEF2C belongs to the Myocyte Enhancer Factor 2 (MEF2) protein family which regulates excitatory synapse number, dendrite morphogenesis and differentiation of post synaptic structures in the brain. MEF2C is the predominant isoform in the developing cerebrocortex and is highly expressed in frontal cortex, entorhinal cortex, cerebellum, dentate gyrus and amygdale. Conditional Mef2c-null mice (Li et al. PNAS 2008;105:9397-9402; Barbosa et al. 2008;105:9391-9396) have hippocampus-dependent learning and memory impairment associated to a dramatic increase in the number of excitatory synapses. These mice also display behavioural phenotypes with abnormal anxiety, decreased cognitive function, and marked paw wringing/clasping stereotypy, resulting in a Rett-like phenotype as observed in mutant Mecp2 mouse models. In conclusion, our results indicate that MEF2C haploinsufficiency caused either by 5q14.3 microdeletion or by mutation is responsible for a severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations.

## 189

**Unravelling the pathogenesis of the *ARX* homeobox mutations; role of IPO13.** *C. Shoubridge[1, 2], M.H Tan[1, 2], T. Fullston[1,2], G. McGillivray[3], G. Mancini[4], J. Gécz[1,2].* 1) Neurogenetics, SA Pathology Women's & Child Hosp, Adelaide, Australia; 2) Department of Pediatrics, University of Adelaide, Adelaide, Australia; 3) Genetics Health Services Victoria, Murdoch Children's Research Institute, Melbourne, Australia; 4) Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, Netherlands.

A frequently mutated gene causing X-linked intellectual disability is the *Aristaless* related homeobox (*ARX*) gene. Nonsense mutations, polyalanine tract expansions and missense mutations in *ARX* are implicated in a range of intellectual disability phenotypes with or without additional features including epilepsy, infantile spasms, hand dystonia, lissencephaly, autism and dysarthria. Severe phenotypes, such as X-linked lissencephaly with abnormal genitalia (XLAG), are frequently observed in individuals with missense or nonsense mutations clustered in the conserved paired-type homeodomain. In this report we have identified a novel point mutation (c.1135C>A, p.R379S) in the homeodomain of ARX in a patient with infantile spasms and intellectual disability. We investigated this and other missense mutations (R332P, R332H, R332C, T333N and R379L) in residues of the nuclear localisation sequences (NLS) flanking either end of the ARX homeodomain, associated with XLAG and Proud syndrome phenotypes. The NLS regions in the ARX homeodomain are required for correct nuclear transport due to binding to the import protein, Importin 13 (IPO13). Moreover, the arginine residues involved with these mutations (position 5 and 52 of the paired-type homeodomain) are invariant in all 26 family members containing this homeodomain and are frequent sites of missense and non-sense mutations associated with a range of diseases. In this study we demonstrate missense mutations in either the N-terminal or C-terminal NLS regions of the ARX homeodomain cause significant disruption to nuclear localisation of the mutant transcription factor protein *in vitro*. Surprisingly, none of these mutations abolished the binding of the mutant protein to the nuclear transport protein IPO13, confirmed by co-immunoprecipitation and immmunoflorescence studies. Instead, endogenous IPO13 remained bound to the mutant ARX proteins, even in the RanGTP rich nuclear environment, which normally causes the release of cargo from IPO13. We conclude in addition to inadequate accumulation and distribution of the ARX transcription factor within the nucleus, the extended sequestration of the ARX cargo with IPO13 most likely contributes to the pathogenesis observed in patients with XLAG, Proud syndrome and infantile spasms caused by mutations in the residues of the NLS regions of the ARX homeodomain.

## 190

***ZC3H14* mutations cosegregate with Non-Syndromic Autosomal Recessive Mental Retardation (NS-ARMR) in two Iranian families.** *M. Garshasbi[1], K. Kahrizi[2], A. Tzschach[1], F. Behjati[2], M. Falah[3], S. Ghasemi Firouzabadi[2], S. Esmaeeli Nieh[1], C. Goswami[1], H.H. Ropers[1], H. Najmabadi[2], A.W. Kuss[1].* 1) Max-Planck Institute for Molecular Genetics, Berlin, Germany; 2) University of Social Welfare and Rehabilitation Sciences, Tehran, Iran; 3) Tehran Welfare Institution, Tehran, Iran.

To date, only five genes have been implicated in NS-ARMR: *PRSS12* (neurotrypsin), *CRBN* (cereblon), *CC2D1A*, *GRIK2* and *TUSC3*. However, ARMR is extremely heterogeneous, and there is reason to believe that the number of underlying gene defects actually goes into the thousands. As part of an ongoing systematic study aiming to identify ARMR genes, we investigated a consanguineous family with three patients with NS-ARMR. Genome-wide SNP typing enabled us to map the relevant genetic defect to a 4.56 Mbp interval on chromosome 14. This interval contains a total of 26 genes, which we screened by Sanger sequencing for exonic and splice site mutations. We thus found a R154X nonsense mutation in *ZC3H14*, a recently described CCCH-type zinc finger gene (Leung et al., Gene 439:71, 2009). This nucleotide change is predicted to cause loss of the nuclear localization signal and was absent in 188 Iranian controls. Mutation screening in an additional NS-ARMR family with three affected sibs and a linkage interval encompassing *ZC3H14* revealed a homozygous 25 bp deletion 16 bp downstream of exon 16. Again, this variant was not found in healthy Iranian controls (n=188). Database screening revealed that the genomic segment harbouring this deletion is transcribed. We could confirm this by RT-PCR, which indicates that this transcript is part of the *ZC3H14* gene, suggesting that the 25 bp deletion is a functionally relevant second mutation. *ZC3H14* is thought to modulate post-transcriptional gene expression, as all of its four protein isoforms bind specifically to polyadenosine RNA through their common Cys3His zinc fingers. By Western blotting and immunostaining we have shown that in cell lines of patients with the R154X mutation, three large *ZC3H14* isoforms are indeed absent from the nucleus. Thus, *ZC3H14* defects may give rise to NS-ARMR by compromising mRNA binding and processing. We are now performing gene expression profiling and CLIP (cross-linking and immunoprecipitation) followed by next generation sequencing to identify specific polyA-RNA targets of *ZC3H14* and to shed more light on its function.

## 191

**Disturbed Wnt signaling due to mutations in CCDC88C causes an unclassified autosomal recessive brain malformation.** *A.B. Ekici[1], D. Hilfinger[1], M. Jatzwauk[1], U. Hehr[2], D. Wenzel[3], I. Lorenz[3], G. Staatz[3], A. Reis[1], A. Rauch[1,4].* 1) Institute of Human Genetics, University Erlangen-Nuremberg, Erlangen, Germany; 2) Department of Human Genetics, University of Regensburg, Germany; 3) Department of Pediatrics, University Erlangen-Nuremberg, Erlangen, Germany; 4) Institute of Medical Genetics, University Zurich, Zurich, Switzerland.

We performed positional cloning in a consanguineous family of North African origin with an unclassified severe autosomal recessive brain malformation. Three of five affected individuals died in early childhood but two could be analyzed, one surviving child and one fetus from a terminated pregnancy after ultrasound detection. MR and US images of the affected individuals showed enlarged ventricles, unilateral diverticle, hydrocephalus, hypoplastic cerebellum and ventrally dislocated cerebellum, mesencephalon and brain stem. Initially a phenotypic overlap with schizencephaly was suspected. Therefore mutations in Emx2 where excluded in affected family members. A genome-wide linkage analysis was performed using Affymetrix250K arrays. This homozygosity mapping resulted in a single 3.4Mb interval encompassing 33 positional candidate genes. After prioritizing for expression in brain and Emx2 involved pathways we screened the family for mutations in the CCDC88C gene (HkRP2), encoding a Hook-related protein with a binding domain for the Wnt signaling pathway protein Dishevelled. This revealed a homozygous substitution in the donor splice site of intron 29 in affected individuals. RT-PCR studies showed a reduced length of amplified product and expected exclusion of exon 29 was confirmed by sequencing. This position is highly conserved throughout evolution and sequencing of 224 controls showed no alteration at this position. Western-blot analysis showed absent protein expression in lymphoblastoid cell lines from affected individuals and reduced expression in the mother, as compared to controls. CCDC88C was ubiquitously expressed in humans, but most prominently in the fetal brain, especially in pons and cerebellum, while only cortex and medulla oblongata were expressed in adult brain. Expression profiling of 84 genes from the Wnt-signaling pathway in peripheral blood from the index patient revealed an overall reduced expression of genes from this pathway with a specific altered expression pattern consistent with a negative feed-back loop via increased beta-catenin expression. We therefore conclude that this is a bona-fide loss-of-function mutation. Finally, mutation screening in 58 other patients with schizencephaly revealed no mutation in this gene, supporting the notion that this is a novel, rare brain malformation disorder caused by mutations in CCDC88C. We thus identified a further essential component of the Wnt-signaling pathway in brain development in humans.

## 192

**A truncating mutation in *BOD1* leads to mental retardation and ovarian failure.** *S. Esmaeeli Nieh[1], I.M. Porter[2], M.M. Motazacker[1], L.R. Jensen[1], S. Freier[1], K. Kahrizi[3], C. Goswami[1], F. Behjati[3], H. Khodai[4], A. Dehghani[4], A. Tzschach[1], M. Garshasbi[1], J.R. Swedlow[2], H.H. Ropers[1], H. Najmabadi[3], A.W. Kuss[1].* 1) Max Planck Institute for Molecular Genetics, Berlin, Germany; 2) Wellcome Trust Centre for Gene Regulation & Expression, University of Dundee, Dundee, United Kingdom; 3) University of Social Welfare and Rehabilitation Sciences, Tehran, Iran; 4) Welfare Rehabilitation Center, Yazed, Iran.

In the course of systematic clinical and molecular studies to identify mental retardation (MR) causing defects that follow an autosomal recessive mode of inheritance (ARMR), we found a single 4.3 Mbp interval on chromosome 5q with a LOD score of 3.8 by autozygosity mapping in a family with four females affected by mild MR and ovarian failure. By sequencing the coding regions of all 27 genes within this region we discovered a nonsense mutation in exon 2 of the *BOD1* gene. This defect co-segregated with the disease and was not found in 720 control chromosomes. All other genes within the interval were not affected by nucleotide changes. *BOD1* is expressed in a wide range of tissues, including brain and ovary. By RT-PCR, we identified two previously unknown isoforms of BOD1 in control fibroblasts and showed expression of all four transcripts in a variety of brain tissues. Quantitative RT-PCR revealed loss of all BOD1 isoforms in patient fibroblasts, including splice variants that did not contain exon 2. This seems to be due to nonsense mediated decay, as it could be abrogated by cycloheximide treatment of the cells. We confirmed the absence of BOD1 protein in cells of the patients by Western blotting experiments. Further studies including live cell imaging showed several abnormalities in patient cell division. These observations are in keeping with previous findings that showed severe biorientation defects as a consequence of BOD1 depletion in Hela cells (Porter et al. J. Cell Biol. 179:187-197, 2007) and may also provide an explanation for the ovarian failure observed in this family. In addition, pull-down and mass spectrometry have enabled us to identify an interacting protein that links BOD1 to gene regulation, and overexpression studies in primary murine neurons indicate an extra-nuclear localization of BOD1 during interphase. The latter suggests an involvement of BOD1 in neuronal information processing and may provide a clue to the pathogenesis of MR in this condition.

## 193

**Mouse chromosome 2 harbors a locus that protects against deafness caused by hypothyroidism.** *S. Camper[1], Q. Fang[1], C. Longo-Guess[3], L.H. Gagnon[3], A. Giordimaina[1], T.W. Gong[2], M. Mustapha[1], D.F. Dolan[2], K.R. Johnson[3].* 1) Dept. Human Genetics, University of Michigan Medical School, Ann Arbor, MI; 2) Dept. Otolaryngology, University of Michigan Medical School, Ann Arbor, MI; 3) The Jackson Laboratory, Bar Harbor, ME.

Congenital hypothyroidism can cause permanent hearing defects in humans and mice if circulating levels of thyroid hormone (TH) are not corrected in neonates. TH has pleiotropic effects on cochlear development, and genomic variation influences the severity of the hearing impairment. The nature of the genetic modifiers and the ineffectiveness of later treatment are not understood. *Pou1f1*$^{dw}$ (*Pit1*$^{dw}$) mutant mice lack pituitary thyrotropin (TSH), which results in severe TH deficiency, and either profound deafness or hearing impairment, depending on the genetic background (Karolyi et al. 2007). The development of the organ of Corti is delayed in *Pou1f1*$^{dw}$ mutants, and there are numerous permanent defects including reduced potassium channel gene expression and function and abnormalities of the tectorial membrane and stria vascularis (Mustapha et al., 2009). To assess the contribution of the maternal environment to hearing impairment, we transferred fertilized eggs from a *Pou1f1*$^{dw/+}$ intercross to surrogate mothers with hybrid vigor (B6D2) and found no improvement in hearing, suggesting that the deafness is intrinsic to the fetus. To determine the genetic complexity of the protective effects, an F1xF1 intercross was generated between *Pou1f1*$^{dw}$ carriers and an inbred strain with excellent hearing, *Mus castaneus*. Approximately 16% of the mutant progeny had normal hearing. A genome scan of these individuals revealed a locus on chromosome 2, named modifier of dw hearing, *Mdwh*, that rescues hearing despite persistent hypothyroidism. This chromosomal region contains a modifier of *Tubby* hearing (*Moth1*) that encodes a protective allele of the microtubule-associated protein *Mtap1a* (Ikeda et al., 2002). We crossed *Pou1f1*$^{dw}$ carriers with two strains that carry protective alleles of *Mtap1a* and found that 129/Ola is protective for dw hearing, while AKR is not. This suggests that a protective allele of *Mtap1a* is not sufficient to rescue dw hearing. Microarray analysis identified cochlear gene expression changes caused by hypothyroidism in *Pou1f1*$^{dw}$ mice. Some of these are positional candidates for the modifier gene. We expect that identification of protective modifiers will enhance our understanding of the mechanisms of hypothyroidism-induced hearing impairment. Supported by NIDCD: DC05188, DC02982, DC05401, DC05053; and NOHR. References: Karolyi et al., Mamm Genome 18:596, 2007. Mustapha et al., J Neurosci. 291212, 2009. Ikeda et al. Nat Genet. 30:401, 2002.

## 194

**Altered expression of *DIAPH3*/*diaphanous* results in deafness as a true dominant trait in humans and impaired response to sound in *Drosophila*.** *M.M. Lesperance[1], C.J. Schoen[2, 3], S.B Emery[1], M.C. Thorne[1], E. Sliwerska[2], J. Arnett[1], H.R. Ammana[4], M. Hortsch[6], F. Hannan[4,5], M. Burmeister[2,7,8].* 1) Dept Otolaryngology/Head/Neck Surg, Univ Michigan, Ann Arbor, MI; 2) Molecular/Behavioral Neuroscience Institute, Univ Michigan, Ann Arbor, MI; 3) Neuroscience Graduate Program, Univ Michigan, Ann Arbor, MI; 4) Dept Cell Biology and Anatomy, New York Medical College, Valhalla, NY; 5) Dept Otolaryngology, New York Medical College, Valhalla, NY; 6) Dept Cell and Developmental Biology, Univ Michigan, Ann Arbor MI; 7) Dept Psychiatry, Univ Michigan, Ann Arbor, MI; 8) Dept Human Genetics, Univ Michigan, Ann Arbor MI.

We previously mapped the AUNA1 locus (deafness, autosomal dominant auditory neuropathy, 1) to 13q14-q21 using linkage analysis. Affected individuals develop delayed onset progressive deafness with functional outer hair cells as demonstrated by preserved otoacoustic emissions. Two homozygous offspring of a consanguineous marriage between affected first cousins had a phenotype of progressive hearing loss without syndromic features, indistinguishable from heterozygotes, providing a rare example of true dominance. To identify candidate genes, whole genome expression array analysis of RNA from lymphoblastoid cell lines from affected and control individuals was performed. Only one gene, diaphanous homolog 3 (*DIAPH3*), with significant expression changes also mapped to the linkage region, upregulated 2-3 fold in affected individuals. Sequencing of *DIAPH3* revealed a missense mutation in the 5'UTR, c.-172G>A, g.48G>A in all affected individuals but absent in 322 control chromosomes. Quantitative reverse transcription-PCR confirmed overexpression of the normal transcript. A number of splice variants were characterized through RT-PCR, but ratios of known splice variants were similar between affected and unaffected individuals. Northern blot and 5' RACE showed several different transcription start sites. As compared to individuals lacking the c.-172G>A mutation, affected individuals express more of the transcript with the longest 5'UTR, in which the sequence extends to the site of the mutation. Reporter assays using a portion of the promoter region and the wild-type or mutated 5' UTR revealed that the c.-172G>A mutation is sufficient to drive increased expression of luciferase in vitro. We then tested whether overexpression of *diaphanous* leads to impaired response to sound in *Drosophila*. Flies that overexpressed UAS-diaCA only in the JO neurons required for hearing had a significant decrease in sound evoked potential (SEP)($p < 0.0001$, Mann Whitney). DIAPH3 is one of 3 diaphanous-related formin proteins, which play a role in regulation of actin and stabilization of microtubules. The demonstration of overexpression in heterozygotes and homozygotes is consistent with a true dominant trait. *DIAPH3* may have an important role in maintenance of the actin cytoskeleton. Alternatively, *DIAPH3* overexpression may interfere with the function of the inner hair cell synapse, perhaps due to abnormal vesicular trafficking.

## †195

**Discovery and functional analysis of a new retinitis pigmentosa gene, *C2orf71*.** *L.M. Baye[1], D.Y. Nishimura[2], C.C. Searby[3], A. Avila[4], C. Ayuso[4], D. Valverde[5], E.M. Stone[3,6], D.C. Slusarski[1], V.C. Sheffield[2,3].* 1) Dept. of Biology, University of Iowa, Iowa City, IA; 2) Dept. of Pediatrics, University of Iowa, Iowa City, IA; 3) Howard Hughes Medical Institute; 4) Dept. of Medical Genetics, Fundacion Jimenez Diaz, CIBERER Madrid, Spain; 5) Dept. Bioquimica, genetica e inmunología, Fac Ciencias, University of Vigo, Spain; 6) Dept. of Ophthalmology, University of Iowa, Iowa City, IA.

Homozygosity mapping was performed with SNP genotyping data generated from a small nuclear family of Spanish decent with retinitis pigmentosa (RP). A large region of homozygosity was identified that spanned 21.9 Mb on the short arm of chromosome 2. A list of RP candidate genes was obtained by comparing retinal expression array data generated from of an RP mouse model lacking photoreceptor cells with retinal expression data from wild-type mice. Based on these data, three candidate genes within the human interval had preferential expression in the photoreceptor cells of the retina. Sequencing of these genes in the proband revealed a homozygous I201F mutation in the *C2orf71* gene. This mutation was not observed in the homozygous state in unaffected family members nor was it detected in over 200 control individuals. To assess functional significance of the *C2orf71* I201F mutation, both the wild-type and mutant form were over-expressed in ARPE-19 cells. Western blot analysis demonstrated that the C2orf71 protein harboring the I201F mutation was expressed at a significantly lower level than that of wild type protein. Inhibition of proteosome degradation by the inhibitor MGC115 was able to restore expression of the mutant C2orf71 protein. These data indicate that protein instability is likely to underlie the pathology of the I201F mutation in *C2orf71*. To examine the *in vivo* function of *C2orf71*, the zebrafish orthologue was identified and analysis by RT-PCR on staged embryonic cDNAs show that expression begins in the zebrafish at 48hpf, a time-point when photoreceptor cells are exiting the cell cycle; moreover, *C2orf71* is also expressed in the adult zebrafish retina. We next examined the effects of gene knockdown in zebrafish using antisense oligonucleotides (morpholinos) directed toward the translation start region of the RNA. Morpholino-injected embryos exhibited defects in vision. Together, these data support a functional role for *C2orf71* in vision and demonstrate the combined utility of small human pedigrees, gene expression data and animal models in identifying rare causes of retinitis pigmentosa.

## 196

**The role of the transcription factor C/EBPδ in a mouse model of inherited photoreceptor degeneration.** *D.C. Martin, M.J. Szego, C.C. Jiang, A. Bramall, R.R. McInnes.* Prog. in Dev & Stem Cell Biol, The Hospital for Sick Children, Toronto, Ontario, Canada.

Inherited photoreceptor degenerations (IPD) in humans result from mutations in more than 140 genes and ultimately cause blindness. In orthologous mouse models of IPDs, the mutant photoreceptors (PR) are at a constant risk of death, but the mechanistic basis of the constant risk and the identity of molecules that mediate the pathogenesis, or that resist it, are largely unknown. To identify such molecules, we performed microarray analysis of a mouse model of IPD, $Rds^{+/-}$, and identified an up-regulation of several members of the IL-6 signaling pathway including the downstream transcription factor gene, *Cebpd*. *Cebpd* encodes a bZIP transcription factor, C/EBPδ, that regulates apoptosis, cellular proliferation and immunity. By qPCR, the mRNA levels of *Cebpd* in three different mouse models of IPD, $Rds^{+/-}$, $Rd1^{-/-}$ and *Tg(RHOP347S)*, were up-regulated 2.6- (n=3, p<0.004), 3.2- (n=3, p<0.01) and 5.2- (n=3, p<0.025) fold compared to wild-type respectively. Analysis of mRNA levels in $Rds^{+/-}$ mice, using laser capture microdissection and qPCR, identified a 6.2-fold up-regulation (n=3; p<0.04) of *Cebpd* mRNA in the PR layer, whereas no significant difference in *Cebpd* expression was identified in the inner nuclear layer (n=3; p=0.34). To determine the significance of the up-regulation of *Cebpd* in IPDs, we generated an $Rds^{+/-}$ mouse with a complete deficiency of C/EBPδ ($Cebpd^{-/-}$). $Rds^{+/-};Cebpd^{-/-}$ mice showed a significant increase in the thickness of the PR layer (28 ± 1.7 μm) compared to $Rds^{+/-};Cebpd^{+/+}$ mice (22.9 ± 0.38 μm) at 8 months of age (n=5, p<0.006). This corresponded to an 18.2 % increase in the thickness of the outer nuclear layer. These findings indicate that 1) a retinal regulatory pathway controlled by C/EBPδ plays an important role, likely in the PRs themselves, in the pathogenesis of PR death in IPDs; and 2) C/EBPδ may be a suitable target for the design of therapies to inhibit PR degeneration in IPDs. Studies to determine whether C/EBPδ contributes to the pathogenesis of PR death in other IPD models, and whether the levels of pro-apoptotic proteins are decreased in the absence of C/EBPδ, are in progress.

## 197

**A Neurodevelopmental Survey of Angelman Syndrome with Genotype-Phenotype Correlations.** *J.K. Gentile[1], W.H. Tan[1], C.A. Bacino[2], S.A. Skinner[3], R. Barbieri-Welge[4], A. Baeur-Carlin[3], A.L. Beaudet[2], T.J. Bichell[5], L.T. Horowitz[3], H.S. Lee[6], T. Sahoo[2,7], S.E. Waisbren[1], L.M. Bird[4], S.U. Peters[2].* 1) Department of Psychiatry, Children's Hospital Boston, Boston, MA; 2) Baylor College of Medicine, Houston, TX; 3) Greenwood Genetic Center, Greenwood, SC; 4) Rady Children's Hospital San Diego, San Diego, CA; 5) Kennedy Center, Vanderbilt University, Nashville, TN; 6) Data Technology Coordinating Center, University of South Florida, Tampa, FL; 7) Signature Genomic Laboratories, LLC, Spokane, WA.

**Background:** Angelman syndrome (AS) is a neurodevelopmental disorder caused by a lack of expression of the maternal copy of UBE3A, due to a deletion of maternal chromosome 15, paternal uniparental disomy (UPD), imprinting defect, or UBE3A mutation. Although all AS patients are globally delayed, the neurocognitive abilities in each molecular sub-class have not been well characterized. As part of the NIH Rare Diseases Clinical Research Network AS Natural History study, we used standardized neurodevelopmental instruments to evaluate the developmental profile in children and adults with different molecular subtypes. We present the baseline data from the first 102 subjects enrolled in this study. **Methods:** All subjects were enrolled at one of four study sites (Rady Children's Hospital San Diego, Texas Children's Hospital, Greenwood Genetic Center, and Children's Hospital Boston), had a molecular diagnosis of AS, and were between 5 months and 26 years old. Subjects with a co-morbid genetic condition or extreme prematurity were excluded. Subjects were evaluated using the Bayley Scales of Infant and Toddler Development, Third Edition, Vineland Adaptive Behavior Scales, Second Edition, and the Aberrant Behavior Checklist, Community version. Deletion sizes were determined using a chromosome 15-specific comparative genomic hybridization microarray. **Results:** The median age was 36 months (80% were between 17 and 60 months). 74% of subjects had deletions, 15% had either imprinting defects or UPD, and 12% had UBE3A mutations. Individuals with deletions were more delayed in cognition, language, motor skills, and adaptive behavior compared to those with other molecular subtypes. Receptive language was more advanced than expressive language, and level of ability for receptive language was generally similar to that for motor skills. There were no consistent differences between subtypes in gross and fine motor abilities. We will present our statistical analysis of the differences in neurocognitive characteristics by molecular subtype. **Conclusions:** Individuals with AS are able to understand more than they can express. Their neurocognitive profile varies according to the molecular subtype.

† Trainee Award Finalist

## †198

**Genetic and biochemical analysis of mutations in the DMXL1 gene causing a Prader-Willi like Syndrome.** *K. Gokhale[1], B. Kulkarni[1,2], E. Chin[1], M. Adams[1], S. Vora[1], J. Schroeder[1], D. Weinshenker[1], S. Leung[3], A. Corbett[3], R. Fuller[4], S.T. Warren[1], M.R. Hegde[1].* 1) Dept Human Gen, Emory Univ, Atlanta, GA; 2) National Institute of Immunohaematology, Parel, Mumbai India; 3) Dept Biochem, Emory Univ,Atlanta,GA; 4) Dept Biol Chem, Univ of Michigan Medical Center, Ann Arbor, MI.

Prader-Willi syndrome (PWS) is a classic imprinting disorder with most cases resulting from paternal deletions of 15q11-q13 or maternal uniparental disomy 15. However, not all patients who present with PWS-like phenotype have identifiable chromosome 15 involvement, suggesting genetic heterogeneity. We have recently identified 14 novel missense mutations in the DMXL1 gene on chromosome 5 in 12.3% (14/114) of patients with a PWS-like phenotype who previously tested negative for known chromosome 15 etiologies. Analysis of parental samples from three cases showed these missense mutations to all be de novo. All but three mutations (79%) replace an amino acid conserved in DMXL1 from human to yeast. Expression of three mutant alleles of RAV1 -M998I, R1323C and ΔAAL in yeast have a dominant effect to varying degree with all the phenotypes tested. Clinical features observed in patients with a mutation in the DMXL1 gene overlap with PWS including; mental retardation, infantile hypotonia, poor suck reflex, growth retardation. The distinctive features observed in the DMXL1 patients include strabismus, childhood spasticity, microcephaly and hyperactivity. To further evaluate the function of the DMXL1 gene, we constructed Dmxl1 knockout mouse model. Homozygous knockout mice are embryonic lethal, while heterozygous mice were noted to have hyperactivity and deficits in social memory. We propose that dominant mutations, possibly impairing vesicular trafficking in the highly conserved DMXL1 gene may be responsible for a new genetic syndrome with features which initially may be suspicious for PWS.

## 199

**Perinatal complications in a U.S. population with Prader-Willi syndrome.** *J.A. Gold[1], E. Simpson[1], V.E. Kimonis[1], S.B. Cassidy[2].* 1) Division of Human and Metabolic Genetics, Department Pediatrics, Univ California Irvine CA, Orange, CA; 2) University of California San Francisco.

Introduction Prader-Willi syndrome (PWS) is characterized by prenatal-onset hypotonia, infantile poor feeding and growth, childhood-onset obesity, short stature, facial dysmorphic features, psychomotor delay, and a distinct behavioral phenotype. Deficiency of imprinted genes in 15q11-q13 cause this disorder. Method We assessed 64 patients over three days for multiple historical and clinical parameters, including detailed history of the pregnancy, labor and delivery obtained from parental report and available record review. This is a fairly unique cohort of subjects with PWS as the recruitment was aimed at equal numbers with UPD 15 and deletion 15q11-13 in order to study genotype-phenotype correlations and data was gathered with investigators blind to the genotype. There were 34 males and 30 females, ages 3-38 years, 34 with deletion on chromosome 15q (53.1%) and 30 (46.9%) with UPD. Results/Discussion From the results of the 63 patients with perinatal information, 16 were born prematurely (delivery prior to 37 weeks) (25.6 %). 32 were term (38 to 42 weeks) (47.6%) and 15 were post-dates (>41 weeks) (23.8%). 21 required pictocin induction (36.2%) and 27 had caesarean section (45.6%). Only seven had placental pathology review: Of these, three were normal, one meconium stained, one “abnormal” (no records), one small, and one had calcifications. Fetal movements were reported as subjectively decreased in 52/62 (86.6%). Mean age at diagnosis was significantly delayed in the UPD group (4.7 years) compared with the deletion group (3.2 years) ($p<0.017$). It is unclear whether this was due to the limitations of the testing that was available at that time or if it is due to a milder phenotype leading to later diagnosis. Conclusion Although the data collected may have been affected by recall bias, there is clearly a high rate of pregnancy and delivery complications, especially: Very high caesarean section rate Decreased fetal movements High induction of labor rate. Similar results were previously reported in Europe in France and the UK.This had not been reported previously in a US population. The cause of these perinatal difficulties is not known, but may relate to hypotonia and/or hormonal differences. These perinatal problems might play a role in causing characteristic early postnatal difficulties (lethargy, poor suck) and subsequent developmental and behavioral manifestations.

## 200

**Comprehensive Analysis of Human Holoprosencephaly Reveals Distinct Genotype-Phenotype Correlations.** *M. Muenke, B. Solomon, D. Pineda, J. Velez, A. Keaton, F. Lacbawan, E. Roessler.* Medical Genetics Branch, NHGRI/NIH, Bethesda, MD.

Holoprosencephaly (HPE), the most common malformation of the human forebrain, may be due to cytogenetic anomalies, teratogenic influences, or mutations in one of at least 12 identified HPE-associated genes. Here we report the results of testing of mutations in the 4 most common HPE-associated genes (SHH, ZIC2, SIX3, and TGIF) as well as additional rare HPE-associated genes and correlate these findings with detailed clinical data as well as with functional analyses. In a comprehensive analysis of approximately 1000 individuals with HPE-spectrum anomalies we find mutations in HPE-associated genes in 25-30% of probands with normal karyotypes. To date, these include 128 families with mutations in SHH, 93 with mutations in ZIC2, 64 with mutations in SIX3, and 19 with mutations in TGIF. While there is a clear overall HPE-specific phenotype, our large cohort allows us to differentiate between the clinical characteristics of patients with mutations in the various genes, and to use functional analyses to make predictions of HPE severity among patients with mutations in the same gene. We find mutations in SHH and SIX3 in multiple multigenerational kindreds, and mutations in SHH can be correlated with specific end-organ anomalies including the liver (i.e. fatty infiltration) and adrenal glands (i.e. hypoplasia). Mutations in ZIC2 are more often de novo than with any other HPE-associated genes, result in typical HPE facies in only 35% of patients, and may also result in dorsal brain and neural tube defects. Additionally, we describe a previously unrecognized facial phenotype in some patients with ZIC2 mutations consisting of bitemporal narrowing, upslanting palpebral fissures, a short anteverted nasal tip, broad and well-demarcated philtrums, and disproportionately large ears. Patients with mutations in SIX3 over-represent the more severe forms of HPE and we can predict HPE severity using a functional analysis in zebrafish. While mutations in TGIF are less common, chromosomal aberrations affecting the region occur relatively frequently and may result in sequelae due to deletions of TGIF and nearby genes. Finally, we are further able to explore alternate models of the genetic pathogenesis of HPE and because of these clinical findings we present new recommendations for the genetic testing of probands with HPE.

## 201

**The genetic burden of craniosynostosis: a 10-year cohort study of 326 patients.** *A.O.M. Wilkie[1,2,3], J.C. Byren[2], J. Jayamohan[2], D. Johnson[2], P.G. Richards[2], S.A. Wall[2], J.A. Hurst[2,3], T. Lester[4], S.J.L. Knight[5], S.R.F. Twigg[1].* 1) Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK; 2) Craniofacial Unit, Oxford Radcliffe Hospitals NHS Trust, John Radcliffe Hosp, Oxford, UK; 3) Dept Clinical Genetics, Oxford Radcliffe Hospitals NHS Trust, Churchill Hospital, Oxford, UK; 4) Genetics Laboratories, Oxford Radcliffe Hospitals NHS Trust, Churchill Hospital, Oxford, UK; 5) Oxford Biomedical Research Centre, Wellcome Trust Centre for Human Genetics, Oxford, UK.

**Background:** Craniosynostosis, the premature fusion of the cranial sutures, is one of the most common and serious disorders of craniofacial development. Genetic abnormalities are overrepresented in more severe cases, including a frequent mutation in the *FGFR3* gene (P250R), first described in 1996. Since molecular genetic testing became possible, no comprehensive analysis has been performed of the impact of genetic disorders on craniosynostosis. **Methods:** We undertook targeted molecular genetic, cytogenetic and array CGH testing in a prospective cohort of 326 children requiring surgery for craniosynostosis, born between 1993 and 2002, presenting to a single craniofacial unit, and followed up until the end of 2007. **Results:** 82 children (and 61 relatives) had a proven genetic abnormality (single gene or chromosomal); the FGFR3 P250R mutation was the single greatest contributor (26% of the genetic cases). Causative mutations were also identified in the *FGFR2* (32%), *TWIST1* (20%), *EFNB1* (7%) and *FAM20C* (1%) genes. Within the cohort born 1998-2002, which showed improved ascertainment of children without syndromic features, genetic diagnoses accounted for 21% of all cases of craniosynostosis. Children with genetic diagnoses had significantly increased risks for many complications including requirement for tracheostomy, hearing aids and upper limb surgery. Cases with single gene mutations had more severe surgical trajectories compared to those with chromosome abnormalities (which comprised 15% of genetic diagnoses), being 3.5 times more likely to require >1 major craniofacial procedure ($P$=0.01). **Conclusions:** This study provides a comprehensive picture of the relative prevalence of different mutations in craniosynostosis. Genetic analysis should be considered in all affected children, but diagnostic yield is very low in those with isolated sagittal, metopic or lambdoid synostosis (comprising half of cases), enabling laboratory testing to be more efficiently targeted. Abnormal karyotypes associated with craniosynostosis are very diverse; this, and the indolent clinical course, suggests that in many of the chromosomal cases the craniosynostosis has a biomechanical origin (reduced transduction of stretch forces exerted by the growing brain). This may undermine attempts to identify specific craniosynostosis “genes” associated with many non-recurrent abnormal karyotypes.

† Trainee Award Finalist

## 202

**Spine Abnormalities in Asymptomatic Children with Neurofibromatosis Type 1 (NF1).** *D. Viskochil[1], L. Armstrong[2], J. Eelloo[3], H. Hanson[1], S. Huson[2], Z. Mughal[2], K. Murray[1], E. Schorry[4], D. Stevenson[1], NF1 Bone Consortium.* 1) Univ Utah, Salt Lake City, UT; 2) University of British Columbia, Vancouver, Canada; 3) University of Manchester, Manchester, United Kingdom; 4) University of Cincinnati, Cincinnati, Ohio.

Children with neurofibromatosis type 1 (NF1) are at risk to develop progressive scoliosis. The etiology of this manifestation is unknown; however, associated findings include vertebral scalloping, rib penciling, dural ectasia, and paraspinal neurofibromas. To identify potential spine features in NF1 patients who develop dystrophic scoliosis, a multi-center cohort of asymptomatic prepubertal children with NF1 were screened for spine abnormalities upon entry to the study. The primary outcome of this 4-year natural history study is to determine the incidence of development of dystrophic scoliosis in a cohort of 110 prepubertal individuals with NF1. Additional bone health indices are determined as secondary outcome measures. Children with NF1 who did not have scoliosis by physical examination were enrolled through 1 of 4 NF Clinics. Each subject was evaluated by thoracic MRI, scoliosis series, DXA, and pQCT. The MRIs and scoliosis series were examined and scored by a single radiologist (KM) for the following: scoliosis >9 degrees, scalloping ratio at T5 and L1, vertebral wedging in the sagittal or coronal planes, spindling of the transverse process, rib penciling, dural ectasia, paraspinal neurofibroma, and meningocele. Eighty prepubertal enrollees have been evaluated. The age range of this cohort is 6 to 9 years, 61% are male, and 52% are sporadic cases. 41 subjects have normal imaging on both spine MRI and PA/lateral spine radiography, and 39 had at least 1 abnormality. These abnormalities include 23 with paraspinal neurofibroma, 7 with dural ectasia, 2 with meningocele, 11 with an abnormal vertebral body, 13 with measurable scoliosis between 10 and 20 degrees, and 1 with rib penciling. In summary, about half of the asymptomatic prepubertal NF1 enrollees without scoliosis on physical examination had at least 1 spine abnormality. Of those with abnormal findings, 59% have at least 1 paraspinal neurofibroma, 18% have dural ectasia, and 33% have scoliosis between 10-20 degrees.

## 203

**Vascular Endothelial Cell Dysfunction in Patients with Neurofibromatosis 1 (NF1).** *K. Jett[1], B. Birch[1], J. Case[3], S. Chan[2], M. Estes[3], T. Kydland[1], E. Lasater[3], GBJ. Mancini[2], L. Smiley[3], D. Ingram[3], J. Friedman[1].* 1) Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Department of Medicine, University of British Columbia, Vancouver, BC, Canada; 3) Department of Pediatrics, Indiana University Purdue University Indianapolis, Indiana, USA.

Vasculopathy can be one of the most serious manifestations of neurofibromatosis 1 (NF1). Although NF1 vasculopathy is often asymptomatic, the first clinical manifestation may be a life-threatening or fatal event. The pathogenesis of NF1 vasculopathy is not clearly understood but is thought to result from haploinsufficiency of neurofibromin in vascular smooth muscle cells, endothelial cells and bone marrow-derived inflammatory cells. Recent studies in genetically engineered mice indicate that Nf1+/- marrow-derived pro-inflammatory monocytic cells directly contribute to vascular disease in this model system. Pro-inflammatory monocytes have been linked to inflammatory cytokine production and increased endothelial transmigration of macrophages in other kinds of vascular disease. Based on the these observations, we used poly-chromatic flow cytometry (PFC) to test for vascular inflammatory cells in the peripheral blood of eight NF1 patients and performed flow-mediated vasodilation (FMD) and glyceryl-trinitrate-mediated dilation (NMD) to assess vascular endothelial and smooth muscle function in these subjects. Three of these individuals had known NF1 vasculopathy, one had hypertension (which may be associated with vasculopathy), and four had no history of cardiovascular disease. All eight NF1 patients had low FMD, indicative of dysfunction of the vascular endothelium. In contrast, NMD, a test of vascular smooth muscle function, was normal in all of the patients studied except one who has atherosclerotic vascular disease as well as NF1 vasculopathy. PFC analysis of peripheral blood from all eight NF1 patients demonstrated a dramatically increased population of pro-inflammatory CD16++CD14+ monocytes not observed in healthy controls. Our findings demonstrate that vascular endothelial function is frequently altered in NF1 patients. Further, we show for the first time that NF1 patients have an unusual circulating pro-inflammatory cell population that may contribute to their development of vascular disease.

## 204

**The utility of screening MRI in the identification and management of clinically significant pathology in patients with NF-1.** *C.P. Morgan[1], G. Vezina[2], M.T. Acosta[3], K.N. Rosenbaum[1], C.J. Tifft[1].* 1) Divisions of Genetics & Metabolism, Children's Natl Medical Ctr, Washington, DC; 2) Diagnostic Imaging, Children's Natl Medical Ctr, Washington, DC; 3) Neurology, Children's Natl Medical Ctr, Washington, DC.

Neurofibromatosis type 1 (NF1) results from mutations in the gene encoding the tumor suppressor neurofibromin. Patients with NF1 are at an increased risk for developing tumors of the brain, especially optic pathway gliomas (OPG) and gliomas of the brain stem. An MRI scan is the best method for the diagnosis of these tumors, but the use of "screening" MRI to identify asymptomatic tumors is controversial. The cost and sedation risk associated with performing the MRI studies in young children has been cited as too high to justify its routine practice. A National Neurofibromatosis Foundation consensus statement in 1997 recommended annual ophthalmologic exams in NF1 patients to identify symptomatic OPG. Neurological exams could be used to identify other symptomatic brain tumors. One problem with this conservative approach is that young patients, at risk for OPG in particular, may be unable to fully cooperate with these exams. In addition, previous studies did not take into consideration the utility of a single MRI to identify both OPG and other CNS tumors. The aim of this study is to determine if screening MRI is useful in identifying and managing clinically significant pathology. Patients less than 18 years of age with a diagnosis of NF1 and who had undergone neuroimaging at CNMC from 2000 through 2007 were included in this study. Patient charts and radiology reports were used to accumulate phenotypic and outcome data related to their NF1 and any brain tumors identified by MRI. Three hundred-seven screening MRI studies were performed in 188 patients at a median age of 6.3 years. Additional scans were performed in follow-up to a positive screening MRI or as a result of patient symptoms. Thirty-five screening MRI identified significant pathology in 34 patients. These included 20 OPG and 9 additional CNS tumors. Of the 20 OPG, 5 required treatment with chemotherapy. None of the 9 additional CNS tumors identified by screening MRI required treatment. In 3 patients a screening MRI identified stenosis of the internal carotid artery. Two of these cases required surgical repair. Screening MRI studies in NF-1 patients can detect CNS pathology requiring medical or surgical intervention in pre-symptomatic patients. The use of screening MRI in centers experienced in both childhood sedation and imaging NF-1 patients should be reconsidered.

## †205

***CNTNAP2* and *NRXN1* are mutated in recessive, severe mental retardation resembling Pitt-Hopkins syndrome and target a common synaptic protein in *Drosophila*.** *C. Zweier[1, 2], E.K. de Jong[2], M. Zweier[1], A. Orrico[3], L.B. Ousager[4], A.L. Collins[5], E.K. Bijlsma[6], M.A.W. Oortveld[2], A.B. Ekici[1], A. Reis[1], A. Schenck[2], A. Rauch[1, 7].* 1) Institute of Human Genetics, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany; 2) Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences, Donders Institute for Brain, Cognition and Behaviour & Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 3) Unita Operativa Medicina Molecolare, Azienda Ospedaliera Universitaria Senese, Policlinico S. Maria alle Scotte, Siena, Italy; 4) Department of Clinical Genetics, Odense University Hospital, Odense, Denmark; 5) Wessex Clinical Genetics Service, Princess Anne Hospital, Southampton, UK; 6) Department of Clinical Genetics, Leiden University Medical Centre, Leiden, The Netherlands; 7) Institute of Medical Genetics, University of Zurich, Zurich-Schwerzenbach, Switzerland.

Heterozygous copy number variations and single nucleotide polymorphisms of *CNTNAP2* and *NRXN1*, two distantly related members of the neurexin superfamily, have been repeatedly reported in association with developmental language disorders, autism spectrum disorder, epilepsy and schizophrenia, thus pointing to a shared molecular basis underlying different neuropsychiatric disorders. We now identified homozygous and compound heterozygous deletions and mutations via molecular karyotyping and mutational screening in *CNTNAP2* and *NRXN1* in four patients with severe mental retardation and variable features such as epilepsy and breathing anomalies, phenotypically overlapping with Pitt-Hopkins syndrome. With a frequency of at least 1% in our cohort of 179 patients, recessive defects in *CNTNAP2* appear to significantly contribute to severe mental retardation. In contrast to the presynaptic adhesion protein NRXN1, evidence for a synaptic function of CNTNAP2 was lacking so far. Using *Drosophila* as a model we now demonstrate that, as was known for Nrx-I, also the CNTNAP2 ortholog Nrx-IV localizes to synapses and can reorganize them by influencing density of active zones, the synaptic domains of neurotransmitter release. Moreover, we show that Nrx-I and Nrx-IV converge on a common molecular target, the presynaptic protein bruchpilot. Thus, the similar phenotypes resulting from defects in both genes in humans may result from an analogous shared synaptic mechanism.

† Trainee Award Finalist

**206**

**Genotype-phenotype correlations in a large cohort with Joubert syndrome.** *D. Doherty[1], D. Knutzen[1], T. Rue[2], M.A. Parisi[3], I.A. Glass[1].* 1) Dept Pediatrics, Univ Washington and Seattle Children's Hospital, Seattle, WA; 2) Department of Biostatistics, University of Washington, Seattle, WA; 3) National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD.

Joubert Syndrome (JS) is an autosomal recessive disorder characterized by a distinctive hindbrain malformation ("molar tooth sign"), hypotonia, developmental delay, ataxia, and variable features: cystic renal disease, retinal dystrophy, hepatic fibrosis and polydactyly. Associations between clinical features (e.g. retinal and renal disease) have been reported in the literature, as have genotype-phenotype correlations (such as *AHI1* mutations with retinal dystrophy); however, these findings have not been tested in a single large cohort with relatively uniform ascertainment. We report clinical features and genetic testing results in 216 families ascertained based on clinical findings of JS with supportive brain imaging features. The prevalence of variable JS features in our cohort is comparable to the literature: occipital encephalocele-6%, retinal dystrophy-29%, coloboma-19%, cystic renal disease-23%, hepatic fibrosis-18%, polydactyly-19%. The most highly correlated features are i) liver disease with coloboma ($r=0.50$, $p<0.001$) and ii) renal disease with retinal dystrophy ($r=0.33$, $p<0.001$), with modest correlation between polydactyly and encephalocele ($r=0.20$, $p<0.01$). Modest negative correlations were observed between liver disease and retinal dystrophy ($r=-0.20$, $p<0.05$) and liver disease and polydactyly ($r=-0.16$, $p<0.05$). The correlations between phenotypes likely reflect the relative importance of shared molecular processes in different tissues. We also report the most striking genotype-phenotype correlation in JS to date: *MKS3/TMEM67* mutations account for 80% of subjects with liver disease and 47% of subjects with coloboma in our cohort. This shared molecular cause may explain the correlation between liver disease and coloboma. In contrast, only 23% of retinal dystrophy is explained by *AHI1* mutations, while renal disease is caused by mutations in multiple genes. Based on data from the literature and our cohort, we propose the following strategy for JS gene sequencing in various clinical settings: 1) Liver disease and/or coloboma: *MKS3/TMEM67*, *CC2D2A* and *RPGRIP1L*; 2) Mild MTS: *NPHP1*; 3) Retinal dystrophy: *AHI1* and *CEP290*; 4) Renal disease: *NPHP1*, *RPGRIP1L* and *CEP290*; 5) Retinal and renal disease: *CEP290*, *CC2D2A*; 6) No other features: *AHI1*, *CC2D2A* and *CEP290*. If no mutations are detected, sequencing the remaining JS genes should be considered. Testing strategies will evolve as new genes are identified and sequencing technology improves.

## †207

**Duplication of Conserved Non-Coding sequence Elements (CNEs) - a Novel Mechanism in the Pathogenesis of Congenital Malformations.** *E. Klopocki[1], K. Dathe[1], A. Brehm[2], C.-E. Ott[1], I. Kurth[3], S. Mundlos[1,2].* 1) Charité Universitätsmedizin Berlin, Institute for Medical Genetics, Berlin, Germany; 2) Max Planck Institute for Molecular Genetics, Berlin, Germany; 3) Institut für Humangenetik, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany.

Two thirds of the sequence conserved among mammals is not protein coding. The precise function of such conserved non-coding sequence elements (CNEs) is unknown but they have been proposed to regulate time and tissue specific gene expression. CNEs that function as cis-regulatory elements are often several hundred bases in size and may exert their influence over distances greater than 1Mb. Using high-resolution array CGH we detected CNE containing duplications in patients with congenital malformations. We identified duplications of 1) a regulatory sequence of *SHH* in triphalangeal thumb-polysyndactyly syndrome (MIM #174500) and in 2) Laurin-Sandrow syndrome (MIM %135750), 3) a 5 kb regulatory element 100kb downstream of *BMP2* in brachydactyly type A2 (MIM #112600), 4) a 1.2Mb element containing several CNEs upstream of *SOX9* in Cooks syndrome (MIM %106995), and 5) a 1.7Mb and a 250kb element, respectively, upstream of *MSX2* in cleidocranial dysplasia (MIM #119600). In all cases the duplications are arranged in tandem. We show that the CNE contained in the duplication at the *BMP2* locus is an enhancer regulating *BMP2* expression in the limbs, thus, functioning as a limb-specific cis-regulatory element. We postulate that duplications of cis-regulatory elements cause an enhanced and/or deregulated expression of the target gene which in consequence disturbs dosage-dependent signalling pathways. These duplications result in unpredictable phenotypes clearly different from those associated with point mutations or deletions of the target gene.

Our data provide the molecular cause for several so far genetically unresolved conditions as Laurin-Sandrow and Cooks syndrome. Furthermore, duplications of CNEs can be considered a novel genetic mechanism for developmental defects. Given the importance of temporal-spatial gene regulation during embryonic development it is to be expected that a large number of developmental defects are caused by mutations affecting such distant enhancers/repressors.

## 208

**Mutations in *GDF5* and in its receptor *BMPR1B* cause Brachydactyly Type A1.** *D.E. Bulman[1], L. Racacho[2], A.M. Byrnes[2], H. Dranse[4], T.M. Underhill[4], J. Allanson[3], S.M. Nikkel[3].* 1) Department of Regenerative Medicine, Ottawa Hospital Research Institute, and the University of Ottawa, Ottawa, ON, Canada; 2) Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, ON, Canada; 3) Department of Genetics, Children s Hospital of Eastern Ontario, and Department of Pediatrics, University of Ottawa, Ottawa, ON, Canada; 4) Cellular and Physiological Sciences, University of British Columbia, Vancouver, BC, Canada.

Brachydactyly A-1 (BDA1; OMIM 112500) is an autosomal dominant disorder characterized by shortness of all middle phalanges of the fingers and toes, occasional terminal symphalangism, shortness of the proximal phalanges of the first digit, and short stature. To date, BDA1 has been described as a fully penetrant, autosomal dominant condition caused by mutations in the gene Indian hedgehog (IHH). A second locus for BDA1 has been reported at chromosome 5p13.3-13.2 (designated BDA1B; MIM 607004); however, a causative gene has yet to be described. Furthermore, an inversion of the Sonic hedgehog (Shh) locus has been shown to yield a murine BDA1 phenotype in heterozygous Short digits (Dsh/+) mice. A consanguineous French Canadian kindred with BDA1 was investigated subsequent to the exclusion of mutations in IHH and linkage to the BDA1B locus at chromosome 5p. Radiographs of the hands and feet were obtained for detailed examination. Sequencing the coding region of GDF5 revealed that a mildly affected individual in the family was heterozygous and severely affected individuals were homozygous for a novel missense mutation at a highly conserved amino acid in the GDF5 protein. Functional analysis demonstrated that while the mutant is able to stimulate chondrogenesis, it is much less effective than its wild-type counterpart. We then chose to investigate the functional candidate gene BMPR1B, a receptor of GDF5, for mutations in those BDA1 families which did not harbor mutations in IHH or GDF5 and subsequently identified a novel missense mutation in the BMPR1B gene. The functional ramifications of this mutation are currently being investigated. Neither mutation was seen in 400 control chromosomes. Neither of these genes has been previously implicated in type A-1 Brachydactyly. These data establishes that heterozygous (mild) or homozygous (severe) mutations in GDF5 are capable of causing BDA1, in addition to mutations in its receptor BMPR1B. Finally, we establish that mutations upstream of IHH signalling can result in brachydactyly A-1.

## 209

**Mutations in the Heparan-Sulfate Proteoglycan Glypican 6 (*GPC6*) Impair Endochondral Ossification and Cause Recessive Omodysplasia.** *R. Savarirayan[1], AB. Campos-Xavier[2], D. Martinet[3], J.F. Bateman[1], D. Belluoccio[1], L. Rowley[1], T. Tan[1], A. Baxova[4], KH. Gustavson[5], Z.U. Borochowitz[6], A.M. Innes[7], S. Unger[8, 9], J.S. Beckmann[3], L. Mittaz[2], D. Balhausen[2], A. Superti-Furga[9], L. Bonafé[2].* 1) Murdoch Childrens Research Institute, Parkville, Australia; 2) Division Molecular Genetics, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland; 3) Service Medical Genetics, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland; 4) Institute of Biology and Medical Genetics, Prague; 5) Dept. Clinical Genetics, Uppsala, Sweden; 6) Simon Winter Institute for Human Genetics, Haifa, Israel; 7) Dept Medical Genetics, Calgary, Canada; 8) Institute of Human Genetics, Freiburg, Germany; 9) Dept. Pediatrics, University of Freiburg, Germany.

Glypicans are a family of glycosylphosphatidylinositol (GPI)-anchored, membrane-bound heparan sulfate (HS) proteoglycans. Their biological roles are only partly understood, although it is assumed that they modulate the activity of HS-binding growth factors. The involvement of glypicans in developmental morphogenesis and growth regulation has been highlighted by Drosophila mutants and by a human overgrowth syndrome with multiple malformations caused by glypican 3 mutations (Simpson-Golabi-Behmel syndrome). We now report that autosomal recessive omodysplasia, a genetic condition characterized by short-limbed short stature, craniofacial dysmorphism,and variable developmental delay, maps to chromosome 13 (13q31.1-q32.2) and is caused by point mutations or by larger genomic rearrangements in glypican 6 (*GPC6*). All mutations cause truncation of the *GPC6* protein and abolish both the HS-binding site and the GPI-bearing membrane-associated domain, and thus loss of function is predicted. Expression studies in microdissected mouse growth plate revealed expression of *gpc6* in proliferative chondrocytes. Thus, *GPC6* seems to have a previously unsuspected role in endochondral ossification and skeletal growth, and its functional abrogation results in a short-limb phenotype.

## †210

**Biochemical, cellular, and tissue analysis of *Crtap-/-* mice and patients with recessive osteogenesis imperfecta.** *D. Baldridge[1], R. Morello[1], E. Homan[1], J. Lennington[1], T. Bertin[1], E. Munivez[1], M. Jiang[1,2], Y. Chen[1,2], D. Keene[3], D. Rimoin[4], D. Krakow[4], D. Cohn[4], S. Pyott[5], P. Byers[5], M. Weis[5], D. Eyre[5], B. Lee[1,2].* 1) Dept Molecular & Human Gen, Baylor Col Medicine, Houston, TX; 2) Howard Hughes Medical Institute, Houston, TX; 3) Shriners Hospital for Children, Portland, OR; 4) Cedars-Sinai Medical Center and David Geffen School of Medicine, UCLA, Los Angeles, CA; 5) University of Washington, Seattle, WA.

Null mutations in *CRTAP* (Cartilage-associated protein) or *LEPRE1* (coding for prolyl 3-hydroxylase 1 (P3H1)) cause recessive forms of osteogenesis imperfecta (OI) and loss of prolyl 3-hydroxylation. CRTAP, P3H1, and Cyclophilin B (a prolyl cis-trans isomerase) form a molecular complex in the endoplasmic reticulum that is responsible for proper collagen post-translational modification and prolyl 3-hydroxylation of specific proline residues. We completed a comprehensive analysis of the phenotype of the *Crtap-/-* mice (a mouse model of recessive OI) to better understand the role of CRTAP during development and in adulthood. Knock-out mice have multiple tissue abnormalities, including those observed in the lungs, kidneys, and skin, consistent with dysregulated production of extracellular matrix and disrupted matrix to cell signaling. Glomerulosclerosis with abnormal collagen deposition was also observed in some *Crtap-/-* glomeruli. Both the lung and kidney show an early onset (P10) increase in cellular proliferation by BrdU staining. To understand the potential collagen defects underlying these tissue abnormalities, we conducted a mass spectrometry survey of 3-Hyp occupancy in different types of collagen, and assessed the impact on these 3-Hyp sites in *Crtap-/-* mice. We identified previously unknown sites where a proline residue is hydroxylated to become a 3-Hyp in types I, II, and V collagens, and we observed that the expected 3-Hyp at one proline residue in $\alpha$2(V) collagen was absent in Crtap nulls. This is consistent with and may explain the lung, kidney, and skin abnormalities seen in *Crtap-/-* mice. We also analyzed the stability of CRTAP/P3H1/Cyclophilin B complex in primary human skin fibroblasts derived from patients with null mutations in *CRTAP* or *LEPRE1*. We demonstrate that *LEPRE1-/-* fibroblasts lose expression of the CRTAP protein and similarly, in *CRTAP-/-* fibroblasts, P3H1 expression is undetectable. These data suggest that both proteins are required to form a stable complex in the endoplasmic reticulum and that lack of either CRTAP or P3H1 causes degradation of the interacting partner. We propose that CRTAP may exert more widespread effects on collagen homeostasis including type V and perhaps type IV in kidney. These studies point to the requirement for prolyl 3-hydroxylation in the regulation of collagen structural and signaling functions and suggest a broader clinical assessment of patients with *CRTAP* or *LEPRE1* mutations.

† Trainee Award Finalist

## 211

**Mutations in *PPIB*, which encodes a prolyl cis-trans isomerase (cyclophilin B), in recessive forms of osteogenesis imperfecta (OI).** *U. Schwarze[1], S. Pyott[1], D. Russell[1], M. Sussman[2], C. Harris[3], P.H. Byers[1].* 1) Dept Pathology, University of Washington, Seattle, WA; 2) Shriners Hospital for Children, Portland, OR; 3) Medical Genetics, University of Illinois, Chicago, IL.

About 90% of all individuals with OI have heterozygous mutations in type I collagen genes. Recently recessive mutations in the *CRTAP* and *LEPRE1* genes were identified in individuals with phenotypes that ranged from lethal in the perinatal period to severe deforming OI. These genes encode proteins that interact in the RER with a third protein, cyclophilin B (CyPB), encoded by *PPIB*. This complex forms, brings CyPB to the nascent chains, and marks the delivery by the hydroxylation of a single prolyl residue at position 986 of the triple helix in the pro$\alpha$1(I) chains of type I procollagen. Loss of function mutations in either *CRTAP* or *LEPRE1* lead to loss of both proteins in the cell, leave the prolyl residue unmodified, and allow slow folding and excessive post-translational modification of the triple helical domain of the type I collagen molecules synthesized. These findings have suggested that delivery of CyPB to the chains is the pivotal function of the complex. We have now identified mutations in *PPIB* in cells from 3 individuals with OI. The first, with a severe form of OI, is homozygous for a 10bp deletion (c.414_423del) that results in a premature termination codon (PTC) that destabilizes the mRNA; the second, with lethal OI is a compound heterozygote for a single nucleotide deletion (c.120delC) that leads to a PTC and nonsense mediated mRNA decay, and a missense mutation (c.313G>A, p.Gly105Arg) that destabilizes the protein; the third, with an OI type IV phenotype, is homozygous for a splice donor mutation (c.343+1G>A, IVS3+1G>A) that leads to inclusion of 27nt of intron 3 in the stable mRNA and also to exon 3 skipping (unstable). In each case cultured dermal fibroblasts (patients 1 and 2) or marrow-derived mesenchymal stem cells (patient 3) make over-modified type I procollagen molecules which, in contrast to those with mutations in type I collagen genes but like those with mutations in the other two genes, have normal thermal stability. Type I procollagen accumulates in the RER. These findings indicate that increased modification is a common effect of dominant and recessive mutations, suggest that substitutions in the triple helical domains could interfere with prolyl *cis-trans* isomerization, and so, suggest that this model provides a common pathogenetic pathway in these disorders. Even with this new gene in hand, it is clear that others remain to be identified.

## 212

**Mutations in a gene encoding a rough endoplasmic reticulum protein causes autosomal recessive progressive deforming osteogenesis imperfecta.** *Y. Alanay[1], H. Avaygan[2], N. Camacho[2], G.E. Utine[1], K. Boduroglu[1], D. Aktas[1], M. Alikasifoglu[1], E. Tuncbilek[1], D. Orhan[1], A. Alanay[1], B. Zabel[3], A. Superti-Furga[3], L. Bruckner-Tuderman[3], P. Byers[4], D. Eyre[4], A.E. Merrill[2,5], D.H. Cohn[2,5], N. Akarsu[1], D. Krakow[2,5].* 1) Hacettepe University, Ankara, Turkey; 2) University of California, Los Angeles, CA, USA; 3) University of Freiburg, Freiburg, Germany; 4) University of Washington, Seattle, Washington, USA; 5) Cedars-Sinai Medical Center, Los Angeles, CA, USA.

Nearly 90% of individuals with OI have dominantly inherited mutations in a type I collagen gene (*COL1A1* or *COL1A2*) causing defects in the structure or synthesis of type I collagen. Type I collagen is subject to post-translational modification in the rough endoplasmic reticulum (RER) during chain synthesis and helix formation. The “collagen prolyl 3-hydroxylation complex” forms with molecular assembly of cartilage associated protein (CRTAP) with protein prolyl 3-hydroxylase-1 (P3H1) and cyclophilin B (CYPB) or peptidylprolyl isomerase B (PPIB) in a 1:1:1 ratio. Recessively inherited OI has been associated with mutations in *CRTAP* and *LEPRE1*,gene encoding P3H1. Five families with a progressive deforming type of OI were ascertained and studied. All families are from a small region on the Black Sea coast of Turkey supporting a common genetic origin for the disease. The skeletal phenotype resembled moderate to severe OI. Sclerae were white or gray, dentinogenesis imperfecta was absent and hearing was normal.Under polarized light, bone histology resembled the characteristic bone lamellation named “fish-scale pattern” previously documented in individuals affected with OI type VI. Autosomal recessive epidermolysis bullosa simplex, due to homozygosity for a Y203X mutation in the keratin 14 gene, cosegregated with OI in all affected individuals.Based on the known location of *KRT14* on chromosome 17, homozygosity by descent mapping identified a 2.4 megabase region on chromosome 17q21-22 shared by all of the affected individuals.Homozygosity for a 33 base pair deletion, c.319-355del, in the gene that encodes the immunophilin molecule FKBP65, was identified in all affected individuals. This mutation predicts an in-frame deletion of 11 amino acids, p.C110_P120del. Subsequently a frameshift mutation was identified in an unrelated family with recurrent progressive deforming OI, c.831_832insC, leading to p.Gly278ArgfsX295.Chaperones are proteins that can recognize non-native proteins, prevent unwanted inter- and intramolecular protein-protein interactions and influence protein folding steps. The rough endoplasmic reticulum-resident FK506-binding protein FKBP65 has a peptidylpropyl cis-trans-isomerase (PPIase) activity and has recently been documented to be a chaperone for type I collagen. However, the exact role of FKBP65 in the posttranslational modification process, and how mutations in this gene produce severe progressive deforming OI, remains unknown.

## †213

**Short rib polydactyly syndrome is due to defects in the ciliary transport protein DYNC2H1.** *A.E. Merrill[1,2], B. Merriman[3], C. Farrington-Rock[2], N. Camacho[1], E.T. Sebald[2], V.A. Funari[2], M.J. Schibler[7,8], M. Priore[2], A.K. Thompson[9], D.L. Rimoin[2,3,4,5], S.F. Nelson[3], D.H. Cohn[2,3,4], D. Krakow[1,2,3,6].* 1) Orthopaedic Surgery, David Geffen School of Medicine at UCLA, Los Angeles, CA; 2) Medical Genetics Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA; 3) Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA; 4) Pediatrics, David Geffen School of Medicine at UCLA, Los Angeles, CA; 5) Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA; 6) Obstetrics and Gynecology, David Geffen School of Medicine at UCLA, Los Angeles, CA; 7) Brain Research Institute, David Geffen School of Medicine at UCLA, Los Angeles, CA; 8) California NanoSystems Institute, UCLA, Los Angeles, CA; 9) Center for Electron Microscopy, University of Southern California, Los Angeles, CA.

The short-rib polydactyly (SRP) syndromes, a heterogenous group of skeletal disorders, are among the most frequent autosomal recessive lethal osteochondrodysplasias. SRPs manifest extremely shortened long bones, a small, narrow thorax, frequent pre- and postaxial polydactyly, and multisystem anomalies including polycystic kidneys. SRPs share phenotypic and radiographic findings with asphyxiating thoracic dystrophy and Ellis-van Creveld, supporting the hypothesis that these diseases comprise a family of ciliary disorders that may be functionally related. Homozygosity by descent mapping in a consanguineous SRP family identified a genomic region containing DYNC2H1, a cytoplasmic dynein involved in retrograde transport in the cilium. Affected individuals in the family were homozygous for an exon 12 missense mutation that predicted the amino acid substitution R587C. Compound heterozygosity for one missense and one null mutation was identified in two additional nonconsanguineous SRP families. Cilia are highly conserved microtubule-based organelles that play diverse roles in cellular motility, sensory transduction, and signaling. Cilia consist of an axoneme made of nine peripheral microtubule doublets arranged around a central core. Elongation and maintenance of the cilia depends on the dynamic process of intraflagellular transport which moves ciliary cargo with the help of motors along the outer microtubules of the axoneme. Anterograde movement of axonemal precursors from the basal body to the distal tip is driven by the heterotrimeric kinesin-II motor. Retrograde transport from the distal tip to the basal body is accomplished by the cytoplasmic dynein 2 complex, which includes DYNC2H1. We found that cilia displayed by SRP chondrocytes have morphological abnormalities characteristic of retrograde transport dysfunction. Scanning electron microscopy of cultured chondrocytes from affected individuals showed shortened cilia with a bulbous distal tip. Immunofluorescent visualization of tubulins, the microtubule building blocks, confirmed these structural defects and showed that cytoplasmic microtubules were abnormally fragmented and heavily acetylated. Increased levels of acetylated $\alpha$-tubulin, confirmed by western blot, implicates altered microtubule network as part of the disease process. Together these findings establish SRP as a cilia disorder and demonstrate the essential role of DYNC2H1 in maintenance of microtubule architecture.

## 214

**Mutations in *IFT139* contribute both causal and modifying alleles across the ciliopathy spectrum.** *E.E. Davis[1,2], Q. Zhang[3], Q. Liu[3], J. Hartley[4], D.M. Muzny[5], A.C. Young[6], P. Cruz[6], J.C. Mullikin[6], P.J. Scambler[7], P.L. Beales[7], J.G. Gleeson[8], F. Hildebrandt[9], E.R. Maher[4], T. Attie-Bitach[10], H. Dollfus[11], C.A. Johnson[12], E.D. Green[6], R.A. Gibbs[5], E.A. Pierce[3], N. Katsanis[1,2,13].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA; 2) Center for Human Disease Modeling, Department of Cell Biology, Duke University Medical Center, Durham, North Carolina, 27710 USA; 3) F.M. Kirby Center for Molecular Ophthalmology, University of Pennsylvania School of Medicine, Philadelphia, PA, 19104; 4) Department of Medical and Molecular Genetics, Institute of Biomedical Research, University of Birmingham, Birmingham, United Kingdom; 5) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA; 6) Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA; 7) Molecular Medicine Unit, Institute of Child Health, University College London, London WC1N 1EH, UK; 8) Department of Neurosciences, Howard Hughes Medical Institute, University of California, San Diego, La Jolla, CA 92093, USA; 9) Departments of Human Genetics and Pediatrics, Howard Hughes Medical Institute, University of Michigan, Ann Arbor, Michigan 48105, USA; 10) Département de Génétique et INSERM U-781, Hôpital Necker-Enfants Malades, Paris Cedex 15, France; 11) Laboratoire de Génétique Médicale EA 3949, Faculté de Médecine de Strasbourg, Université Louis Pasteur, 67085 Strasbourg, France; 12) Section of Ophthalmology and Neurosciences, Leeds Institute of Molecular Medicine, St. James's University Hospital, Leeds, United Kingdom; 13) Wilmer Eye Institute and Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore Maryland 21205, USA.

Perturbation of cilia leads to a broad range of overlapping phenotypes in humans, termed collectively as ciliopathies. This grouping is also underscored by genetic overlap, since some ciliopathy genes can also contribute modifying alleles to related, but clinically distinct disorders. Here we show that mutations in *IFT139*, a component of the retrograde intraflagellar transport (IFT) complex, and the causal locus of the *alien* mouse mutant, cause ciliopathies that range from isolated nephronophthisis (NPHP), to neonatal lethal Jeune Asphyxiating Thoracic Dystrophy (JATD), and Meckel-Gruber Syndrome (MKS). Moreover, systematic medical resequencing of a large, clinically diverse ciliopathy cohort and matched controls showed a significant enrichment of rare nonsynonymous alleles in patients. In vitro and in vivo evaluations showed most variants to be pathogenic and suggested that *IFT139* contributes null and hypomorphic alleles to as much as 5% of ciliopathy patients, with a marked excess of pathogenic alleles in severe phenotypes. Our data illustrate how genetic lesions in a ciliary transcript can be causally associated with the breadth of ciliopathies in humans and highlight how saturated resequencing and functional analysis of all variants can inform the genetic architecture of related clinical phenotypes.

† Trainee Award Finalist

## 215

**OFD1 is mutated in X-linked Joubert syndrome and interacts with *LCA5*-encoded lebercilin.** *A.P.P de Brouwer[1], K.L.M. Coene[1], D. Doherty[2], B. Afroze[3], H.Y. Kroes[4], S.J.F. Letteboer[1], H.L. Ngu[3], B. Budny[5], E. van Wijk[1], M. Azhimi[1], C. Thauvin-Robinet[6], J.A. Veltman[1], M. Boink[1], T. Kleefstra[1], F.P.M. Cremers[1], H. van Bokhoven[1], R. Roepman[1].* 1) Dept of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 2) Division of Genetics and Developmental Medicine, University of Washington Children's Hospital and Regional Medical Center, Seattle, USA; 3) Division of Clinical Genetics, Pediatrics Institute, Kuala Lumpur Hospital, Kuala Lumpur, Malaysia; 4) Dept of Medical Genetics, University Medical Centre Utrecht, Utrecht, The Netherlands; 5) Dept of Medical Genetics, Poznan University of Medical Sciences, Poznan, Poland; 6) Centre de Génétique, Hôpital d'Enfants, Dijon, France.

We ascertained a multi-generation family with Joubert syndrome (JBTS) from Malaysia. The presence of a-symptomatic obligate carrier females suggests an X-linked recessive inheritance pattern. Affected males presented with mental retardation accompanied by postaxial polydactyly and retinitis pigmentosa. Brain MRI scans showed the presence of a so-called 'molar tooth sign', which classifies this syndrome as classic JBTS with retinal involvement. Linkage analysis showed linkage to Xpter-Xp22.2, with a maximum LOD score of 2.06 for marker DXS8022. The linkage interval is delimited by the telomere of the p-arm and marker DXS8036. The remainder of the X chromosome could be excluded (LOD score <-2). The 14 Mb interval contains 66 annotated genes (NCBI Map Viewer build 36.3), four of which encode ciliary proteins: *NLGN4X*, *HCCS*, *OFD1*, and *PRPS2* (www.ciliaproteome.org). Mutation analysis revealed a frameshift mutation, p.K948fsX8, in exon 21 of *OFD1*. In an isolated male with JBTS, a second frameshift mutation, p.E923fsX3, in the same exon was identified. These data show that we have found the first JBTS gene on the X chromosome and the eight JBTS gene in total. OFD1 has previously been associated with oral-facial-digital type 1 (OFD1) syndrome, which is a male-lethal X-linked dominant condition, and with X-linked recessive Simpson-Golabi-Behmel syndrome type 2 (SGBS2). By a yeast two-hybrid screen of a retinal cDNA library, OFD1 was identified as an interacting partner of the *LCA5*-encoded ciliary protein lebercilin. Mutations in *LCA5* cause Leber congenital amaurosis, an inherited condition of very early onset childhood blindness due to retinal degeneration that also belongs to the ciliopathy spectrum. Yeast growth and liquid β-galactosidase assays showed that X-linked recessive mutations in OFD1 reduce the interaction with lebercilin to a limited extent, whereas X-linked dominant OFD1 mutations abolish binding with lebercilin. Ciliary localization was concomitantly affected. Recessive mutations in OFD1 did not affect the localization of the recombinant protein to the basal bodies of cilia in hTERT-RPE1 cells, whereas this localization was lost completely when OFD1 carried a dominant mutation. These findings offer a molecular explanation for the phenotypic spectrum observed for *OFD1* mutations that now includes OFD1 syndrome, SGBS2, and JBTS.

## 216

**Exon capture and large-scale sequencing of 828 ciliopathy candidate genes in patients with nephronophthisis, Senior Loken-, and Joubert syndrome.** *E. Otto[1], J. MacDonald[1], H. Harville[1], M. Chaki[1], W. Zhou[1], F. Hildebrandt[1,2,3].* 1) Dept Pediatrics, Univ Michigan, Ann Arbor, MI; 2) Human Genetics, Univ Michigan, Ann Arbor, MI; 3) Howard Hughes Medical Institute.

The number of ciliopathies with a mutation in a gene expressed at the cilia-centrosome complex is steadily increasing. Defects in cilia have been implicated mostly in recessive disorders like nephronophthisis (renal cysts, kidney failure), Senior Loken syndrome (kidney disease, retinitis pigmentosa), Joubert syndrome (cerebellar vermis hypoplasia, mental retardation) and many others. To identify disease causing mutations, we custom designed NimbleGen sequence capture hybridization arrays (385K) to capture all 13,221 exons of 828 selected ciliopathy candidate genes in 5 patients. Candidate genes were chosen based on published ciliary or centrosomal protein localization, animal models, and our own Y2H protein-protein interaction data. DNA capture was performed through the NimbleGen capture service. Subsequently, captured DNA fragment size was reduced by exo- and endonuclease treatment to be suitable for high throughput sequencing on a Solexa/Illumina platform. Bioinformatic evaluation was performed with the CLC Genomics Workbench software. On average we obtained about 9 million reads (36 bases) per lane with 33% of reads on target. The median sequencing depth was 49 with 99% of the targeted exons covered at least 5-fold. Sequencing revealed a homozygous missense mutation (c.770G>A; p.R257H) in the cAMP dependent protein kinase catalytic subunit alpha (*PRKACA*) gene in a patient with nephronophthisis. The patient developed end-stage renal disease at age 7 years. In a whole-genome search for linkage in this patient from a consanguineous kindred from Turkey (A395), linkage to known *NPHP* loci had been excluded. The mutation identified in *PRKACA*, is absent from 95 Caucasian and 85 ethnical matched control individuals from Turkey. The amino-acid residue arginine at position 257 is evolutionarily conserved in mammals, birds, frogs, fish, *Ciona*, and *Drosophila*. Mutation analysis in a cohort of 102 additional ciliopathy patients was negative. PRKACA is involved in cell cycle regulation. Interestingly, by 2 different proteomic approaches PRKACA has been identified in mouse photoreceptor sensory cilia and in human centrosomes. Furthermore, knockout mice show growth retardation and defective sperm motility indicating a ciliary phenotype (Skalhegg, Mol Endocrinol 16:630, 2002). This approach of using a "ciliopathy candidate exon capture array" with consecutive large-scale sequencing may strongly facilitate the identification of new causative genes for ciliopathies.

## 217

**Identification of multiple novel genetic variants which predispose to prostate cancer by genome-wide association in the PRACTICAL consortium.** *Z. Kote-Jarai[1], A. Amin Al Olama[2], G. Giles[3], M. Guy[1], G. Severi[3], J. Hopper[3], F. Hamdy[4], D. Neal[5], J. Donovan[6], K. Muir[7], J. Morrison[2], D. Easton[2], R. Eeles[1, 8], The PRACTICAL Consortium.* 1) Translational Cancer Genetics, The Institute of Cancer Research, Sutton, Surrey, United Kingdom; 2) CRUK Genetic Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom; 3) Centre for Mol, Env, Genetic and Analytic Epidemiology, The University of Melbourne, Melbourne, Australia; 4) Nuffield Department of Surgery, University of Oxford, Oxford , United Kingdom; 5) Surgical Oncology, University of Cambridge, Cambridge, United Kingdom; 6) Department of Social Medicine, University of Bristol, Bristol, United Kingdom; 7) University of Nottingham Medical School, Queens Medical Centre, Nottingham, United Kingdom; 8) The Royal Marsden NHS Foundation Trust, London, United Kingdom.

Prostate cancer (PrCa) is the commonest solid tumour in men in developed countries. There is epidemiological evidence of genetic predisposition to the disease. To identify common PrCa susceptibilility alleles, we conducted a genome-wide association study in 1,854 PrCa cases with clinically detected PrCa diagnosed at <60 years or a family history of disease, and 1,894 population screened controls from the UK. 541,129 SNPs were genotyped using the Illumina Infinium platform and putative associations were typed in a further 3,245 cases and 3,329 controls from the UK and Australia. We identified 7 novel prostate cancer susceptibility loci (P=2.2x10-8 to P= 8.1x10-29). We also confirmed previous reports of common prostate cancer susceptibility loci at 8q24 (P=2.8x10-17) and TCF2 at 17q (P=10-12). Three of the novel loci contain strong candidate susceptibility genes, MSMB, LMTK2 and KLK3. Follow up genotyping of 47 000 SNPs from 4000 PrCa cases and 4000 controls from the UK and Australia identified 116 SNPs at p<10-6 of which 12 had P values of 10-6 to 10-12 in ten regions not previously reported. These 12 SNPs were genotyped by an international consortium, PRACTICAL in over 30 000 DNA samples from diverse populations in 21 groups. This revealed that 9 SNPs at 7 new loci were at genome-wide significance, all of which contain plausible candidate genes, some of which are in pathways amenable to therapeutic intervention. There are now 25 genetic variants which predispose to PrCa, of which 16 have been found by this experiment. This will aid SNP risk assessment profiling for targeted PrCa screening and provides hits for therapeutic targets. Acknowledgements This study was funded by Cancer Research UK, The Prostate Cancer Research Foundation, and the NIHR to the Biomedical Research Centre.

## 218

**Fine-mapping of 8q24 based on resequence analysis and tag SNPs provides new insights into prostate cancer risk.** *M. Yeager[1,2], J. Gonzalez Bosquet[2], K. Jacobs[1,2,3], N. Chatterjee[2], R.B. Hayes[2], L.A. Burdett[1,2], P. Kraft[4], S. Wacholder[2], S. Berndt[2], K. Yu[2], A.A. Hutchinson[1,2], Z. Wang[1,2], Z. Deng[1,2], M. Tucker[2], R.N. Hoover[2], J.F. Fraumeni, Jr.[2], D.J. Hunter[4,5], G. Thomas[2], S.J. Chanock[2].* 1) Core Genotyping Facility, Advanced Technology Program, SAIC-Frederick, Inc., NCI-Frederick, Frederick, MD 21702; 2) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Heath, Department of Health and Human Services, Bethesda, Maryland 20892, USA; 3) Bioinformed Consulting Services, Gaithersburg, Maryland 20877, USA; 4) Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA; 5) 5Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.

Recent genome-wide association studies (GWAS) have identified that a ~600kb segment of chromosome 8q24 contains multiple independent variants that are associated with the risk of prostate, colon, breast, bladder, and other cancers. To date, three independent regions of 8q24 ("regions 1, 2, and 3") have been implicated in prostate cancer. As a part of the National Cancer Institute's (NCI) Cancer Genetics Markers of Susceptibility (CGEMS) project, in order to thoroughly investigate 8q24 in prostate cancer, we have taken a multi-tiered approach that includes dense resequencing for polymorphism discovery which identified SNPs for regional fine mapping, and genotyping in ~10,000 prostate cancer cases and ~10,000 controls to refine the association signals. This presentation outlines the CGEMS 8q24 prostate cancer follow-up strategy, initial findings including the discovery of a fourth prostate cancer-association region ("region 4") previously implicated in breast cancer risk, refinement of the association signal in region 2, and the nomination of variants for functional studies from regions 1 and 3. Further, when considering all independent loci for potential non-multiplicative effect, we observe departures from a multiplicative odds ratios model between SNPs in regions 1 and 4. Our findings underscore the importance of exploring regions of association with dense resequencing and genotyping in follow-up studies. Funded by NCI Contract No. HHSN261200800001E.

## 219

**A Functionally Polymorphic, Long-Range Enhancer of MYC in the 8q24 Region Associated with Prostate Cancer.** *C. Carlson, O. Kahsai, A. McDavid, J. Potter.* Public Health Sci, Fred Hutchinson Cancer Res Ctr, Seattle, WA.

Multiple recent genome wide association studies have identified polymorphism in the chromosome 8q24 region as associated with a variety of solid tumor cancers, including prostate, colon, breast, bladder and lung. The specific polymorphisms associated with each cancer show some overlap, but the data support the presence of multiple, independent cancer risk alleles in this region, a gene desert 5' of the MYC oncogene. We used bioinformatic approaches to predict an enhancer element in a region showing an association with prostate cancer, on the basis of evolutionary conservation and transcription factor motif analyses. Within this region, the rs16901966 polymorphism is in perfect LD with the prostate cancer associated rs6983561 in the HapMap CEU samples. Using the core MYC promoter and allelic versions of the predicted enhancer, transient transfection assays in prostate cancer cell lines showed significantly elevated MYC expression from the cancer risk allele. Gel shift assays confirm allele specific transcription factor affinity in PC3 prostate cell line nuclear extract. Our results suggest that at least one of the prostate cancer associations in the 8q24 region is attributable to long-range regulation of MYC by a functionally polymorphic enhancer element.

## 220

**Genetic Susceptibility to Type 2 Diabetes is Associated with Prostate Cancer Risk.** *B.L. Pierce, H. Ahsan.* Health Studies, Univ Chicago, Chicago, IL.

INTRODUCTION: Epidemiologic research suggests that diabetes mellitus (DM) is associated with reduced prostate cancer (PCa) risk. The purpose of this analysis was to examine the collective effects of recently-identified type 1 (T1D) and type 2 diabetes (T2D) risk alleles on PCa. METHODS: Using data on 1,171 non-Hispanic white, PSA-screened, PCa cases and 1,101 matched controls (from the Cancer Genetic Markers of Susceptibility study (CGEMS)), we generated genetic risk scores (a "risk allele count" and a "genetic relative risk") for both T1D and T2D based on 14 and 18 single nucleotide polymorphisms (SNPs) known to associate with T1D and T2D, respectively. We then employed logistic regression models (adjusted for 5-year age groups and genetic ancestry derived from principal components analysis) to estimate odds ratios (OR) and 95% confidence intervals (CI) for associations between these risk scores and PCa risk in the CGEMS dataset. RESULTS: Both T2D risk scores, but neither T1D score, showed an inverse association with PCa (p<0.01). These associations remained significant after excluding HNF1B SNP rs4430796 (a known genetic risk factor for PCa) from the analysis. The highest quartile of T2D allele count (>20 risk alleles) was associated with reduced PCa risk (OR=0.77; CI: 0.60-0.99) compared to the lowest category (<17 risk alleles). For individual T2D SNPs, 13 of the 18 risk alleles had ORs <1.00 for PCa risk. CONCLUSION: These results suggest that individuals with increased genetic susceptibility to T2D have decreased risk for PCa. This association is consistent with the hypothesis that individuals with T2D are at decreased risk for prostate cancer. Future studies of these DM variants and PCa risk should incorporate DM phenotype data.

## 221

**The Nicotinic Acetylcholine Receptor region on chromosome 15q25 associates with adenocarcinoma of the lung in African-Americans.** *C.I. Amos, I.P. Gorlov, Q. Dong, X. Wu, H. Zhang, E. Lu, P. Scheet, M.R. Spitz.* Dept Epidemiology, MD Anderson Cancer Ctr, Houston, TX.

Lung cancer is the most common cause of cancer mortality in the U.S. and is 30% more frequent in African-American men than in Caucasian men. Last year, several groups including ours identified a region of chromosome 15q24-25.1 that had a highly reliable and significant association with lung cancer risk for all histologies. Unfortunately, this region, which encompasses several nicotinic acetylcholine receptor subunit loci (CHRNA3, CHRNA5, CHRNB4)as well as a proteasome subunit (PSMA4) contains many SNPs that are in strong linkage disequilibrium so that identifying the specific causal factor(s) has been difficult. The linkage disquilibrium pattern in African-Americans is more punctate in this region facilitating gene localization efforts. Toward that aim, we performed dense genotyping of 34 SNPs across the associated region of chromosome 15q on 467 African-American cases and 388 African-American controls. SNPs across the region only showed significant effects among former or current smokers. The single most significant SNP was rs10519203 (p=0.0002, OR=1.60) near PSMA4, but several SNPs, including rs16969968, a functional SNP in CHRNA5 (p=0.003, OR=1.98) and rs1051730 near CHRNA3 (p=0.0004, OR=1.85) were also highly significantly associated with lung cancer risk, even though these markers and several others in the region have low joint LD levels. Analysis by histology shows significant effects only in patients presenting with adenocarcinomas (p=0.0008, OR=2.30 for rs1051730), and a much less significant effect in patients with squamous cancers (p=0.10, OR=1.60for rs1051730). These results suggest effects from several loci in the associated region on chromosome 15q25.1 and suggest that unlike Caucasians, there are histology-specific effects of variants in this region on lung cancer risk in African-Americans. Results underscore the value of studying minority populations that show different LD patterns from Caucasians and that may have different sets of risk alleles, affecting the disease phenotype.

## 222

**Nevus and pigmentation loci are associated with melanoma in genome-wide association study of European and Australian populations.** *D. Bishop[1], F. Demenais[2], M.M. Iles[1], M. Harland[1], J.C. Taylor[1], E. Corda[2], B. Bakker[3], P.A. Kanetsky[4], K.M. Brown[5], A.M. Goldstein[6], E.M. Gillanders[7], D.E. Elder[8], N.A. Gruis[9], N.K. Hayward[10], J.H. Barrett[1], J.A. Newton Bishop[1], The GenoMEL Consortium.* 1) Gen Epidemology Lab, St James Univ Hosp, Leeds, United Kingdom; 2) Fondation Jean Dausset-CEPH, 75010, Paris, France; 3) Department of Clinical Genetics, Center of Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands; 4) Centre for Clinical Epidemiology & Biostatistics and Department of Biostatistics & Epidemiology, 219 Blockley Hall, University of Pennsylvania, USA; 5) Melanoma Genomics Laboratory, The Translational Genomics Research Institute (TGen), 445 N. Fifth Street, Phoenix, AZ 85004, USA; 6) Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute. NIH, Bethesda, MD 20892-7236, USA; 7) Inherited Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, Maryland 21224, USA; 8) Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA; 9) Department of Dermatology, Leiden University Medical Centre, Leiden, The Netherlands; 10) Queensland Institute of Medical Research. 300 Herston Rd, Herston, QLD 4029, Australia.

On behalf of the GenoMEL melanoma genetics consortium we conducted a genome-wide association study of melanoma based on 317k SNPs in 1650 genetically-enriched cases and 4336 controls from centres across Europe and Australia. Three loci were identified and replicated using two further data sets (one of 1149 genetically-enriched cases and 964 controls and one of 1163 population-based cases and 903 controls): 16q24 encompassing *MC1R* (lead SNP rs258322, $p=2.5\times10^{-27}$), the *TYR* pigmentation gene on 11q14-q21 ($p=2.4\times10^{-14}$ for rs1393350), and 9p21 adjacent to *MTAP* and the familial melanoma gene *CDKN2A* ($p=4.0\times10^{-7}$ for rs7023329), a locus also identified in an independent genome-wide association study of nevus count in UK twins. Multiple logistic regression analysis showed an independent effect of 6 SNPs at these loci on melanoma risk and no evidence of departure from a joint multiplicative effect. Despite variation in allele frequencies, and large differences in sun exposure, effect sizes were homogeneous across centres. We also replicated two further melanoma loci, not reaching genome-wide significance here, but identified in independent studies: a region on chromosome 20 close to *ASIP* ($p=1.3\times10^{-8}$ for rs1885120) and one on chromosome 22 previously associated with nevus count ($p=2.4\times10^{-9}$ for rs2284063). We investigated these three pigmentation loci (*MC1R*, *TYR* and *ASIP*) and two nevus-related loci (*MTAP*/*CDKN2A* and chromosome 22) in a UK population-based extensively-phenotyped melanoma case-control study and observed that the effect of the SNPs on melanoma risk is partially, but not entirely, attenuated by the related phenotype. Further work is underway to impute genotypes genome-wide to help identify further signals for follow-up, to fine-map these loci and to further characterise their effect on melanoma risk taking into account phenotype and environmental exposures.

## 223

**Common breast cancer susceptibility alleles and the risk of breast cancer in *BRCA1* and *BRCA2* mutation carriers: implications for risk prediction.** *A.C. Antoniou[1], G. Chenevix-Trench[2], D.F. Easton[1], Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA).* 1) Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom; 2) Queensland Institute of Medical Research, Brisbane, Australia.

Genome-wide association studies have identified multiple SNPs that are associated with increased breast cancer risks in the general population. In a previous study we demonstrated that the minor alleles of three of these SNPs, in *FGFR2*, *TNRC9* and *MAP3K1*, also confer increased risks of breast cancer for *BRCA1* or *BRCA2* mutation carriers. However, the effects of SNPs at other susceptibility loci on breast cancer risk for *BRCA1* and *BRCA2* carriers are unknown. We evaluated the association of three further SNPs rs13281615 in 8q24, rs3817198 in *LSP1* and rs13387042 in 2q35, with breast cancer risk in 9,442 *BRCA1* and 5,665 *BRCA2* mutation carriers from 33 studies. The minor allele of rs3817198 was associated with increased breast cancer risk in *BRCA2* mutation carriers (Hazard Ratio (HR)=1.16, 95%CI:1.07-1.25, p-trend=$3\times10^{-4}$), but not *BRCA1* carriers. The best fit for the association of SNP rs13387042 in 2q35 with breast cancer risk was a dominant model for both *BRCA1* and *BRCA2* mutation carriers (*BRCA1*: HR=1.14, 95% CI:1.04-1.25, p=.0047; *BRCA2*: HR=1.18 95%CI:1.04-1.33, p=.008). SNP rs13281615 in 8q24 was not associated with breast cancer in either *BRCA1* or *BRCA2* mutation carriers, but the HR in *BRCA2* carriers was consistent with that in the general population (HR=1.06,95%CI:0.98-1.14). These results are consistent with a general pattern whereby breast cancer susceptibility loci confer a similar relative risk in *BRCA2* carriers and non-carriers, but only a limited subset confer a risk in *BRCA1* carriers. These polymorphisms appear to interact multiplicatively on breast cancer risk for *BRCA2* mutation carriers. Because *BRCA1* and *BRCA2* mutations confer high risks of developing breast cancer, the modest hazard ratios associated with these SNPs translate into marked differences in absolute risks between genotypes at these SNPs. Based on the joint genotype distribution of the *FGFR2*, *TNRC9*, *MAP3K1*, *LSP1* and 2q35 SNPs, and assuming they interact multiplicatively on breast cancer risk, the 5% of *BRCA2* mutation carriers at lowest risk are predicted to have a probability of 45-49% of developing the disease by age 80. In comparison, the corresponding probability for the 5% of *BRCA2* mutation carriers at highest risk is 82-94%. Such risk differences may be sufficient to influence the clinical management of mutation carriers and suggest that this may be one of the first clinically useful impacts of common, low penetrance variants identified through genome wide association studies.

## 224

**Dissecting the genetic components of gene expression in breast carcinoma.** *S. Nordgard[1], W. Sun[2], P. Van Loo[3,4], B. Naume[5], O.C. Lingjærde[6], A-L. Børresen-Dale[1,7], V.N. Kristensen[1,8].* 1) Department of Genetics, Institute for Cancer Research, Oslo, Oslo, Norway; 2) Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA; 3) Department of Molecular and Developmental Genetics, Vlaams Instituut voor Biotechnologie, Leuven, Belgium; 4) Department of Human Genetics, Katholieke Universiteit Leuven, Leuven, Belgium; 5) Oncology, Oslo University Hospital-Radiumhospitalet, Oslo, Norway; 6) Department of Informatics, University of Oslo, Oslo, Norway; 7) Faculty Division The Norwegian Radiumhospital, University of Oslo, Oslo, Norway; 8) Institute for Clin Epidemiology and Molecular Biology, University of Oslo, Oslo, Norway.

A series of publications has demonstrated the effects of genetic variation on mRNA expression. Given the significant role mRNA expression patterns play in breast cancer, we examined to what extent germline and somatic variation may influence expression levels in breast carcinomas. *Cis* and *trans* eQTL analyses were performed using a linear regression model, and corrected for multiple testing by Bonferroni. eQTL analysis of germline SNPs found 86 significant hits in 45 genes (70 SNPs). The top two findings were *cis* associations for the *LRAP* gene (Pval=$2x10^{-30}$). Defects in the expression of this aminopeptidase are known to cause improper antigen processing, which may possibly enable the tumor to escape the immune surveillance. Utilizing the LogR and BAF we elucidated the copy number for each allele (A and B). eQTL analysis on total copy number (A+B) returned 573 hits (318 probes), with the most significant result for a probe in the alkaline phytoceramidase gene *PHCA* (Pval=$4.2x10^{-33}$), a regulator of cell proliferation and survival. Probes within the *ERBB2* gene were also found significant (Pval=$8.5x10^{-26}$, *cis*). We identified the functional categories of the genes harboring these significant probes by using the Gene Ontology(GO) database, and found that significantly enriched GO categories include amongst others hormone biosynthesis. When studying the allele specific influence on expression in the tumor (A-B) the most significant finding was seen for a SNP in *PHCA* (Pval=$2x10^{-29}$). We found 86 hits in 33 genes (70 SNPs). There was a 37% overlap between the tumor SNPs influencing expression in an allele specific manner and the germline SNPs significantly associated with expression. Interestingly, none of the findings from the eQTL analysis for A+B overlapped with the eQTL analysis of the germline SNPs. We have already reported that the germline genetic background of breast cancer patients may either predispose to, or be co-selected with the genomic aberrations seen in their tumors. Our preliminary analysis implies the existence of skewness in breast tumors with respect to which allele is amplified or deleted, and that their association to variation in expression level may be the driving force behind this selection. These results imply that the germline genetic background may play a significant role in the expression pattern observed in the tumor, as well as total copy number and allele specific aberrations of the tumor.

## 225

**Neoadjuvant Therapy with Cisplatin in BRCA1-Positive Breast Cancer Patients.** *J. Lubinski[1], T. Byrski[1], T. Huzarski[1], R. Dent[2], J. Gronwald[1], D. Zuziak[3], R. Wisniowski[3], C. Cybulski[1], B. Gorski[1], S.A. Narod[4].* 1) International Hereditary Cancer Center, Pomeranian Medical University, Szczecin, Poland; 2) Sunnybrook Odette Cancer Center, Toronto, Ontario, Canada; 3) Regional Oncology Center, Bielsko-Biala, Poland; 4) Women's College Research Institute, Toronto, Ontario, Canada.

**Background** Neoadjuvant chemotherapy is administered to control disease, make surgical resection possible and increase the possibility of breast tissue conservation. A further advantage of neoadjuvant therapy is that it helps to assess chemo-sensitivity to a particular agent. Induction of a pathological complete response (pCR) is one of the primary goals of neoadjuvant therapy in order to achieve a better disease-free and overall survival. Experimental data suggest that BRCA1 related breast cancer may have increased sensitivity to platinum-based chemotherapy, but clinical data are limited. The aim of this study was to evaluate the frequency of complete pathologic response after neo-adjuvant treatment with cisplatin chemotherapy in women with breast cancer and a BRCA1 mutation. **Methods** Twenty five women with breast cancer and a BRCA1 mutation with stage I, II and III breast cancer between December 2006 and December 2008 were entered into this study. Patients were treated with cisplatin 75 mg/m2 intravenously every three weeks for four cycles. After chemotherapy, patients underwent surgery and were assessed for pathologic response in both the breast and axillary lymph nodes. Complete pathologic response was defined as no residual invasive disease in both the breast and axilla, however ductal carcinoma in situ was allowed. **Results** Twenty five patients were enrolled in the study. Thirteen patients had tumors of greater than two centimeters (52%) and seven patients had positive lymph nodes at diagnosis (28%). Twenty two patients completed four cycles of cisplatin (88%) and three patients completed two cycles (12%). Clinical complete response was observed in eighteen patients (72%). Pathologic complete response was observed in eighteen patients (72%). **Conclusions** Platinum-based chemotherapy is effective in a high proportion of patients with BRCA1-associated breast cancers. Clinical trials are warranted to determine the optimum treatment for this subgroup of breast cancer patients.

## 226

**The Genome Wide Association Studies (GWAS) Controversy - Is It Time To Consider A New Hypothesis To Explain The Genetics Of Multifactorial Common Diseases Such As Cancer?** *B. Gottlieb[1,2], L.K. Beitel[1,2,3], M. Trifro[1,2,3].* 1) Lady Davis Institute for Medical Research, Montreal, PQ, Canada; 2) Department of Human Genetics, McGill University. Montreal, PQ, Canada; 3) Department of Medicine, McGill University, Montreal, PQ, Canada.

Since the beginning of the Human Genome Project and throughout subsequent follow-up studies, such as the ongoing 1000 genome project, it has been assumed that the genetic principles governing the inheritance of multifactorial diseases are essentially the same as those governing locus specific diseases. These assumptions are also the basis of the genome wide association studies (GWAS) whose goal is to identify the genetic factors that contribute to common disease phenotypes. Recently, controversy has arisen as it has become apparent that cancer-associated variant genes identified from GWAS, have so far been found to carry only a modest risk for cancer. This has led to the suggestion that perhaps it is time to reconsider whether such studies are really worth the effort and resources that are presently being devoted to them. A clue as to why such studies are not leading to the expected breakthroughs has recently been suggested when we used ultra-deep pyrosequencing to examine prostate and breast cancer tissues "in depth" by "over-sequencing" the androgen receptor gene (AR) up to 50,000 times. We examined the AR CAG repeat length polymorphism within both diseased and non-diseased tumor tissues and found an extensive number of different repeat lengths (somatic mosaicism) with a clear majority variant, and almost 30 minority variants in tissue samples of c2500 cells. Even more remarkably non-diseased tissue, in one case, healthy prostate tissue from a I yr old, also exhibited multiple CAG repeat lengths. Such subtle gene heterogeneity can only be seen by "over sequencing" many thousands of times and so the arrival of next generation sequencing is likely to reveal a whole new level of tissue genetic detail. If tissue somatic mosaicism is indeed widespread it could possibly explain the lack of effectiveness of GWAS. These observations have led us to consider that the critical process in cancer ontogeny is selection of pre-existing genetic variants rather than somatic mutation of disease-associated genes. Thus, it seems appropriate that on the bicentenary of Charles Darwin's birth to propose that natural selection operates at the cellular as well as organismal level. If this is true this would suggest that understanding cancer ontogeny and its treatment lies in analysing the affects of tissue microenvironments on cancer-susceptible cells and identifying factors within these microenvironments that select pre-existing cancer-promoting genetic variants.

## 227

**Children and Young Adults with Germline *PTEN* Mutations: Phenotypic Description and Key Features to Early Diagnosis.** *J. Mester, C. Eng.* Genomic Med Inst, Cleveland Clinic Foundation, Cleveland, OH.

While PTEN Hamartoma Tumor Syndrome (PHTS) was once thought to only describe persons with Cowden syndrome (CS) and Bannayan-Riley-Ruvalcaba syndrome (BRRS) with germline *PTEN* mutations, the phenotypic spectrum continues to expand. Clinical diagnostic criteria have been defined for CS. However, many features do not manifest until adulthood, creating a diagnostic challenge for children. The National Comprehensive Cancer Network recommends that individuals with CS begin thyroid cancer surveillance at 18 years and women initiate clinical breast exams at 25 and imaging at 30, making an early diagnosis essential. After a total enrollment of 2,625 accruing by CS and CS/BRRS-like features or presence of ≥ 5 gastrointestinal (GI) polyps (≥ 1 being hyperplastic or hamartomatous), 38 children and young adults (age < 21 years) have been identified with deleterious *PTEN* alterations. Most patients had either one or two base pair insertions or deletions leading to a frameshift (12/38, 32%) or nonsense (10/38, 26%) or missense mutations (10/38, 26%); fewer had mutations causing splice site alterations (3/38, 8%) or whole or partial gene deletions (3/38, 8%). 24 (63%) were male and the mean age at study referral was 9.6 years (range 1.9 - 20.3). Macrocephaly was found in all 37 patients where OFC was measured. Other common features include developmental delay/mental retardation/autism (25/38, 66%) and lipomas (14/38, 39%). Significantly more females (7/14, 50%) than males (2/24, 8%) were reported to have pathognomonic mucocutaneous manifestations (p=0.006) with no correlation with age observed. Other minor CS criteria observed at moderate frequency were GI polyps (8/38, 21%), thyroid nodules/goiter (7/38, 18%), and congenital genitourinary anomalies (6/38, 16%). Seven patients were reported to have hemangiomas (one CNS) and 5/28 (18%) males were noted to have hyperpigmented macules of the glans penis. Three patients had been diagnosed with malignant tumors: two females with papillary thyroid carcinoma at 11 and 17 years, and one male with a metastatic mixed nonseminomatous germ cell tumor at 18 years. These data emphasize the importance of benign skin lesions and careful physical examination in children with possible PHTS. This is especially important for females, whom these data show are more likely to exhibit subtle dermatologic features, and for whom early diagnosis is key to initiate appropriate breast cancer surveillance.

## 228

**The clinical phenotype of FG syndrome: An algorithm for diagnostic testing.** *R.D. Clark[1], J.M. Graham Jr[2], M.J. Friez[3], J.J. Hoo[4], K.L. Jones[5], C. McKeown[6], J.B. Moeschler[7], F.L. Raymond[8], R.C. Rogers[3], C.E. Schwartz[3], A. Battaglia[9], M.J. Lyons[3], R.E. Stevenson[3].* 1) Department of Pediatrics, Loma Linda University School of Medicine, Loma Linda, CA; 2) Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, CA; 3) Greenwood Genetic Center, Greenwood, SC; 4) Department of Pediatrics, University of Toledo, NW Ohio Regional Genetics Center, Toledo, OH; 5) Department of Pediatrics, University of California at San Diego School of Medicine, San Diego, CA; 6) West Midlands Regional Genetics Service, Birmingham Women's Hospital, Birmingham, UK; 7) Dartmouth -Hitchcock Medical Center, Lebanon, NH; 8) Cambridge Institute of Medical Research, University of Cambridge, Cambridge, UK; 9) Stella Maris Clinical Research Institute for Child and Adolescent Neuropsychiatry, Calambrone (Pisa), Italy.

To define the phenotype of FG syndrome, an X-linked cognitive impairment and multiple congenital anomaly syndrome, we identified 23 males from 10 families, whose disease status was confirmed with *MED12* gene analysis for the p.R961W mutation. The clinical phenotype of FG syndrome in these patients includes a large and dolicocephalic head, tall forehead and small ears. In addition to early hypotonia and feeding problems, most affected males had one or more malformations. A positive family history of X-linked mental retardation, early death of male children or multiple miscarriages was seen in all affected families. All affected males had cognitive impairment, but 1 had a normal IQ of 84. Although all mothers were carriers, they were clinically unaffected. The phenotype of these patients was compared to 38 patients previously diagnosed with FG syndrome but lacking the p.R961W mutation in *MED12*.The most discriminating traits were chosen to develop an algorithm to facilitate testing for the p.R961W *MED12* mutation. The algorithm requires male sex, no affected females, intellectual disability and at least 4 of 6 specific clinical and behavioral traits. Fewer features are required in the neonate, who cannot demonstrate the expected behaviors, and more features are required without an X-linked or compatible family history. The algorithm is 100% sensitive and 90% specific in the group studied to date. As such it may be useful to the clinician evaluating patients with suspected FG syndrome. (rclark@long.cnc.net).

## 229

**Genotype-facial phenotype correlation in Wolf-Hirschhorn syndrome.** *P. Hammond[1], M. Suttie[1], K. Devriendt[4], J. Vermeesch[4], F. Hannes[4], F. Forzano[5], F. Faravelli[5], S. Williams[2], S. South[3], J. Carey[3], O. Quarrell[2].* 1) Inst Child Hlth, Univ Col London, London, United Kingdom; 2) Sheffield Children's Hosp, Sheffield, United Kingdom; 3) The University of Utah, Salt Lake City, UT, United States; 4) Center Hum Gen, University Hosp Gasthuisberg, Leuven, Belgium; 5) Galliera Hosp, Genoa, Italy.

Wolf-Hirschhorn syndrome (WHS) has characteristic facial features resulting from terminal 4p deletions which may be small or large; or result from unbalanced translocations. Those with small deletions are less likely to have congenital anomalies. We used 3D imaging and dense surface modelling to investigate whether facial features correlate with the aetiology of the 4p deletions. 3D facial images and routine karyotype results were collected from patients attending support group meetings. Excluding adults and unusable images, dense surface models computed from the 3D facial images of 100 Caucasian WHS cases (mean age of 7.9 years, range 1.0 - 18.9 years) were compared with 200 controls (mean age of 8.7 years, range 0.2 - 20.7 years). The WHS cohort consisted of 28 large deletions (breakpoint proximal to 4p16.3); 23 small deletions (breakpoint within 4p16.3); 2 instertitial deletions; 15 (4;8) translocations; 13 other translocations and 19 unclassified mainly because the breakpoint was recorded as 4p16. Pattern recognition algorithms (closest mean, linear discrimination analysis and support vector machines) supported a classification rate of 100% between WHS and control images using multifolded testing of 20 dense surface models generated from randomly selected 90% training and 10% unseen subsets. Two atypical children with a mild WHS phenotype and 4 others from different ethnic backgrounds were tested unseen against the 20 dense surface models; all were classified as more similar to the WHS subgroup. Compared to controls (n=208) and children with other developmental disorders such as Noonan (n=62), Williams (n=85) and 22q11DS (n=84) syndromes, there was a significant time lag in facial growth in WHS that does not seem to depend on deletion size. Those with larger deletions had greater facial dysmorphology and greater facial asymmetry than those with smaller deletions. There is increasing interest in those few children with the smallest 4p deletions which define the WHS critical regions. We suggest 3D imaging and dense surface modelling are used for objective assessment of facial dysmorphology in these crucial cases.

## 230

**Expression as a cellular phenotype for Proteus syndrome.** *M.J. Lindhurst[1], T.C. Huber[1], A. Elkahloun[2], L.G. Biesecker[1].* 1) GDRB, NHGRI/NIH, Bethesda, MD; 2) CGB, NHGRI/NIH, Bethesda, MD.

Proteus syndrome (PS) is rare, sporadic disorder that is characterized by progressive, disproportionate, asymmetric postnatal overgrowth that is mosaic in nature. Characteristic manifestations include cerebriform connective tissue nevi (CCTN), linear epidermal nevi, disproportionate overgrowth of the limbs, and vascular malformations. Happle hypothesized that a genetic alteration occurs post-zygotically, which causes growth dysregulation in tissues with mutant cells. Because PS is not inherited, traditional methods to identify the molecular etiology have not been successful. Complicating efforts to study this disorder is the fact that while lesions can be identified at the level of the tissue, a cellular phenotype has yet to be discovered. We have compared the transcriptomes of cells cultured from affected and unaffected tissues of Proteus patients and also compared them to the transcritomes of non-Proteus cells. RNA from 29 cell lines derived from five Proteus patients and seven non-Proteus individuals was hybridized to Affymetrix GeneChip Human Gene 1.0 ST arrays. Twenty cell lines were from dermal fibroblasts and were analyzed using GeneSifter Analysis Edition software. A comparison of fibroblast transcriptomes from affected Proteus cells and non-Proteus cells showed over 500 genes with altered expression. We found up regulation of six genes that made biological sense given the nature of the lesions commonly found in PS. They are: *COL14A1*, *COL15A1*, *COL6A3*, *COL21A1*, *FBN2*, and *CD9* and their up-regulation has been confirmed by quantitative PCR. In addition to the qPCR validation of the array results, preliminary data show that circulating leukocytes have increased levels of CD9 protein by FACS (fluorescence activated cell sorting). CCTNs found in PS contain massive amounts of collagen. FBN2 is a component of connective tissue microfibrils and may be involved in elastic fiber assembly. CD9 is a cell surface molecule with many functions including modulation of cell adhesion and motility. These gene products are all components of the extracellular matrix or are involved cell growth and migration both of which are abnormal in PS. Identification of these and other transcripts that are mis-regulated in affected PS cells is an important step in developing tools for the isolation of pure populations of cells that can be used to identify the molecular lesion responsible for PS and to develop more rigorous diagnostic criteria for the disorder.

## 231

**Quantitative assessment of endocrinologic and cognitive abnormalities in a cohort of 53 patients with Bardet-Biedl syndrome.** *D. Ng[1], P.P. Feuillan[1], J.C. Sapp[1], A. Bottar[1], L. Guevara[1], E. Wiggs[2], B. Brooks[3], L.G. Biesecker[1].* 1) Genet Disease Research Branch, NHGRI, NIH, Bethesda, MD; 2) Office of the Clinical Director, NINDS, NIH, Bethesda, MD; 3) Ophthalmic Clinical Genetics Section, NEI, NIH, Bethesda, MD.

Bardet-Biedl syndrome (BBS, OMIM #209900) is a disorder of primary cilium. BBS is a rare pleiotropic genetic disorder affecting the development of multiple organ systems and is characterized by cone-rod dystrophy, postaxial polydactyly, truncal obesity, cognitive impairment, male hypogonadism, abnormal genital tract in females, and renal malformations. Past studies have found a high frequency of endocrine related disorders among BBS patients and include obesity (72%), irregular menses, male hypogenitalism (89%), and pubertal delay (31%). We report a detailed cross-sectional assessment of the cognitive, endocrine and metabolic perturbations in BBS to include BMI, fat distribution, glucose metabolism (assessed with oral glucose tolerance testing [OGTT]), characterization of hypogenitalism, ultrasound measurements of ovaries, uterus, testes and formal neuropsychological evaluations. In the first 36 months of this study, 40 families containing 53 affected individuals have been enrolled. Twelve of these families had two or more affected siblings. The age range of the BBS subjects was 1.5-31 yr (mean 10.4 ± 6.8 yr). Physical findings include retinal rod/cone dystrophy (100%), polydactyly (78%), renal anomalies (60%), and microphallus (30%). The majority of affected subjects were obese (mean total body fat 41± 8%; normal, 22± 9%). The exceptions were two children with BBS6/MKKS mutations who had normal BMI. Most of the 53 patients had elevated serum glucose and insulin at baseline and between 30-120 min during the OGTT. Many of these subjects had evidence of dyslipidemia; serum cholesterol > 200 mg/dL (19%), HDL < 40 mg/dL (44%), and triglycerides > 150 mg/dL (40%). Among the fifty-three patients who underwent neuropsychological evaluations, thirty-eight full-scale IQ (FSIQ) scores were obtained. The mean FSIQ was 75.87 (range 41 to 112). These scores were matched with the patients' visual acuities and there was no correlation between poor visual acuity and IQ. Patients affected with BBS have an increased risk for developing elements of the "metabolic syndrome". By characterizing the metabolic phenotype of BBS patients, we hope to gain a better understanding of the relationship between ciliopathies and hyperhagia; determine if there are any differences among the 12 BBS genotypes, and develop effective treatment for this hyperphagia-induced obesity/metabolic syndrome.

## 232

**A Novel Mutation of LAMB2 Resulting in Non-Congenital Retinal Disease and Nephrotic Syndrome: a new variant of Pierson Syndrome.** *N. Lindor[1], B.G. Mohney[2], J.S. Pulido[2], D.H. Morton[3], E. Puffenberger[3], M.C. Hogan[4], M.B. Consugar[4], J. Peters[4], L. Eldahdah[5], P.C. Harris[4].* 1) Dept Medical Genetics /E7B, Mayo Clinic, Rochester, MN; 2) Ophthalmology, Mayo Clinic Rochester; 3) Clinic for Special Children, Strassburg, Pennsylvania; 4) Division of Nephrology, Mayo Clinic Rochester; 5) Reproductive Genetics Institute, Chicago, Illinois.

**Background:** Pierson Syndrome is a rare disorder characterized by congenital renal failure due to nephrotic syndrome, bilateral microcoria (small pupil with risk for glaucoma), and neurologic impairment in long term survivors, conferred by biallelic inactivating mutations in the laminin beta 2 gene, *LAMB2*. Recent reports of missense mutations suggest a broader phenotype is possible [Hasselbacker et al.,Kidney International 2006]. We have identified a family with a homozygous missense mutation in *LAMB2* with a milder and variant eye/renal phenotype. **Methods:** We examined the blood pressure, eyes, urine, and genetic material from over 50 members of an extended Mennonite family that had presented with bilateral retinal detachments and neprhotic syndrome in two members under the age of 10 years. Homozygosity mapping on 5 affected cases, plus 11 unaffected family members using a 300K SNP array was performed. Candidate genes within the common area of homozygosity were identified and analyzed. **Results:** The mapping studies mapped the disease gene to a 9Mb interval on chromosome 3. Considering the eye and renal phenotypes, *LAMB2* was a strong candidate within the interval. A homozygous missense mutation was found in *LAMB2* (H147R) in 9 affected family members but not in unaffected members or in 91 controls from outside the Mennonite population. Within the Pennsylvania Mennonite population a carrier frequency of 2.2% was found, suggesting a founder effect for this *LAMB2* missense mutation. Although a relatively conservative change, the substitution is at a highly conserved residue. The retinal findings in affected individuals were unique and included band keratopathy, cataracts, pigmentary retinopathy, retinal detachments, and optic atrophy. Diagnosis of eye changes preceded diagnosis of renal disease. No family members had microcoria. **Conclusions:** Pierson syndrome had not been considered likely in this family prior to study because the eye features differed from published cases. Homozygosity for *LAMB2* H147R mutation causes a non-congenital onset variant of Pierson syndrome and this mutation has a high carrier frequency in the Mennonite population.

## 233

**Phenotypic Characterization of Poland Sindrome Based on a Series of 180 Patients.** *A. Baban[1,2], M. Torre[3], M.T. Divizia[1], A. Buluggiu[3], S. Bianca[4], S. Gimelli[5], N. Catena[6], L. Bricco[1], S. Costanzo[3], G. Gimelli[1], F.M. Sénès[6], M.G. Calevo[7], M. Valle[8], R. Ravazzolo[1,9], V. Jasonni[3], M. Lerone[1].* 1) Molecular Genetics Unit, Gaslini Children Hospital, Genova- Italy; 2) Cardiovascular Department, Gaslini Children Hospital, Genova- Italy; 3) Pediatric Surgery Department, Gaslini Children Hospital, Genova- Italy; 4) Center of Genetic and Teratologic Counselling, Fetomaternal Department, ARNAS Garibaldi-Nesima, Catania- Italy; 5) Department of Genetic Medicine and Development, University of Geneva Medical School, and University Hospitals, 1211 Geneva, Switzerland; 6) Orthopedic and Traumatology Department, Gaslini Children Hospital, Genova- Italy; 7) Epidemiology and Biostatistics Section of Scientific Direction, Gaslini Children Hospital, Genova- Italy; 8) Radiology Department, Gaslini Children Hospital, Genova- Italy; 9) Department of Pediatrics and CEBR, University of Genova- Italy.

Poland Syndrome (PS) consists of unilateral hypoplasia/ aplasia of the pectoralis muscles in isolated entity or in association with ipsilateral upper limb or thoracic anomalies (ULA or TA). Associated ULA are highly variable from normal to mild hypoplasia or typical symbrachydactyly and syndactyly of short fingers. TA can be different: agenesis or hypoplasia of the ribs, pectus carinatum or excavatum, variable degrees of breast involvement, high scapula, and others. The phenotypic expressivity of PS is highly variable and its characterization has not been described in international literature on such a wide patient series. We have considered the following aims in designing our project: Describe the disease phenotype in a wide range of patients. Verify the current data present in literature. Classify disesase severity according to clinical features and identify risk factors according to gender, affected side and other phenotypic charecteristics. Define associated malformations or syndromes. Obtain best methods in management of patients from diagnostic, therapeutic and prognostic points of view. Obtain specific radiologic standards through pectoral muscles ultrasound. Possibly identify new etiopathogenetic hypotheses (genetic versus environmental factors) and validate those present in literature. We have studied 180 patients (60.7% M - 39.3% F) with monolateral hypoplasia or agenesis of pectoralis muscles in the period 2003 and 2009. The management of these patients was based on multidisciplinary approach. At the first phase of the study all included patients had undergone Specialistic Counselling (Genetic, Psychologic, Radiologic, Surgical, and Orthopedic). The second phase was based on medical indication and included high resolution karyotyping or array-CGH. Moreover, other investigations included chest X-ray, Echocardiography, Abdominal ultrasound, and thoracic CT scan. We shall focus in this work to present the preliminary results regarding different phenotypic, array-CGH and radiologic characterizations. In collaboration with AISP (Italian Association of Poland Syndrome) This Spoken presentation was awarded young researcher prize at the Italian Society of Human Genetics Conference, Montecatini - 2007. This spoken presentation was a finalist young investigator awards for the European Society of Human Genetics, Barcelona- 2008.

## 234

**15q25.2 Deletion Syndrome: A Novel Recurrent Microdeletion Syndrome Characterized by Failure to Thrive, Developmental Delay and Increased Risk for Congenital Diaphragmatic Hernia.** *D.A. Scott[1], M.J. Wat[1], V.B. Enciso[2], C.J. Fernandes[1], A. Johnson[1], K.P. Lally[3], D. Tibboel[4], A. de Klein[4], E.R. Roeder[2], D. Freedenberg[5], P. Stankiewicz[1], S.W. Cheung[1].* 1) Baylor College of Medicine, Houston, TX; 2) University of Texas Health Science Center, San Antonio, TX; 3) University of Texas Health Science Center, Houston, TX; 4) Erasmus Medical Center, Rotterdam, The Netherlands; 5) Department of State Health Services, Austin, TX.

Array comparative genome hybridization has proven to be a powerful tool for identifying recurrent chromosomal deletion and duplications associated with sporadic structural birth defects such as congenital diaphragmatic hernia (CDH). To identify novel genomic regions harboring genes that play a role in diaphragm development, we screened a cohort of patients with isolated and non-isolated CDH. In this screen we identified an ~2.3Mb deletion on chromosome 15q25.2 in a thirteen year old male with CDH, failure to thrive, developmental delay, cryptorchidism and a marfanoid habitus. This region contains 15 genes and is flanked by low copy repeats (LCRs). A search for additional reports of CDH associated with deletions of this region in a cohort of 12,000 cases referred for microarray testing revealed two reported cases; a fetus with CDH who had a large de novo deletion that extended to 15q26.2, and a fetus with CDH and mild hydrocephalus with a de novo ~2.5 Mb deletion of 15q25.2 flanked by the same proximal LCR seen in our patient but extending to a slightly more distal LCR. Two other individuals with overlapping de novo 15q25.2 microdeletionsone previously reported and one recently identified by our groupdid not have CDH but presented with failure to thrive, developmental delay and a variety of other structural defects including polysplenia, severe portal vein stenosis, cleft palate, and a coronary artery fistula. Common features present in these patients allow us to define a new recurrent microdeletion syndrome characterized by failure to thrive, developmental delay and increased risk for CDH. Interestingly, deletions of an adjacent, but non-overlapping, region of 15q25.2 flanked by LCRs have been described in two individuals with autism and two with schizophrenia suggesting the possible existence of two distinct microdeletion syndromes located on 15q25.2.

## 235

**Prospective comparison of 6 genome wide array genomic hybridization platforms for the detection of copy number variants in mental retardation.** *T. Tucker[1], A. Montpetit[2], D. Chai[3], S. Chan[4], S. Chénier[2], A. Delaney[4], P. Eydoux[3], S. Langlois[2,5], E. Lemyre[2], M. Marra[4], H. Qian[4], D. Vincent[2], J. Michaud[2], JM. Friedman[1,5].* 1) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Centre de Recherche, CHU Sainte-Justine, Montréal, Canada; 3) Children's & Women's Hospital, Vancouver, Canada; 4) Genome Sciences Centre, BC Cancer Agency, Vancouver, Canada; 5) Child & Family Research Institute, Vancouver, Canada.

Many clinical laboratories are adopting array genomic hybridization as a standard clinical test to detect pathogenic copy number variants (CNVs) in patients with mental retardation (MR), autism and other birth defects. A number of whole genome array genomic hybridization systems are commercially available, but little is known about the comparative performance of these platforms in a clinical context. We prospectively studied 30 children with idiopathic MR and both unaffected parents using Affymetrix 500K GeneChip® SNP arrays, Agilent® Human Genome 244K oligonucleotide arrays and NimbleGen® 385K Whole-Genome oligonucleotide arrays. We also determined whether CNVs called on these platforms were detected by Illumina® Hap550 arrays or SMRT® 32K BAC whole genome tiling arrays and tested 15 of the 30 trios on Affymetrix 6.0 arrays. Samples were processed according to each manufacturer's specifications, and CNVs were identified using each manufacturer's software with default settings. We restricted our comparison to the autosomes. Using the Affymetrix 500K, Agilent and Nimblegen platforms, we identified 145 autosomal de novo CNVs, but only 30 (21%) were identified on more than one platform and only 10 (7%), by all 3 platforms. Five deletions (78 Kb-9.8 Mb) are likely to be pathogenic for the child's MR. One child has a 657 Kb deletion on chromosome 17q21.31 that has been found in other MR patients, and another child with clinical features of Rubinstein-Taybi syndrome has a 78 Kb deletion that encompasses the first exon and upstream sequence of CREBBP. The other pathogenic deletions involved 6q21-q22.31, 9q34.3, and 22q11.21. We also identified 10 other CNVs for which the clinical significance is uncertain. Although not included in the comparison, a 9 Mb pathogenic duplication was identified on Xq21-q21.1. All 6 de novo CNVs were identified by the 3 main platforms and Illumina, but only 3 were identified by the SMRT BAC array. Two of these children were also tested on the Affymetrix 6.0 array, and their pathogenic CNVs but no additional pathogenic CNVs were identified on this platform. The large number of apparently false positive CNV calls on each of the platforms studied supports the need for validating suspected pathogenic CNVs with a different method.

## 236

**PATIENTS ENTERING THE HEALTHCARE SYSTEM FOLLOWING DIRECT-TO-CONSUMER GENETIC TESTING.** *M. Giovanni[1], M. Fickie[2,3], L. Lehmann[3], L. Meckley[4], M. Murray[1].* 1) Division of Genetics, Brigham & Women's Hospital, Boston, MA; 2) Harvard Medical School Medical Genetics Training Program, Boston, MA; 3) Center for Bioethics, Brigham and Women's Hospital, Boston, MA; 4) Center for the Evaluation of Value and Risk in Health, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA.

**Context:** Direct-to-consumer (DTC) genetic testing has received significant attention in the popular press. Consumers incur out-of-pocket costs in an effort to learn about their risk of developing disease, susceptibility to allergy, or any number of other traits. Little is known about the economic and clinical impact of "consumer genetics" on healthcare delivery systems. **Objective:** To identify cases where DTC genetic testing prompted consumers to become patients and to understand potential costs and benefits associated with this phenomenon. **Study Design:** A case series identified through a survey of genetic professionals who reported their experience with consumers seeking consultation following DTC genetic testing. **Main Outcome Measure:** Analysis of costs and benefits associated with identified cases of consumer to patient transition. Costs are based on Medicare reimbursement standards using national averages. **Results:** In this study, we report 24 cases of consumers becoming patients following DTC genetic testing. The majority of patients self-referred to genetics professionals, though approximately 25% were referred by a primary care physician or other provider. The estimated cost to the healthcare system ranged from $40 to $20,604 while the out-of-pocket healthcare cost for the patient ranged from $20 to $5,565. The health concerns which prompted consultation included cancer, celiac disease, hemochromatosis, nutrigenetics, and genetic issues with non-healthcare focus (e.g. ancestry). The geneticist judged the testing to have been useful in half of the cases. Referral for additional consultation or testing clustered in 65% of cases. **Conclusions:** The healthcare cost generated for some consumers turned patients exceeds the out-of-pocket expense for DTC genetic testing. The healthcare system incurred significant, and sometimes appropriate, costs based on consumer-driven genetic testing. Further research is needed in order to: [1] characterize the magnitude of the trend in genetic testing and the relevant cost and benefits of such testing, [2] guide policy related to this use of technology, and [3] prepare healthcare systems for the optimal management of these consumers turned patients.

## 237

**Estimating the number of unseen genetic variants in the human genome.** *I. Ionita-Laza[1,2], C. Lange[1], N. Laird[1].* 1) Dept of Biostatistics, Harvard University, Boston, MA; 2) Dept of Biostatistics, Columbia University, New York, NY.

The different genetic variation discovery projects (including the HAPMAP and the 1000 Genomes Projects) aim to identify as much as possible of the underlying genetic variation in various human populations. The question we address here is how many new variants are yet to be found. We regard this question as an instance of the species problem in ecology, where the goal is to estimate the number of species in a closed population. We use a capture-recapture approach that allows us to calculate, based on data for a small number of individuals, the expected number of new variants with a desired minimum frequency to be discovered in a future dataset of a specified size. The method is also useful in follow-up studies of genome-wide association studies, as it allows prediction of the minimum number of individuals necessary to be sequenced in a region in order to capture all or a fraction of the underlying variation in the region. We show applications to several datasets, including the ENCODE dataset, the SeattleSNPs dataset and the NIEHS SNPs dataset. Although these applications concern SNPs, the same approach may be applied to copy-number variation data to predict the number of still to be identified copy-number variable regions. We highlight here several results from our applications. Consistent with previous findings, our results show the African population to be by far the most diverse in terms of the number of variants expected to exist (especially with respect to rare variants), the Asian populations the least diverse, with the European population in-between. In addition, our results show a clear distinction between the Chinese and the Japanese populations, with the Japanese population being the less diverse. Finally, we present results of two new non-parametric tests that show that these between-population differences are highly statistically significant.

## 238

**The potential benefits of trans-ethnic fine mapping.** *J. Byrnes, E. Hechter, S. Myers, G. McVean.* Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom.

Both the risks and benefits of using genotype data from multiple populations in large-scale association studies are fairly well understood. However, there are few examples of using trans-ethnic data in fine mapping analysis, and the benefits of this approach have yet to be thoroughly investigated. It is hypothesized that using data from multiple populations, each with distinct haplotype structure, may aid in localizing an association signal further. We have performed a detailed simulation study to compare an experimental design in which all individuals are sampled from a single ethnic population to a design in which an equivalent number of samples are taken from two or more distinct populations. Using phased ENCODE haplotype data from CEU, JPT+CHB, and YRI as surrogate fine mapping regions, we simulate either 6,000 cases and 6,000 controls from a single population, or 2,000 cases and 2,000 controls from each of the three populations. Data was simulated under an additive disease model assuming a single causal variant with various levels of relative risk. We then apply two approaches to testing for disease association. Under the first approach, we combine all data, regardless of population identity and calculate the Bayes factor of association under an additive disease model for each SNP in the region. Under the second approach, Bayes factors are calculated for the data under each population separately and then multiplied together. We summarize the evidence of association across a region by constructing credible sets that contain all SNPs comprising 95% of the posterior probability of disease association within each region. Finally we compare the credible sets, in terms of set size and presence/absence of the causal variant, under both sampling schemes and under both testing approaches. Our results demonstrate significant improvement in localizing a fine mapping signal when using trans-ethnic data and a testing approach that respects population identity. This approach produced consistently smaller credible sets that contained the causal variant at least as frequently as the approach that ignores population identity. These results are likely to have significant impact on future fine mapping studies by providing a new, more effective methodology.

## 239

**Sequencing a small reference panel in an isolated population for imputation using Identity-By-Descent.** *I. Pe'er[1], M.J. Shah[1], E.E. Kenny[1,4], C. Lee[2], H. Peckham[2], Q. Doan[2], J.K. Lowe[3,5], J. Salit[4], M.J. Daly[3], M. Stoffel[6], D.M. Altshuler[3], J.M. Friedman[4], J.L. Breslow[4], A. Gusev[1].* 1) Dept of Computer Science, Columbia University, New York, NY; 2) Applied Biosystems Inc. 500 Cummings Center, Beverly, MA; 3) Broad Institute of Harvard and MIT, 7 Cambridge Center, Cambridge MA; 4) The Rockefeller University 1230 York Avenue New York, NY; 5) UCLA, 695 Charles E Young Dr Los Angeles, CA; 6) ETHZ, Wolfgang-Pauli-Str. 16 Zürich.

High-throughput sequencing allows genetic investigation beyond previously restricted studies of common, known alleles assayed on SNP arrays. Still, whole genome sequencing a sizeable sample is prohibitively expensive, motivating the generation of relatively inexpensive genotype data in conjunction with a sequenced reference panel to impute the un-sequenced individuals. We investigate the utility of an isolated population for such studies relying on two upfront advantages: Population bottlenecks may allows direct, more powerful imputation than statistical methods used in other cohorts; furthermore, in isolated populations causal alleles of strong effect that are unknown and otherwise rare may drift to appreciable frequency and show detectable association. We focus on a cohort from the Pacific Island of Kosrae, Micronesia, where we previously collected SNP array and rich phenotype data for most of the population. We report identification of long regions that were historically co-inherited between pairs of individuals and are therefore identical by descent (IBD). The totality of these regions reveals pervasive presence of IBD sharing, with an average indigenous sample sharing the average locus with 10.85% of the entire 3,000 person. We present a novel method for leveraging such high degree of shared genetic content for imputation of large fraction of the "population genome" from sequence data. Briefly, we assume that a sequenced individual can propagate the sequence content to all individuals with which it shares an IBD segment, at the respective location in the genome. Using this strategy, we have developed an efficient algorithm for quantifying the exact amount of unique information that an arbitrary set of individuals infer to the population, coupled with a greedy heuristic that prioritizes individuals for sequencing to maximize imputation power. Our findings show that by sequencing only 7 individuals (0.2% of the population), we can infer sequence for at least 50% of the population at the average locus. We have collected a pilot data-set of whole genome sequences from four Kosraean individuals, with limited coverage of one SOLiD per sample (approximately 4X). Analysis of the sequence concordance in IBD shared regions within these pilot samples is informative with respect to the overall strategy, the computational method and practical considerations of sequencing-and-imputation. Such information may guide subsequent studies in additional populations.

## 240

**Squeezing Value from Next Generation Imputation Reference Sets.** *L. Jostins, J.C. Barrett.* Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

Genotype imputation allows high-density reference genotype data to be used in inferring unobserved genotypes in low-density samples, such as produced for a GWAS. With the advent of increasingly powerful genotyping and sequencing platforms, new reference sets are becoming available. We investigate what improvements one such reference set, the HapMap3 haplotype map, can grant to imputation, with particular emphasis on rare variants, which have previously been challenging to impute.

A pronounced improvement in the efficiency of imputing the genotypes of UK samples is observed between the HapMap2 and HapMap3 reference sets, when imputed using the program Beagle. Using the larger HapMap3 CEU reference set grants a modest improvement over the HapMap2 CEU set, and integrating the diverse HapMap3 populations grants additional improvements. Overall, this next generation panel gives a 14% increase in the number of SNPs confidently called, and a 96% drop in highly overconfident calls. We investigate what drives this improvement, and conclude that sample size, SNP density and genotyping quality each contribute, with increased sample size being the most significant contributing factor. This effect is strongest for SNPs with a low minor allele frequency.

We test the hypothesis that the improvements are also driven by the increased diversity of the reference panel, allowing better representation of rare alleles. To investigate the trade-off between specificity and diversity in the reference panel, we constructed equal-sized European reference sets with varying degrees of diversity. We find that the imputation of common SNPs is highly robust to reference panel diversity, but a diverse panel grants a small but significant improvement in the imputation of rare SNPs.

These results show how increases in genome reference sets can give increased performance now, as well as boding well for the future, indicating that large and diverse sequencing projects such as the 1000 Genomes Project will allow increasingly accurate imputation of rare variants.

## 241

**TAKING ADVANTAGE OF DISTANT RELATEDNESS: GENOTYPE IMPUTATION IN THE RESEQUENCING ERA.** *Y. Li, G.R. Abecasis.* Dept Biostatistics, Univ Michigan, Ann Arbor, MI.

Genomewide association studies have successfully identified many complex disease susceptibility loci by genotyping a subset of common genetic variants across many individuals. With the rapid deployment of next generation sequencing technologies, it is anticipated that future genetic association studies will be able to more comprehensively survey genetic variation, both to identify new loci that were missed in the original round of genomewide association studies and to finely characterize the contributions of identified loci. Next generation sequencing technologies can aid genetic association studies in two ways. Directly, they can be used to characterize the genomes of individuals sampled in disease studies and studies of other medically interesting traits. Indirectly, they can be used to discover and genotype SNPs in reference samples that will then be used to guide analyses of disease studies. We have previously developed a Hidden Markov Model that identifies short stretches of chromosome shared between individuals and uses these stretches to impute genotypes that are missing in one individual but are observed in other carriers of the same haplotype. We have now extended this model to the analysis of shotgun sequence data so that it can combine short read sequences and their corresponding quality scores across many individuals. Using simulations, we show that by sequencing 200 - 1000 individuals at 4-8x per individual error rates of <99.9% in individual genotype calls can be obtained at polymorphic sites; comparable to the accuracy that might be obtained by sequencing a single individual at 20-30x depth. To evaluate our method in practice, we analyzed preliminary shotgun sequence data generated by the 1000 genomes project on 52 CEU individuals sequenced at ~4.7x average depth per individual. We then used estimated haplotypes for these individuals to fill in missing genotypes in a case-control study of type 2 diabetes. Even with this small reference panel, missing alleles in our type 2 diabetes case-control study could be imputed with an average error rate of ~2.6%. As the size and quality of the reference panel increases, we expect this average error rate to improve rapidly. Our preliminary results validate the utility of low depth shotgun sequence for discovering and genotyping SNPs across many individuals. Further, they show that the resulting data can be used to accurately impute missing genotypes in existing genomewide association studies.

## 242

**Shared Genomic Segment Analysis: A Powerful Method for Detecting Rare Risk Variants.** *S. Knight, R. Abo, N. Camp.* Univ Utah, Salt Lake City, UT.

Shared genomic segment (SGS) analysis is a method that uses dense SNP genotyping in extended high-risk pedigrees to identify regions of sharing between distantly related terminal cases (at the end of pedigree branches). Sharing is established for all terminal cases (n), (n-1), (n-2) and (n-3) cases, with significance assessed empirically. Here, we test the power of this approach and compare to the popular case-control association design to identify rare risk variants. We used simulation to generate extended high-risk pedigrees. Pedigrees were simulated to be 5 generations. Simulated genotypes were based on the Illumina 610K SNP array for chromosome 21 with allele frequencies and LD structure determined from the HapMap CEPH. We placed a disease SNP (dSNP) on the chromosome with an allele frequency of 0.0005, replacing one of the rarer SNPs on the array. Founders in the pedigree were assigned haplotypes and a gene-drop performed according to an established genetic map. Disease status was assigned based on a sporadic rate of 2% and a relative risk of 20 for carriage of the dSNP. The dSNP was then removed from the data. A high-risk pedigree was determined to be one with ≥17 cases to represent excess clustering (3-fold higher than the sporadic rate). Only high-risk pedigrees that contained at least 15 meioses between the terminal cases were selected. We generated 100 such high-risk pedigrees and performed SGS analysis. For each analysis, we identified whether the segment containing the true risk locus was identified as the longest shared region, and if so, whether it was statistically significant. The average number of terminal cases in a pedigree was 12 (on average 3 were sporadic). The (n-3) SGS sharing analysis performed the best: 47% of the pedigrees correctly captured the risk locus on the longest shared segment; however, in only 32 was the length significant. Hence, under our model parameters, the power of a single pedigree analyzed using SGS (n-3) was 32%. Therefore, a resource of ≥5 high-risk pedigrees will attain >85% power. The power in a case-control design for such model parameters is driven by the LD between the unobserved risk allele and the best tSNP in the array. For our model parameters, this is very low and the power for a case-control study is less than 1%, even for excessively large samples sizes. In conclusion, SGS analysis appears to be a promising and powerful method for detecting rare risk variants.

## 243

**Variation in relatedness estimates.** *B.S. Weir[1], C.C. Laurie[1], W.G. Hill[2], L.J. Bierut[4], A.D. Anderson[3], The GENEVA Invetigators.* 1) Dept Biostatistics, Univ Washington, Seattle, WA; 2) Institute of Evolutionary Biology, Univ Edinburgh, Edinburgh, UK; 3) Department of Mathematics Western Washington University, Belingham, WA; 4) School of Medicine, Washingon University St Louis, MO.

There are many reasons to estimate relatedness for pairs of individuals, including our current activities in cleaning GWAS data for the GENEVA project. We concentrate on estimating k0 and k1, the probabilities that individuals share zero or one pair of alleles identical by descent, and we use a maximum likelihood method that avoids bias problems with current moment-based methods. We see large variation in estimates around the values expected for full-sibs, half-sibs, first-cousins etc but our new estimates for the variances and covariances of the estimates show that the variation is generally at the expected level. This variation, caused by evolutionary sampling, makes it difficult to decide when individuals are related versus unrelated or to distinguish between relationships such as half-sibs and first-cousins. Further examination of variation in relatedness estimates along the genome may allow the detection of the effects of natural selection or association with disease genes., but only if the expected variation is taken into account. We have found it useful to translate relatedness estimates into measures of connectedness among sets of individuals that may represent family clusters. We illustrate our methods with GENEVA data.

## 244

**High resolution detection of identity by descent with linkage disequilibrium modelling.** *S.R. Browning, B.L. Browning.* Dept Statistics, Univ Auckland, Auckland, New Zealand.

Individuals are identical by descent (IBD) at a locus if they share genetic material due to co-inheritance from a common ancestor. Usually the concept of IBD is applied to individuals of known relationship, and the common ancestor is only a small number of generations in the past. However, the IBD concept is also applicable to individuals without known relationship ("unrelated" individuals), as all pairs of individuals share common ancestors at some time in the past. The larger the number of generations since a common ancestor, the smaller the IBD regions will be. For example, after 10 generations, the average IBD tract length is 5cM, whereas after 50 generations, the average length is 1cM. Detection of identity by descent (IBD) in "unrelated" individuals has important, wide-ranging applications including relationship inference, population-based linkage analysis (IBD mapping), improved haplotype inference, imputation of ungenotyped variants, genotype error detection, and detection of deletion structural variants.

Existing approaches to IBD detection either require markers to be in linkage equilibrium or rely on observed length of identity by state. Typically, such methods can detect IBD regions of length >5cM. We have recently developed a new approach to IBD detection that is based on our localized haplotype clustering model (implemented in BEAGLE). Our method estimates posterior probabilities of IBD based on haplotype probabilities, and thus accounts for linkage disequilibrium. We use 1958 British Birth Cohort data to show that our method can reliably detect very small (≤ 2 cM) IBD regions in data from the Illumina 500K or Affymetrix 500K platforms. There is a high level of overlap between the detected regions found using the two platforms. With data from a 1M SNP platform, even smaller regions (< 1 cM) can be detected.

**245**

**When do summary odds ratios and association p values reveal membership of individuals in genome-wide association scans?** *K. Jacobs[1,2,3], M. Yeager[1,2], P. Kraft[4], S. Wacholder[2], M. Tucker[2], R. Hoover[2], S. Chanock[2], N. Chaterjee[2].* 1) Core Genotyping Facility, Advanced Technology Program, SAIC-Frederick, Inc., NCI-Frederick, Frederick, MD; 2) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Cancer Institute, National Institutes of Health, Bethesda, MD; 3) BioInformed LLC, Gaithersburg, MD; 4) Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health, Boston, MA.

Genome-wide association studies (GWAS) have successfully identified novel disease susceptibility loci. Sharing of GWAS data for the purposes of combined and meta-analysis has been a high priority for many groups (e.g., dbGaP, CGEMS, WTCCC). Until recently aggregate data, such as genotype frequencies, odds ratio estimates, and association p-values were thought to reveal negligible information regarding any given individual's participation within the study. Homer et al. showed, however, that SNP probes can detect an individual's contribution to an admixed DNA sample using SNP probe data. We have shown that if an individual's genotypes are known, genotype frequency data from the scan can reveal with confidence whether an individual or their close relative participated in a GWAS and, sometimes, their phenotypic status. We develop a novel test statistic and demonstrate the specificity and sensitivity with which one can determine whether an individual or their close relative participated in a GWAS and their phenotype status, utilizing odds ratio estimates or even just the direction of the effects from the study. As with the previously reported techniques, the specificity and sensitivity depend on the size of the study, the number of genetic variants reported, and the frequency of the genetic variants. In addition, we show that the most significantly associated variants (i.e. those with smallest p values) provide more information than randomly chosen variants. These observations have profound implications for reporting and sharing GWAS association results. We provide guidance on the limits of how much information can be made available publicly without jeopardizing individuals' privacy.

## 246

**Using Linkage Disequilibrium Patterns, Short-Read Haplotypes, and Population Structure to Improve Data Quality in the 1,000 Genomes Project.** *B. Howie[1], A. Auton[2], G. McVean[2], J. Marchini[2], M. Stephens[1], The 1,000 Genomes Project.* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Statistics, University of Oxford, Oxford, UK.

The 1,000 Genomes Project is a worldwide effort to produce nearly complete genome sequences for hundreds of human individuals from each of several major geographic regions. This project made the important design choice to sequence a large number of people at relatively low coverage rather than sequencing fewer people at high coverage. High-coverage study designs include redundancy that can pinpoint errors in the sequencing experiments, but they are inherently unable to detect rare alleles (e.g., those with worldwide or population frequencies of 1% or less) due to sample size limitations imposed by current sequencing costs. Conversely, low-coverage designs are less able to distinguish true novel variants from sequencing errors, but they expand the range of genetic variants that could, in principle, be observed in a study. One way to mitigate these trade-offs is to use sophisticated analytical methods to improve the quality of low-coverage sequence data. It has already been recognized that models of LD patterns, such as those used in haplotype phasing and genotype imputation algorithms, can serve as "priors" that penalize putative variants that violate population-genetic expectations. In this work, we extend these ideas to capture other information that will be included in the 1,000 Genomes data: short-read haplotypes, which provide phase information when a single template DNA molecule spans two or more heterozygous sites, and population structure, which is an inherent part of the project's geographical sampling scheme. We begin by evaluating the ability of short-read haplotypes to improve phasing in a "best-case" scenario based on two trios from the project that are being sequenced at high coverage. We then describe an integrated framework for combining short-read haplotypes, a sophisticated LD model, and a simple model of population structure to phase haplotypes and call genotypes in the 1,000 Genomes low-coverage sequencing data. We focus particularly on the prospects for detecting rare variants, calling their genotypes, and accurately assigning rare alleles to their haplotype backgrounds. Our analyses focus primarily on the 1,000 Genomes data, but the results may have implications for the design and analysis of many other studies that rely on short-read sequencing technologies.

## 247

**A genome wide association analysis of 132 phospholipids.** *C. Van Duijn[1], Y. Aulchenko[1], P. Ugocsai[2], P. Pramstaller[3], G. Liebisch[2], J. Wilson[4,5], A. Johansson[6], I. Rudan[7], A. Demirkan[1], A.C. Janssens[1], A. Hofman[1], T. Axenovitch[9], B. Oostra[10], T. Meitinger[8], A. Hicks[3], C. Hayward[5], A. Wright[5], U. Gyllensten[6], H. Campbell[4], G. Schmitz[2,11].* 1) Epidemiology & Biostatistics, Erasmus MC, Rotterdam, Netherlands; 2) Institute for Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Regensburg, Germany; 3) Institute of Genetic Medicine, European Academy Bolzano (EURAC), Bolzano, Italy; 4) Centre for Population Health Sciences, University of Edinburgh, Edinburgh, UK; 5) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Edinburgh, UK; 6) Department of Genetics and Pathology, Ludbeck Laboratory, Uppsala University, Uppsala, Sweden; 7) Croatian Centre for Global Health, Faculty of Medicine, University of Split, Split, Croatia; 8) Helmholtz Zentrum München, Neuherberg, Munich, Germany; 9) Institute of Cytology and Genetics, Siberian Division of Russian Academy of Sciences, Novosibirsk, Russia; 10) Clinical Genetics, Erasmus MC, Rotterdam, Netherlands; 11) For the EUROSPAN research group.

Phospholipids are major components of biological membranes and play a key role in cell survival. We performed a genome-wide association (GWA) analysis of circulating levels of 132 distinct phospholipid species (63 phosphatidylcholines (PCs), 20 lysophosphatidyl-cholines (LPCs), 30 phosphatidylethanolamines (PEs) and 19 PE-based plasmalogens (PE-pls)) in 5 European family-based studies (N=3,938). The study identified 10 genetic regions with multiple independent loci affecting plasma phospholipid concentrations, including three regions with candidate genes based on protein function (the FADS1-FADS2-FADS3 gene cluster, LIPC and the ELOVL2 gene), one triglyceride associated gene (GCKR) and 7 new loci (RAB3II1, PDXDC1, TMEM39A, PAQR9, C2ORF16, HLA-B/LOC442200 and C14ORF83). The individual variants explain up to 12% of the variance in the traits. We performed a cluster analysis to identify the SNPs affecting the same phospholipid pathway. The strongest clustering was observed for the effects of SNPs in TMEM39A, GCKR, and LIPC,which was explained by the synergistic effect of these three genes on the PE38:1 and PE38:2 distributions. A second cluster shows a clear overlap in function of the SNPs in the FADS cluster with the ELOVL2 polymorphisms, together determining PC38:5 and PE38:5 distributions. Finally, a third cluster involves the FADS region and PDXDC1, which jointly influence the PC38:3 and PE38:3 distributions. In an independent population-based follow-up study, the Rotterdam study (N= 5,974), we found that variations in four of the genes (RAB3II1, LIPC, ELOVL2, and PDXDC1) are jointly associated with increased risk of type 2 diabetes (p=5*10-7). In conclusion, our study shows that the chromosome 11 FADS/RAB3IL1 cluster, the LIPC region, the ELOVL2 gene and the newly identified PDXDC1 gene, are key genetic regulators of phospholipid metabolism and may play an important role in diabetes and related traits.

## 248

**HDL-associating allelic variants of F13A1 affect its expression in adipose tissue.** *P-P. Laurila[1], J. Naukkarinen[1], S. Söderlund[2], J. Saharinen[1], S. Ripatti[1], I. Surakka[1], M. Gentile[1], H. Yki-Järvinen[2], M. Jauhiainen[1], M-R. Taskinen[2], L. Peltonen[1,3].* 1) Public Health Genomics Unit, NIH of Finland, Helsinki, Finland; 2) Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland; 3) The Wellcome Trust Sanger Institute, Cambridge, UK.

Inverse correlation between HDL-cholesterol level and atherosclerosis has been established. To identify potentially novel, rare variants contributing to exceptionally high or low serum HDL-cholesterol levels, we genotyped 450 Finns (Illumina370K) from EUFAM population sample, either having extremely high (>90th percentile) or low (<10th percentile) HDL-C levels. Out of these, subcutaneous fat biopsies were obtained from 54 individuals, and their global expression profiles were analyzed using Affymetrix microarrays. Two best HDL level associating SNPs were located at CETP, as in previous studies. However, the 3rd and 10th best hits (rs7766109, p=10-5 and rs4959377, p=3.45x10-5) were located within F13A1, a coagulation factor with no previously reported associations with serum lipid levels. We tested 54 additional SNPs within the F13A1 region and detected associations for 10 of them, all of them located within introns 3-5. The associations of F13A1 SNPs in introns 3-5 for HDL were replicated in a Finnish population sample with normal distribution of HDL (n=890). To obtain some functional evidence for our genetic link between HDL level and F13A1 we measured F13A1 transcript levels in fat biopsies and observed them to be higher in adipose tissue of individuals with low HDL (p=0.007), and a dose-dependent effect of rs7766109 genotype on both F13A1 expression (p=0.02) and HDL-levels (p=0.004) was observed. We also analyzed the adipose tissue expression profiles of 5 insulin sensitive (IS) and 5 insulin resistant (IR) before and after a euglycemic insulin clamp. In line with the results of the high/low HDL study, F13A1 transcript level was 6-fold higher in IR subjects compared to IS individuals (p=0.003). Moreover, an insulin-induced decrease in F13A1 expression was observed in IR (p=0.047) but not in IS subjects. Here we show the value of extreme end phenotypes and the combination of global SNP and expression analysis in the identification of new, potentially high impact genes regulating serum lipid levels.

## 249

**_INSIG1_ influences obesity-related hypertriglyceridemia in humans.** *Y. Zhang[1], E.M. Smith[1], T.M. Baye[1], S. Gawrieh[1], R. Cole[1], J. Blangero[2], M.A. Carless[2], J.E. Curran[2], T.D. Dyer[2], L.J. Abraham[3], E.K. Moses[2], A.H. Kissebah[1], L.J. Martin[4], M. Olivier[1].* 1) Med Col Wisconsin, Milwaukee, WI; 2) Southwest Foundation for Biomedical Research, San Antonio, TX; 3) University of Western Australia, Crawley, Australia; 4) Cincinnati Children's Hospital, Cincinnati, OH.

Insulin-induced genes (INSIGs) are critical feedback mediators of cholesterol and fatty acid synthesis in animals, but their role in lipid regulation in humans is unclear. We identified a QTL for plasma TG levels (LOD=3.6) on human chromosome 7q36 in a family-based cohort of 2,209 individuals of Northern European descent. We examined 29 single nucleotide polymorphisms (SNPs) across INSIG1, which is located in this QTL interval, and identified a promoter SNP rs2721 that is associated with TG levels (p=0.002) in the original MRC-OB linkage cohort. This association was replicated in a second cohort comprised of unrelated individuals (p=0.0008; n=920). In the original cohort, individuals homozygous for the T allele of rs2721 had 8% higher TG levels. To test the potential interaction between INSIG1 and its only homolog in humans, INSIG2, we genotyped two INSIG2 SNPs that have been shown previously to be prominently associated with obesity. In the MRC-OB cohort, the variant rs7566605 in INSIG2 enhanced the effect of rs2721 on TG levels (p=0.00117), suggesting an epistatic interaction between the two genes. The SNP rs2721 in INSIG1 alters a putative binding site for the forkhead transcription factor FoxL1. Consequently, we examined the functional effect of the two SNP alleles on transcription. The T allele demonstrates differential binding in electrophoretic mobility shift assays (EMSAs) to nuclear proteins isolated from HepG2 liver cells. In surgical liver biopsy samples, individuals homozygous for the T allele expressed two-fold higher levels of INSIG1 mRNA. These data suggest the T allele of SNP rs2721 modulates the expression levels of INSIG1, possibly by binding FoxL1.

## †250

**Functional characterization of GALNT2, a novel GWAS locus associated with HDL-C.** *A. Edmondson[1], A. Khera[1], S. Khetarpal[1], S. Kathiresan[2], D. Rader[1].* 1) Institute of Translational Medicine and Therapeutics, University of Pennsylvania, Philadelphia, PA; 2) Broad Institute and Massachusetts General Hospital, Boston, MA.

Genome wide association studies (GWAS) for quantitative traits, such as blood lipid concentrations, have successfully identified many genes previously implicated in rare Mendelian disorders of lipoprotein metabolism and have also mapped many novel loci. One such novel locus for plasma high density lipoprotein cholesterol (HDL-C) is on chromosome 1 and contains the GALNT2 gene with the most significant SNP located in the first GALNT2 intron. GALNT2 is involved in the first step of O-linked glycosylation of proteins and was never previously linked to HDL metabolism. To investigate the effect of GALNT2 on plasma HDL-C levels, we created adeno-associated virus serotype 8 (AAV2/8) vectors encoding murine and human GALNT2 driven by the liver-specific thyroxine-binding globulin promoter. We also created an AAV2/8 vector to provide sustained delivery of an shRNA to knock-down mouse GALNT2 driven by the polymerase III U6 promoter. Hepatic overexpression of mGALNT2 and hGALNT2 both significantly decreased HDL-C by approximately 20% at 28 days. Knock-down of endogenous mGALNT2 (88% - 95%) resulted in a dose-dependent increase in HDL-C of 24% - 37% at 28 days. FPLC separation and subsequent analysis of the lipoprotein fractions were consistent with changes in the cholesterol content of the HDL particles without changes in HDL particle size. These results support the hypothesis that GALNT2 is the causal gene at this locus and that changes in hepatic GALNT2 expression are associated with inverse changes in plasma HDL-C levels. The focus of our ongoing studies is to determine the molecular mechanisms by which hepatic GALNT2 expression influences HDL metabolism.

## 251

**The mouse model helps identify the HDL gene from human genome wide association studies.** *M. Leduc[1], R. Korstanje[1], K. Walsh[1], S. Sheehan[1], S. Amend[1], M. Orho-Melander[3], S. Kathiresan[2], B. Paigen[1].* 1) Jackson Lab, Bar Harbor, ME; 2) Massachussetts General Hospital, Boston, MA; 3) University Hospital Malmo, Lund University, Malmo, Sweden.

Plasma high-density lipoprotein cholesterol (HDL) level is a complex trait associated with risk of cardiovascular events. We previously mapped the homologous genes from the human genome wide association studies (GWAS) hits (Kathiresan et al., Nat. Genet., 2008) onto the mouse HDL QTL map and found that 85% of these human loci lay within 20 Mb of HDL QTL peaks in the mouse. In this report, we used the mouse to help choose the responsible gene for 2 newly identified human loci on Chr 11 and 12 that contained multiple genes within the haplotype block indicated by the GWAS hit. These loci were validated in further human GWAS, but the identification of the HDL gene within each multigene locus was challenging due to the lack of available tools in human populations. Both of these loci were homologous to HDL QTL found in the mouse on Chr 5 at 115 Mb and Chr 19 at 12 Mb. The bioinformatics tools used in the mouse include comparison of haplotype differences in the strains used in the QTL studies, amino acid changes in the coding region, significant expression changes between the strains, and other lines of evidence. Using these tools, we were able to identify one highly likely gene within each locus. On Chr 19, the locus contained Fads1, Fads2, Fads3, and Fen1. We identified Fads1 (Fatty Acid Desaturase 1) as the HDL gene because it was highly differentially expressed between the 2 strains that detected the QTL. Fads2 contains one amino acid changing polymorphism (I/V), but this polymorphism is not located in a conserved region of the gene and is not believed to be deleterious. On Chr 5, the locus contained Mvk (Mevalonate Kinase) and Mmab (Methylmalonic Aciduria -Cobalamin Deficiency- cblB Type). Neither gene was differentially expressed among the strains used to detect the Chr 5 HDL QTL, but Mvk possessed one non-synonymous polymorphism in a conserved region that segregated with the high and low allele HDL strains in each cross. However, due to the proximity of Acads (acyl-Coenzyme A dehydrogenase, C-2 to C-3 short chain) in the mouse, we cannot exclude that Acads may also be responsible for the QTL at the Chr 5 locus. In vitro and in vivo experiments are necessary to further ascertain the role of each gene in HDL metabolism. We conclude that the mouse bioinformatics tools are a powerful asset to add evidence to the human GWAS and identify complex trait genes.

## 252

**Genome-wide loci associated with plasma Lp(a), relationship to Kr IV-type 2 repeats, and links to thrombosis.** *D.I. Chasman[1,2], G. Paré[1,2], R.Y.L. Zee[1,2], S. Badola[3], A.N. Parker[3], P.M Ridker[1,2].* 1) Preventive Medicine, Brigham & Women's Hospital, Boston, MA; 2) Harvard Medical School Boston, MA; 3) Amgen, Inc. Cambridge, MA.

Lipoprotein(a) [Lp(a)] is a plasma lipoprotein fraction having composition similar to LDL but with the addition of a single molecule of apolipoprotein(a) [apo(a)] covalently bound to apolipoprotein B, which is common to both. While the function of Lp(a) is uncertain, plasma Lp(a) levels have been associated with incident cardiovascular disease (CVD), possibly due to cholesterol bearing aspects of Lp(a) particles but possibly also due to a role in thrombosis implied by sequence homology between apo(a) and plasminogen. We performed a genome-wide scan for plasma Lp(a) levels among 22,054 participants in the Women's Genome Health Study (WGHS) across 339,913 SNPs. Ninety-three SNPs had genome-wide significance for Lp(a); all mapped to a 1.4Mb segment at chromosome 6q25.3-6q26 (160.3-161.7 Mb) including the *LPA* gene encoding apo(a) as well as the *LPA2*, *PLG*, *IGF2R*, *MAP3K4*, *SLC22A1*, and *SLC22A2* genes. SNPs with known strong association with LDL cholesterol, for example rs676210 in the *APOB* gene and rs646776 near the *SORT1* gene, were not associated with Lp(a) even at nominal significance. The strongest association was in the *LPA* gene with rs3798220, a non-synonymous SNP encoding the Ile4399Met substitution that was previously described for association with both Lp(a) levels and incident cardiovascular disease. The associated risk of CVD could be offset by aspirin, an anti-thrombotic agent (Atherosclerosis 203:371). Exploring variation linked to rs3798220, we find that haplotypes of *LPA* containing the minor allele (MAF=1.8%) predominantly encode 17 Kringle (Kr) IV-type 2 repeats and have higher Lp(a) levels (median ~205 mg/dl) than the already elevated levels of haplotypes with 17 Kr IV-type 2 repeats but the major allele (median ~125 mg/dl). Thirty percent of the variance in plasma Lp(a) could be explained by a non-redundant subset of 42 SNPs from 6q25.3-6q26. A second genome-wide scan in the WGHS with Lp(a) levels adjusted for these SNPs identified variation with genome-wide significance solely at the *APOH* gene (rs1801689) encoding beta-2-glycoprotein I. Like apo(a), beta-2-glycoprotein I has been implicated not only in lipid metabolism (binding to phospholipids), but also in thrombosis based on the presence of 5 complement control modules (sushi and sushi_2) in its amino acid sequence. Thus, the identification of *APOH* as a determinant of plasma Lp(a) levels strengthens dual functions of Lp(a) in lipid metabolism and thrombosis.

† Trainee Award Finalist

## 253

**Identification of novel genetic variants associated with coronary artery disease by combined analysis of genome-wide variability and gene expression.** *T. Zeller[1], A. Schillert[2], P. Wild[1], S. Szymczak[2], M. Rotival[3], F. Cambien[3], L. Tiret[3], A. Ziegler[2], S. Blankenberg[1].* 1) Dept. of Medicine II, Medical University Mainz, Mainz, Rheinland-Pfalz, Germany; 2) Insitute of Medical Biometry and Statistics, Medical Universtity Schleswig-Holstein, Universtity at Luebeck, Luebeck, Germany; 3) INSERM UMRS 937, Pierre and Marie Curie University (UPMC) and Medical School, Paris, France.

Identifying genes that influence common human diseases is one of the most important efforts undertaken in the field of genetics research. The use of genome-wide association studies (GWAS) has led to the discovery of genetic variants that associate with common diseases. Nevertheless, understanding of the effect of genetic variations, and thus the biological basis is still incomplete. Gene expression is a quantitative phenotype linked to genetic variants and can be affected by these variants. Investigating both genome-wide association and expression studies may enhance our understanding of genetics in complex diseases and facilitate the characterization of the biological basis. To gain more insight into the genetics of coronary artery disease (CAD), we conducted a GWAS of CAD (n=5,030) in combination with analysis of the monocytes transcriptome, a key cell in atherosclerosis and inflammation and thus CAD. First, 603,066 SNPs were tested for association to CAD resulting in 1,438 associated SNPs ($p<10^{-3}$). Second, an association analysis of these SNPs and 12,808 genes expressed in monocytes resulted in 34 SNPs associated with the expression of 25 different transcripts ($p<10^{-8}$ using ANOVA). Further steps of analyses included replication of significant findings. The most interesting results were 2 SNPs located within the *LIPA* gene found to be strongly associated with the expression of *LIPA* itself ($p=2.5x10^{-102}$) and *CACYBP* ($p=1.2x10^{-10}$), respectively. *LIPA*, encoding a lysosomal acid lipase is involved in cholesterol metabolism and hydrolysis of cholesteryl esters and triglycerides. Deficiency in LIPA activity results in cholesterol ester storage disease and a role of LIPA in the reduction of atherosclerotic plaques has been described. *CACYBP* encodes a calcyclin binding protein involved in $Ca^{2+}$dependent ubiquitination and proteosomal degradation and plays a role in cardiomyogenic differentiation and protection during hypoxia/reoxygenation. Further results include an association of *TRPV6* expression with a SNP located directly within *TRPV6* ($p=1.4x10^{-8}$) and a SNP located within *EPHB6* ($p=8.8x10^{-9}$). *TRPV6* encodes a transient receptor potential cation channel, and members of the TRP family have been associated with the development of cardiovascular diseases. This study demonstrates with the example of CAD that combining genome-wide association and gene expression studies will lead to a more complete biological picture of genetics of common diseases.

## 254

**Genome-wide meta-analysis in 25,330 individuals identifies multiple loci influencing resting heart rate.** *P.I.W. de Bakker[1,2], M. Eijgelsheim[3], C. Newton-Cheh[2,4,5], N. Sotoodehnia[6,7], M. Müller[8,9,10], A.C. Morrison[11], A.V. Smith[12], A. Isaacs[3], P. Navarro[13], K.V. Tasarov[14], P.P. Pramstaller[15], S.B. Felix[16], V. Gudnason[12], A. Pfeufer[17], S.R. Heckbert[7], B.H.Ch. Stricker[3], E. Boerwinkle[18], C.J. O'Donnell[5] on behalf of the RRGEN Working Group of the CHARGE Consortium.* 1) Division of Genetics, Dept Med, Brigham & Women's Hosp, Harvard Med Sch, Boston, MA; 2) Prog Medical and Population Genetics, Broad Institute, Cambridge, MA; 3) Dept Epidemiology, Erasmus Med Ctr, Rotterdam, The Netherlands; 4) Ctr for Human Genetic Res, Cardiovascular Res Ctr, Mass Gen Hosp, Boston, MA; 5) Natl Heart, Lung and Blood Inst's Framingham Heart Study, Framingham, MA; 6) Division of Cardiology, Dept Med, Univ of Washington School of Med, Seattle, WA; 7) Cardiovascular Health Res Unit, Univ of Washington, Seattle, WA; 8) Dept Med I, Klinikum Grosshadern, Münich, Germany; 9) Inst Epidemiology, Helmholtz Ctr Münich, Münich, Germany; 10) Institute of Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Münich, Germany; 11) Human Genetics Ctr, Univ of Texas, Health Science Ctr, Houston, TX; 12) Icelandic Heart Association Res Inst, Kopavogur, Iceland; 13) MRC Human Genetics Unit, Inst Genetics and Molecular Med, Western Gen Hosp, Edinburgh, Scotland; 14) Lab Cardiovascular Sci, Intramural Res Prog, Natl Inst on Aging, Natl Inst Hlth, Baltimore, MD; 15) Institute of Genetic Medicine, European Academy Bozen/Bolzano (EURAC), Bolzano, Italy; 16) Dept Internal Med B, Ernst-Moritz-Arndt-University, Greifswald, Germany; 17) Inst Human Genetics, Helmholtz Center Münich, Münich, Germany; 18) Human Genetics Ctr, Univ Texas, Hlth Sci Ctr, Houston, TX.

**Background.** Higher resting heart rate is associated with increased cardiovascular mortality and sudden death, independent of traditional risk factors. Though genetic factors play a substantial role in population variation of resting heart rate, accounting for up to 30% of heart rate variation, little is known about specific loci involved.

**Methods.** We performed a meta-analysis of eight genome-wide studies estimating the association between age, sex, and body mass adjusted RR interval (inverse heart rate) and 2.5 million SNPs genotyped in individuals of European descent. Individuals were excluded for prevalent myocardial infarction, heart failure, atrial fibrillation, 2nd or 3rd degree atrio-ventricular block, current use of heart-rate-altering medications (defined as use of β-blockers, non-dihydropyridine calcium antagonists or digoxin), or heart rate <50 or >100 beats per minute. Replication evidence for selected loci was collected from five independent cohorts.

**Results.** Meta-analysis of genome-wide association study results from a total of 25,330 individuals and replication results from 9,299 individuals revealed four novel loci with genome-wide significance: the *GNB4* region at 3q26 (P=4.3e-08), the *SLC35F1-PLN* region at 6q22 (P=2.5e-09), the *BCAT1* region at 12p12 (P=2.5e-09), the *MYH6* region at 14q12 (P=4.8e-12) and an additional locus near *GJA1* at 6q22 recently reported in Asians (P=8.7e-15).

**Conclusion.** In this genome-wide association study for resting heart rate, four novel associated loci have been identified and one recently reported locus was replicated. These findings advance our understanding of the genetic determinants of heart rate, which can now be tested for an influence on cardiovascular mortality.

## 255

**Genome-wide association study of homocysteine levels in Filipinos identifies a new locus (*CPS1*) and evidence for a stronger *MTHFR* effect in young adults than in their mothers.** *L.A. Lange[1], A.F. Marvelle[1], D. Croteau-Chonka[1], C. Kuzawa[2], T.W. McDade[2], Y. Li[3], S. Levy[4], M. Daniels[5], J. Borja[6], E.M. Lange[1,7], L.S. Adair[5], K.L. Mohlke[1].* 1) Dept. of Genetics, Univ. of North Carolina, Chapel Hill, NC; 2) Dept. of Anthropology, Northwestern Univ., Evanston, IL; 3) Dept. of Biostatistics, Univ. of Michigan, Ann Arbor, MI; 4) Dept. of Biomedical Informatics, Vanderbilt Univ., Nashville, TN; 5) Dept. of Nutrition, Univ. of North Carolina, Chapel Hill, NC; 6) Office of Pop. Studies, Univ. of San Carlos, Cebu City, Philippines; 7) Dept. of Biostatistics, Univ. of North Carolina, Chapel Hill, NC.

Increased homocysteine (Hcy) levels are associated with cardiovascular disease and may induce vascular damage, a pre-cursor for atherosclerosis. While the estimated heritability for Hcy is high (51-63%), the only gene established to be associated with Hcy is methylene tetrahydrofolate reductase (*MTHFR*), and much of the inherited variability remains to be explained. We performed a genome-wide association study for Hcy levels in 1,787 unrelated Filipino women from the Cebu Longitudinal Health and Nutrition Survey (CLHNS). A total of 2,073,674 genotyped or imputed SNPs were tested for association with log-transformed Hcy levels. All 21 SNPs with $p < 5x10^{-8}$ were located in or near the carbamoyl phosphate synthetase 1 (*CPS1*) gene. The strongest associated SNP, rs7422339, encodes Thr1405Asn ($p = 2.2 \times 10^{-12}$) and explains 3% of the variability in log-Hcy levels. In comparison, the widely studied *MTHFR* C677T SNP rs1801133 was less significant ($p = 8.0 \times 10^{-7}$) and explained only 0.6% of the variance in log-Hcy. *CPS1* is a novel locus associated with Hcy; the encoded protein catalyzes the first committed step of the hepatic urea cycle, and this variant has been reported previously to be associated with fibrinogen level, nitric oxide level, necrotizing enterocolitis, and pulmonary hypertension. We then genotyped the *CPS1* and *MTHFR* SNPs in n=1,679 young adult (aged 20-22 years) CLHNS offspring. The *MTHFR* C677T SNP was strongly associated with Hcy ($p = 6.8 \times 10^{-22}$) and explained approximately 5.2% of the total variability in log-Hcy, in contrast to the *CPS1* variant, which was significant in females only ($p = 0.11$ in all; $p=0.0087$ in females). A combined analysis using general linear mixed models to account for the correlation among mother-child pairs confirmed that the *MTHFR* variant is more strongly associated with Hcy in the offspring. Furthermore, while there was evidence for a positive synergistic effect between these SNPs in the offspring (interaction $p = 0.0046$), there was no significant evidence for an interaction in the mothers ($p = 0.55$). Together, these data suggest that *CPS1* is a novel locus for Hcy levels and that genetic effects on Hcy may differ across developmental stages, possibly due to the influence of lifelong environmental exposures.

## 256

**A meta-analysis of eight hematological parameters identifies 22 associated loci and extensive disease pleiotropy on chromosome 12q24.** *N. Soranzo[1,2], C. Gieger[3] for the HaemGen Consortium.* 1) Wellcome Trust Sanger Inst, Hinxton, United Kingdom; 2) King's College London, St Thomas' Hospital Campus, Lambeth Palace Rd, London SE1 7EH; 3) Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85764 Neuherberg.

The number and volume of cells in the blood are the most commonly measured blood parameters in clinical practice and impact on a wide range of disorders including cancer, cardiovascular, metabolic, infectious and immune disorders. We carried out a meta-analysis of eight different hematological parameters in 13,943 samples from six European population-based studies. We identified 22 genetic loci associated with hematological parameters, including 6 associated with red blood traits (e.g. the *HBS1L-MYB* region on 6q23-q24 (rs9402686, $P = 7.42 \times 10^{-42}$), the C282Y amino acid change in *HFE* at 6p21.3 (rs1800562, $P = 1.44 \times 10^{-23}$) and *TMPRSS6* ($P = 9.5 \times 10^{-10}$)), one associated with white blood cell parameters (*GSDM1/ORMDL3*, $P = 9.41 \times 10^{-9}$) and 15 associated with platelet counts and volume (eg. *BET1L*, $P = 1.3 \times 10^{-14}$, *BCL2L7P1/BAK1*, $P = 3.7 \times 10^{-10}$ and *JMJD1C*, $P = 3.30 \times 10^{-21}$). The novel loci identify known and novel regulators of hematopoiesis, underscoring the high level of connectivity in stem cell fate determination. We further used whole-genome expression data in stem-cell derived blood lineages to study patterns of gene expression, and identified the first platelet eQTL near *BET1L* ($P = 3.1 \times 10^{-5}$). We finally tested associations with coronary artery disease in 10,838 cases and 12,131 controls and identified one long-range haplotype at 12q24 carrying 10 SNPs associated with this disease (OR = 1.189, 95% CI = 1.136-1.244, $P = 9.3 \times 10^{-14}$ for the best SNP). We show that the haplotype, which also contains known risk loci for type 1 diabetes, hypertension and celiac disease, underlies a locus of high disease pleiotropy. Using evolutionary analyses we show that the haplotype has been spread by a selective sweep which began ~4,300 years ago and was specific to European and nearby populations.

## 257

**Rare Structural Variants found in Attention-Deficit Hyperactivity Disorder are Preferentially Associated with Neurodevelopmental Genes.** *X. Gai[1], J. Elia[2,3], H. Xie[1], J.C. Perin[1], E. Geiger[4], J.T. Glessner[5], M. D'arcy[1], R. deBerardinis[2], E. Frackleton[5], C. Kim[5], F. Lantieri[4], B.M. Muganga[1], L. Wang[1], T. Takeda[2], E.F. Rappaport[6], W. Berrettini[3], M. Devoto[4,7,8,9], T.H. Shaikh[4,7], H. Hakonarson[5,7,10], P.S. White[1,7,11].* 1) Center for Biomedical Informatics, Children's Hospital of Philadelphia, Philadelphia, Philadelphia, PA, 19104, USA; 2) Department of Child and Adolescent Psychiatry, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA; 3) Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA, 19104, USA; 4) Division of Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA; 5) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA; 6) Joseph Stokes, Jr. Research Institute, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA; 7) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA, 19104, USA; 8) Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA, 19104, USA; 9) Dipartimento di Medicina Sperimentale, University La Sapienza, 00185 Rome, Italy; 10) Division of Pulmonary Medicine, The Children's Hospital of Philadelphia, 19104, USA; 11) Division of Oncology, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA.

Attention-Deficit/Hyperactivity Disorder (ADHD) is a common and highly heritable disorder, but specific genetic factors underlying risk remain elusive. To assess the role of structural variation in ADHD, we identified 222 inherited copy number variations (CNVs) within 335 ADHD patients and their parents that were not detected in 2,026 unrelated healthy individuals. The CNV-associated gene set was significantly enriched for genes reported as candidates in studies of autism, schizophrenia, and Tourette syndrome, including A2BP1, AUTS2, CNTNAP2, and IMMP2L. The ADHD CNV gene set was significantly enriched for genes known to be important for psychological and neurological functions, including learning, behavior, synaptic transmission, and central nervous system development. Phenotypes observed in mice from null or transgenic models of genes in the ADHD CNV gene set were significantly enriched for behavioral traits consistent with ADHD symptoms. Four independent deletions were located within the protein tyrosine phosphatase gene PTPRD recently implicated as a candidate gene for restless legs syndrome, which frequently presents with ADHD. A deletion within the glutamate receptor gene GRM5 was found in an affected parent and all three affected offspring whose ADHD phenotypes closely resembled those of the GRM5 null mouse. Together, these results suggest that rare inherited structural variations play an important role in ADHD development and indicate a set of putative candidate genes for further study in the etiology of ADHD.

## 258

**Microduplications of 16p11.2 are Associated with Schizophrenia.** *S.E. McCarthy[1], V. Makarov[1], G. Kirov[2], A. Addington[3], J. McClellan[4], R.A. Kumar[1], F. McMahon[5], A.K. Malhotra[6], J.B. Potash[7], N.R Mendell[8], M.C. O'Donovan[2], T.H. Shaikh[9], E. Susser[10], L.E. DeLisi[11], P.F. Sullivan[12], C.K. Deutsch[13], J. Rapoport[3], D.L. Levy[14], M.C. King[5], J. Sebat[1], Wellcome Trust Case Control Consortium.* 1) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA; 2) Department of Psychological Medicine, School of Medicine, Cardiff University, Cardiff, UK; 3) Child Psychiatry Branch, National Institute for Mental Health, National Institutes of Health, Bethesda, Maryland, USA; 4) Department of Psychiatry, University of Washington, Seattle, Washington, USA; 5) Genetic Basis of Mood and Anxiety Disorders Unit, National Institute for Mental Health, National Institutes of Health, Bethesda, Maryland, USA; 6) Department of Psychiatry Research, The Zucker Hillside Hospital, Glen Oaks, New York, USA; 7) Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, Maryland, USA; 8) Department of Applied Mathematics and Statistics, State University of New York, Stony Brook, New York, USA; 9) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 10) College of Physicians and Surgeons of Columbia University, Columbia University, New York, USA; 11) Department of Psychiatry, Harvard Medical School, Boston, Massachusetts, USA; 12) Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA; 13) Eunice Kennedy Shriver Center, University of Massachusetts Medical School, Waltham, Massachusetts, USA; 14) Psychology Research Laboratory, McLean Hospital, Belmont, Massachusetts, USA.

The evidence implicating rare copy number variants (rCNVs) in the etiology of psychiatric disease is rapidly accumulating. We and others have shown that the genomewide burden of rCNVs is significantly increased (1.15-3 fold) in schizophrenia cases compared to healthy controls. Furthermore, rare recurrent CNVs at 1q21, 15q11, 15q13 and 22q11 significantly increase the risk of schizophrenia. Here, we present the association of rare 16p11.2 microduplications with schizophrenia in a cohort of 1921 cases and 4062 controls the 16p11.2 microduplication was observed ($P = 1.4x10^{-5}$, $OR = 25.4$). This association was replicated using an independent sample of 2645 schizophrenia cases and 2420 healthy controls ($P = 0.022$, $OR = 8.3$) from the Genetic Association Information Network (GAIN) supported "Genome-Wide Association Study of Schizophrenia" (*phs000021.v2.p1*). In a meta-analysis of psychiatric disorders, the 16p11.2 microduplication was consistently observed at a significantly greater frequency in schizophrenia and autism cases than in corresponding controls, 0.3% vs. 0.03% ($P = 5x10^{-7}$) and 0.46% vs. 0.02% ($P = 1.9x10^{-7}$) respectively. The microduplication was also observed at a higher frequency in bipolar disorder (0.1% vs 0.03%, $P = 0.055$). In contrast, the reciprocal 16p11.2 microdeletion increased the risk of autism specifically (0.78% vs. 0.02%, $OR = 38.7$, $P = 2.3x10^{-13}$). Furthermore, analysis of quantitative patient clinical data showed that standardized head circumference based on the Centers for Disease Control growth charts was significantly larger in microdeletion carriers and moderately smaller in microduplication carriers (1.25 vs. -0.28, $P = 0.0007$). Our findings add 16p11.2 to the growing list of rCNV hotspots that considerably increase the risk of schizophrenia. The spectrum of diseases associated with 16p11.2 is consistent with phenotypic heterogeneity at other schizophrenia risk loci and suggests that common neurobiological pathways maybe shared between phenotypes associated with 16p11.2 rearrangements. Finally the association of 16p11.2 rearrangements with head circumference point to a potential genetic basis for early brain overgrowth in autism and observations of smaller brain volume in schizophrenia.

## 259

**Large CNVs in an isolated population show enrichment in neuropsychiatric diagnoses including an enrichment of three schizophrenia associated CNVs overlapping developmental genes.** *O.P.H. Pietilainen[1,2,3], T. Paunio[2,4], A. Tuulio-Henriksson[5,6], J. Suvisaari[5], J. Haukka[5], T. Varilo[2,3], K. Rehnström[2,3], E. Jakkula[2], J. Wedenoja[2,3], A. Loukola[2], J. Suokas[4,5], L. Häkkinen[5], S. Ripatti[2], S. Ala-Mello[7], M.-R. Jarvelin[8,9,10], M. Isohanni[11], J. Lönnqvist[4,5], H. Stefansson[12], N.B. Freimer[13], L. Peltonen[1,2,3,14].* 1) The Wllcome Trust Sanger Institute, Cambridge, Cambridge, United Kingdom; 2) FIMM, Institute for Molecular Medicine Finland, Helsinki, Finland; 3) University of Helsinki, Department of Medical Genetics, Helsinki, Finland; 4) University of Helsinki and Helsinki University Central Hospital, Department of Psychiatry, Helsinki, Finland; 5) National Institute of Health and Welfare, Department of Mental Health and Alcohol Research, Helsinki, Finland; 6) University of Helsinki, Department of Psychology, Helsinki, Finland; 7) Helsinki University Central Hospital, Department of Clinical Genetics, Helsinki, Finland; 8) Department of Epidemiology and Public Health, Imperial College London, St. Mary's Campus, Norfolk Place, London, UK; 9) Department of Child and Adolescent Health, National Public Health Institute (KTL), Oulu, Finland; 10) Institute of Health Sciences and Biocenter Oulu, University of Oulu, Oulu, Finland; 11) Department of Psychiatry, Institute of Clinical Medicine, University of Oulu, Finland; 12) deCODE genetics, 101 Reykjavik, Iceland; 13) Center for Neurobehavioral Genetics, Gonda Center, University of California Los Angeles, Los Angeles, California, USA; 14) The Broad Institute of MIT and Harvard University, Cambridge, Ma, USA.

Emerging evidence from recent studies has linked Copy number variations (CNVs) with cognition and variable neurpsychiatirc disorders such as schizophrenia and autism. We applied two partially distinct approaches to address the impact of large CNVs in neurodevelopment and functioning. In schizophrenia the role of CNVs has been to date emphasized merely in sporadic cases. Consequently, we analyzed a special Finnish study sample of 196 schizophrenia cases and 199 controls emerging from a "high risk region of Finland" with enrichment of familial form of schizophrenia. Our specific interest was to identify moderate to high risk CNV alleles that as the result of recent population bottle necks have become enriched in this founder population. We discovered three large (> 50 kb) CNV alleles on chromosomes 9p24.3, 17p13.3, and 22q11.22 significantly enriched to isolate schizophrenia cases (p<.05). Both permutation and analysis of ~2800 SNPs with similar frequencies as the observed CNVs suggested the associations not to occur due to chance alone (p<.009). After screening additional 5278 Finnish population samples, as well as 2601 schizophrenia cases and 30091 controls of European origin, the three CNVs were found to be significantly enriched in the Finnish population generally (p <1.9*10-9), but particularly in the isolate (p <2.0*10-4). Within the isolate, the CNVs were approximately 3-times more frequent in families with schizophrenia compared to the general population of the isolate (p<.001). Each CNV was ~1.5-times more frequent among family members with any mental disorder as among unaffected individuals and accompanied with increased penetrance within the carrier families. We identified three individuals homozygous for the 22q11.22 deletion all of whom had cognitive disability. The CNVs harbor three developmental genes: DOCK8, ABR, and TOP3B, not previously associated with schizophrenia. In the second approach we used a "reverse genomics" and scanned systematically >1 Mb CNVs in a Finnish population sample of 4932 individuals. The sample was drawn from a birth cohort of all in 1966 born individuals in the two most north provinces of Finland and was genotyped using the Illumina HumanHap 370 platform. We identified 118 individuals with >1 Mb CNV and total of 30%; of them had neurological or psychiatric diagnosis. We are on our way studying their impact in neurodevelopment and functioning using over 100 relevant phenotypes collected from the participants.

## 260

**Copy Number Variation Discovery in Autism Spectrum Disorder.** *A.T. Pagnamenta[1], D. Pinto[2], H. Khan[2], V.J. Vieland[3], A.S. Le-Couteur[4], S.W. Scherer[2], A.P. Monaco[1], The Autism Genome Project Consortium.* 1) Wellcome Trust Centre for Human Genetics, Oxford University, Oxford, UK; 2) The Centre for Applied Genomics, Hospital for Sick Children, Toronto, Canada; 3) Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and The Ohio State University, Columbus, Ohio, USA; 4) Child and Adolescent Mental Health, Institute of Health and Society, Newcastle University, Sir James Spence Institute, Newcastle upon Tyne, UK.

Autism spectrum disorders (ASDs) are a genetically complex group of neurodevelopmental disorders with incidence of ~0.6%. Despite the high heritability indicated by twin studies, cytogenetic abnormalities and coding mutations in genes (often those involved with synaptic connections) have been detected only in a small minority of families. However, the increasing resolution of array-based studies is now facilitating detection of additional genomic variants associated with ASD susceptibility. The international Autism Genome Project (AGP) consortium (http://autismgenome.org) has carried out whole-genome analysis, using the 1M Infinium SNP array (Illumina). Copy number variation (CNV) calling has been undertaken in a cohort of ~1172 ASD family trios using QuantiSNP and iPattern algorithms. Merging these two CNV predictions gave an average of ~15 validated CNVs per proband, with a mean size of 110 kb. CNV findings were further confirmed using a combination of experimental validation, genotype information and a third CNV calling algorithm. Multiple variants are present at *DOCK8*, *SHANK2* and other novel loci, highlighting previously implicated pathways. Integration of CNV data with SNP association results will help prioritize follow-up studies. Over 40% of the AGP families are multiplex, reflecting earlier interest in linkage analysis. Therefore, ongoing analysis is focused on inherited CNVs as well as *de novo* events. As an example, we present detailed molecular characterisation of a rare deletion of the *CDH8* gene on chr16q21. This region was initially identified through linkage analysis in the AGP sample with a PPL (posterior probability of linkage) of 87%. PCR amplification across the breakpoints was used to determine segregation of this deletion. The 1,642,256 bp loss had been transmitted from the mother to 3/3 affected and 0/4 unaffected siblings, consistent with the involvement of cadherin 8 in the ASD phenotype seen in this family. RT-PCR results indicate that multiple *CDH8* isoforms are expressed in various brain regions, with highest expression in parietal cortex.

In conclusion, rare variants collectively contribute substantially to autism risk. There is evidence of overlapping genetic etiologies between ASD and other psychiatric disorders. The example shown here also highlights the importance of segregation analysis of rare CNVs in large well-phenotyped families to complement case-control style studies.

## 261

**Copy-number variation reveals both known and new susceptibility loci for autism spectrum disorders in a population isolate.** *K. Rehnström[1,2,3], H. Kilpinen[1,2,4], E. Jakkula[1,2], E. Gaál[1,2], T. Varilo[1,2], L. von Wendt[5], I. Hovatta[4], L. Peltonen[1,2,3,6,7].* 1) Institute for Molecular Medicine Finland (FIMM), Helsinki, Finland; 2) Unit of Public Health Genomics, National Institute for Health and Welfare, Helsinki, Finland; 3) Department of Medical Genetics, University of Helsinki, Helsinki, Finland; 4) Research Program of Molecular Neurology, University of Helsinki, Helsinki, Finland; 5) Unit of Child Neurology, Hospital for Children and Adolescents, Helsinki, Finland; 6) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 7) The Broad Institute, Cambridge, USA.

Autism spectrum disorders (ASDs) are early onset neuropsychiatric disorders with varying severity. They are characterized by symptoms in three domains including communication, social interaction and a behavioral domain. Recent genome-wide association studies have revealed common genetic risk variants, but other studies have provided compelling evidence for the role of rare genetic variants in the etiology of ASDs. One form of rare variation in ASDs is copy-number variants (CNVs), which have repeatedly been associated with these disorders. Here, we have used a founder population, where genetic risk factors for ASDs could be more homogenous than in large, outbred populations, to assess for the role of CNVs in the etiology of ASDs in Finnish families. We genotyped a total of 128 individuals with autism and 29 individuals with AS using the Illumina HumanHap 300 or 550 BeadChip. We used PennCNV to identify CNVs. To exclude rare, normal CNVs present in the general population, we used a set of approximately 5400 Finnish individuals from a population cohort as controls. We defined CNV-regions of interest (CNV-ROIs) as CNVs spanning at least one gene, covered by ≥10 SNPs, which were identified in cases but no more than 50% overlap with a CNV reported in the database of genomic variants. We also excluded all regions identified in two or more controls in our large Finnish control dataset. In the autism dataset, we identified five large CNVs, all gains, exceeding 1 Mb in size. Three of these amplifications were located on 15q11-13, resulting in a frequency of 2.4% for duplication of 15q11-13 in our Finnish nation-wide autism sample. Two other large duplications, one at 1q42, spanning 2.4Mb, and another at 9q21, spanning 1.7Mb were identified in male probands affected with autism. A total of 27 smaller CNV-ROIs were identified. One deletion spanning NRXN1 was identified, as well as one duplication of 22q13, including SHANK3, both known ASD susceptibility loci. In addition, several novel regions of interest were identified, several of which have been reported in the Autism Chromosomal Rearrangement Database supporting the role of these CNVs in the etiology of ASDs. In the AS dataset, a similar frequency of CNV-ROIs (n= 7) were identified. CNV sizes ranged from 21 to 152 kb. Only one CNV overlapped with CNV-ROIs in the autism dataset. These results suggest that CNVs also play a role in the etiology of AS, but the loci are distinct from those predisposing for autism.

## 262

**Annotation of neuronal enhancers in candidate loci for Autism spectrum disorders.** *M. Bucan[1, 2], R. Lui[1], S. Wadhawan[1], S. Hannanhalli[1,2].* 1) Dept Genetics, Univ Pennsylvania, Philadelphia, PA; 2) Penn Center for Bioinformatics.

Autism spectrum disorders (ASD) comprise a group of highly heritable developmental disorders that manifest in the first 3 years of life and are characterized by impairments in the core areas of communication and reciprocal social interaction, as well as the presence of restricted behaviors, interests and activities. Our findings that the vast majority of CNVs, those that are more prevalent in ASD cases than in controls, were found in intergenic regions of neuronal cell-adhesion genes (Glassner et al., Nature 459, 569-73, 2009; Bucan et al., PLoS Genetics, in press) and that a cluster of ASD-associated common variants map to the 2 Mb CDH9-CDH10 intergenic region (Wang et al., Nature 459, 528-33, 2009), underscores a major challenge in the evaluation of functional consequences of disease risk factors. Our understanding of variation in non-coding, and presumably regulatory regions, is in its infancy, despite predictions that regulatory mutations are more likely to lead to complex diseases and/or quantitative traits. To prioritize these intergenic, presumably regulatory, genomic elements for resequencing and experimental validation, we performed systematic computational analysis of genomic regions surrounding ASD susceptibility loci. Our effort focused on identification of regulatory elements and cis-acting polymorphisms resulting in the disruption or modulation of synaptic development and function. To develop a computational approach for a comprehensive characterization of cis-regulatory elements in genes involved in synaptic development and function, we analyzed multi-species conserved sequences (MCS) in the vicinity of nine presynaptic genes, which are highly and specifically expressed in brain regions, to identify clusters of enriched transcription factor binding sites, i.e. cis-regulatory modules. Sixteen transcription factor binding motifs were over-represented in these MCEs. Based on a combined occurrence for these enriched motifs, MCEs in the vicinity of 107 presynaptic genes were scored and ranked. We then experimentally validated the scoring strategy by showing that 12 of 16 (75%) high scoring MCEs functioned as neuronal enhancers in a cell based assay. This scoring strategy combined with experimental validation has been applied to the CDH9-CDH10 intergenic region (80 MCSs) and genomic regions around over 100 loci implicated in ASD susceptibility in the CNV analysis, leading to a comprehensive annotation of ASD-associated regions.

## 263

**A recurrent 16p12.1 microdeletion modifies neurobehavioral phenotypes.** *S. Girirajan[1], J.A. Rosenfeld[2], P. Siswara[1], F. Antonacci[1], T. Walsh[3], B.C. Ballif[2], S. McCarthy[4], J.M. Kidd[1], G.M. Cooper[1], B. Salbert[5], Y. Lacassie[6], L. deLisi[7], T.H. Shaikh[8], J. Sebat[4], M-C. King[1,3], L.G. Shaffer[2,9], E.E. Eichler[1,10].* 1) Department of Genome Sciences, University of Washington, Seattle, WA. USA; 2) Signature Genomic Laboratories, Spokane, WA, USA; 3) Departments of Medicine and Medical Genetics, University of Washington School of Medicine, Seattle, WA, USA; 4) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA; 5) Geisinger Medical Center, Danville, PA, USA; 6) Division of Genetics, Department of Pediatrics, Louisiana State University Health Sciences Center and Children's Hospital, New Orleans, LA, USA; 7) Department of Psychiatry, New York University, Millhauser Laboratories, New York, New York, USA; 8) Department of Pediatrics and Genetics, University of Pennsylvania, and the Children's Hospital of Philadelphia, Philadelphia, PA, USA; 9) School of Molecular Biosciences, Washington State University, Spokane, WA, USA; 10) Howard Hughes Medical Institute, University of Washington School of Medicine, Seattle, WA.

We report the identification of a novel 560 kbp microdeletion mediated by segmental duplications on 16p12.1. We found a significant enrichment ($p$= 0.001, odds ratio=4.64) of 16p12.1 microdeletion in 20/11393 cases evaluated for mental retardation and developmental delay compared to 4/10566 controls. Clinical evaluation showed a high penetrance but variable expressivity of phenotypes including developmental delay (100%), microcephaly (64%), congenital heart defects (40%), and seizures (50%) in these cases. Nevertheless, dysmorphic evaluation demonstrated consistent facial features including deep-set eyes, broad nasal bridge, downslanting palpebral fissures, smooth philtrum, and low-set ears. Unlike other genomic disorders, most of the probands (9/10) inherited this microdeletion, one case was confirmed *de novo*, and inheritance in 10 cases was unknown. Subsequent parental re-evaluation revealed overt neurobehavioral phenotypes among all parents segregating with the deletion (learning disabilities, depression, bipolar disorders, and seizures) but not in non-segregating parents ($p$=0.01). Among the children with mental retardation and developmental delay, the 16p12.1 microdeletion was the only large pathogenic copy-number variant (CNV) observed in 13 cases. However, 7/20 cases also documented co-occurrence of another pathogenic CNV, a significant excess of "double hits" ($p=5.5 \times 10^{-9}$) compared to that in normal controls (9/2500). In these seven cases with "double hits" the clinical features were more severe, overriding classically described phenotypes for the co-occurring pathogenic CNVs such as dup15q11.2-q11.3, dup22q11.2 and del22q11.2. We propose that 16p12.1 microdeletion predisposes to a variety of neurobehavioral deficits in the general population and that a "two hit" model is required during development to result in a clinical manifestation of mental retardation.

## 264

**SNPs and CNVs in celiac disease: New loci and a new gene-network assisted CNV analysis.** *G. Trynka[1], PC. Dubois[2], J. Romanos[1], A. Zhernakova[3], R. Gwilliam[4], P. Deloukas[4], R. McManus[5], P. Saavalainen[6], DA. Van Heel[2], C. Wijmenga[1,3], L. Franke[1,2].* 1) Gen Dept, UMC Groningen, Groningen, Netherlands; 2) Institute of Cell and Molecular Science, Barts and The London School of Medicine and Dentistry, London, UK; 3) Department of Biomedical Genetics, University Medical Centre Utrecht, Utrecht, The Netherlands; 4) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 5) Departments of Clinical Medicine, Institute of Molecular Medicine, Trinity College Dublin, Dublin, Ireland; 6) Department of Medical Genetics, and Research Program of Molecular Medicine, University of Helsinki, Finland.

Celiac disease (CD) is a common (1% prevalence) small intestinal inflammatory condition induced by dietary protein present in wheat, rye and barley. CD has a considerable heritability and strong association to the HLA locus. In a previous genome-wide association study (GWAS) and its follow-ups we identified 12 risk loci, pointing to altered innate and adaptive immunity pathways in CD. We have now completed an extended GWAS of 4.533 cases and 10.750 controls (from the UK, the Netherlands, Italy and Finland), comprising existing GWAS data (Hap300 / Hap550 platform) and new GWAS data (Human670 / Human1.2M platform). All 12 previously reported CD genes could be replicated and eight new, genome-wide significant loci were identified (of which five contain clear immune-related genes). 77 regions showed association with P<10-5, of which the majority harbour immune-related candidate genes. Replication of these associations is currently underway in an additional ~4500 cases and ~5500 controls. Parallel to the SNP analysis we conducted a CNV association analysis on >10,000 samples (~5000 samples on the Human670 platform, ~5000 samples on the Human1.2M platform). We used an improved TriTyper algorithm (CNVTyper) to identify and impute common deletions and duplications that span as little as a single SNP. We identified over 4000 common multiallelic SNPs (SNPs with extra deletion or duplicaton alleles) on the Hap670 platform, of which 70% map within known CNV regions. The Hap1.2M platform analysis identified over 7500 common multiallelic SNPs (of which 5000 are estimated to reflect CNVs). Through this extensive reference panel and CNV imputation catalogue we could impute CNVs in cases and controls, permitting highly robust association analysis of these variants. In order to identify rare deletions and duplications (as well as to contrast the CNVTyper results with an independent method) we applied PennCNV. To assess their potential role in celiac disease we employed a gene network, assuming that causal structural variants are likely to map within or near genes that are involved in the same gene networks, such as the innate and adaptive immunity pathways. We will present the results on the replication analysis of both SNPs and CNVs in celiac disease.

## 265

**The first map of dispensable regions in the human genome.** *T. Vrijenhoek[1], C. Webber[2], R. Konst[1], M. Steehouwer[1], A. Marques[2], R. Makkinje[1], C. Gilissen[1], M. Rijpkema[3], G. Fernandez[3], H.G. Brunner[1], A. Geurts van Kessel[1], B. Franke[1,4], C.P. Ponting[2], J.A. Veltman[1].* 1) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands; 2) MRC Functional Genomics Unit, University of Oxford, Department of Physiology, Anatomy and Genetics, South Parks Road, Oxford, OX1 3QX, United Kingdom; 3) Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, the Netherlands; 4) Department of Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, the Netherlands.

Recent microarray and resequencing studies show that genomic deletions occur throughout the human genome. These deletions can occur in homozygous state in apparently healthy individuals, indicating a complete loss of part of our genome without apparent phenotypic consequences. The aim of the current study is to create a first map of validated homozygous deletions in healthy individuals. This map will give insight into the functional and dispensable parts of our genome and can ultimately result in the definition of the minimal human genome.

Genome-wide copy number analysis was performed in 600 DNA samples of healthy Dutch volunteers using the Affymetrix SNP 6.0 array. After quality control, we identified over 2,000 homozygous deletions of more than 10kb in size, distributed over 75 distinct regions. Approximately 65% of these homozygous deletions are recurrent, of which one third occurs frequently (in ≥ 5% of the samples). After validation by PCR we define 3.7 Mb of genome sequence that is subject to homozygous deletions in these subjects. The regions contain 39 protein-coding genes and 175 non-coding RNA (ncRNA) loci. Constrained ncRNAs and other functional sequences are depleted in homozygous deletions e.g. a 24% depletion in phastCons sequence ($p = 0.075$) showing that a complete loss of functional sequence is selected against. In addition, genes encompassed by homozygous deletions are significantly smaller ($p = 0.007$), and contain fewer introns ($p = 0.007$), which are characteristics of 'environmental' (e.g. olfactory and immunity) genes.

Our data support the notion that deletion alleles preferentially segregate in the human population only when they do not encompass functional elements, or else those that are less essential for viability. In conclusion, our data show that at least 0.1% of our genome is dispensable without apparent deleterious effect, thus providing a first indication of the size of the minimal human genome.

**266**

**Towards a CNV atlas for human development.** *C.L. Martin[1], E.B. Kaminsky[1], B. Bunke[1], D. Kunig[1], A. DeLorenzo[1], V. Kaul[1], K. Huang[1], D. Saul[2], D.L. Pickering[3], D.M. Golden[3], W. Sanger[3], S. Aradhya[2], D.H. Ledbetter[1].* 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 2) GeneDx, Gaithersburg, MD; 3) Human Genetics Laboratory, University of Nebraska Medical Center, Omaha, NE.

Genome-wide oligonucleotide microarrays have been successfully implemented in clinical cytogenetic diagnostics. These arrays have a 10-100 fold improvement in resolution compared to G-banding analysis. Clinical interpretation of the chromosomal imbalances detected by arrays in individuals with abnormal phenotypes has been aided by studies of copy number variation in normal populations. We have analyzed data from 8,738 whole-genome arrays from three clinical diagnostic laboratories that are members of the International Standard Cytogenomic Array (ISCA) consortium. Goals of this consortium include the standardization of cytogenetic array testing and deposition of data into a public database for use by the clinical and research communities. In accordance with ISCA standards, arrays were custom designed with genome-wide backbone coverage in addition to high density coverage of known clinically relevant regions and genes. Array analysis was carried out on individuals with phenotypes including mental retardation, autism, developmental delay, dysmorphic features, and/or multiple congenital anomalies. Of the 3,084 imbalances detected within the reporting criteria, we found pathogenic abnormalities in 53.5% (15.9% of patients) based on known clinically relevant regions, gene content, inheritance, and size. We observed many of the recently described recurrent disorders such as 16p11 microdeletions (38 cases) and 17q12 imbalances (5 deletions and 7 duplications). For 23.4% of the imbalances (7.6% of patients), the abnormality was of uncertain clinical significance at this time. For example, duplications of 16p11.2 were considered uncertain since evidence is still emerging for their association with an increased risk of autism. As expected, many common, known benign copy number variants were identified, including imbalances within segmental duplications, such as the *DEFB* gene cluster on 8p23.1. Of the 881 imbalances with known inheritance, 347 imbalances (39.4%) were determined to be *de novo* and most likely pathogenic. Of the inherited cases, many were inherited from a similarly affected parent or a parent carrying the balanced form of the rearrangement. Comparison of data such as these generated from clinical testing of individuals with abnormal phenotypes to that generated in normal populations will help to establish a gene dosage map of copy number variation for normal and abnormal human development.

## 267

**Autoimmune diseases molecular interaction network identifies novel candidate genes.** *S. Karamohamed[1], G. Pirela[1], D.L Nicolae[1,2], T.C. Gilliam[1].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Medicine and Department of Statistics, University of Chicago, Chicago, IL.

Purpose: Our hypothesis is that disease genes, such as those for autoimmune diseases (AD), map proximal to one another in a molecular interaction network. Thus, identifying a network of AD genes will revel new closely related genes and might provide clues to the AD genes pathways and mechanisms. The network would help understand the functional relationships among AD genes and provide novel candidate genes. Methods: We used nine previously reported genes that connect six correlated AD; Crohn's (CD), celiac (CeD), type-1 diabetes (T1D), systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), and multiple sclerosis (MS). Pathway, molecular and biological network interactions between the seed genes were explored using Ingenuity Pathway Analysis (IPA) software. These seed genes were overlaid onto a global molecular network developed by IPA. Networks of these seed genes were then algorithmically generated based on their connectivity. The network, up to thirty five molecules, will harbor the maximum number of seed genes and connecting genes "candidate genes" that helps bringing maximum number of seed genes together in a single network. We examined the allelic association of the network candidate genes in previously published GWAS data sets; MS and SLE data sets were available from NCBI dbGAP and CD, RA, and T1D were from The Wellcome Trust Case Control Consortium. CeD GWAS and summary statistics were kindly shared by Dr. Van Heel, D.. Results and Conclusions We present an integrated approach that combines molecular and biological network for a set of disease related genes to infer other genes causing disease. Using IPA, we developed a single AD-network harboring twenty-six candidate genes that connects nine previously identified AD-genes. AD-network has pointed out candidate genes previously reported in AD, and also pointed to novel locus. For example, strong accumulative evidence supports the association of a novel gene with CD, CeD, and T1D. In this gene, two intronic SNPs from different linkage disequilibrium (LD) blocks were associated with CD and CeD. A SNP in a separate LD block reveled association to T1D. This finding illustrate that the use of molecular interaction networks can be effective in identifying risk loci. Furthermore, the network approach provides a unique insight into gene-gene interactions in terms of their contribution to one or more disease/phenotype.

## 268

**The genetic architecture of autoimmunity.** *C. Cotsapas[1,2], B. Voight[1,2], K. Lage[1], E. Rossin[1,2], B. Neale[1,2], S. Rich[3], D. Hafler[2,4], M. Daly[1,2] on behalf of the FoCIS Network of Consortia.* 1) CHGR, Mass Gen Hosp, Boston, MA; 2) MPG, Broad Institute, Cambridge MA; 3) Center for Public Health, University of Virginia, Charlottesville VA; 4) Dept of Neurology, Brigham and Women's Hospital, Boston MA.

Recent genome-wide association (GWA) studies have identified numerous replicable genetic associations influencing risk of common autoimmune/inflammatory (AI) diseases. Moreover, several loci have been independently identified to influence risk to more than one AI disease, suggesting commonality in the processes underlying disease progression. Here, we compare 48 AI-associated loci across GWA studies of six AI diseases and find that 21 of these show statistically convincing evidence for association to multiple diseases. We further show that patterns of disease association group loci into functional clusters suggestive of discrete molecular mechanisms. We validate these observations by investigating 6 predicted associations to MS in an independent cohort and finding strong evidence of replication. Finally, we examine the genetic overlap between Multiple Sclerosis (MS) and Crohn's Disease (CD) in more detail by comparing genome-wide meta-analyses of the two diseases across 629,000 SNPs and find association to components of the IL23R-mediated signaling pathway in MS, suggesting a role for variation in TH17 cell function in the pathogenesis of both diseases.

## 269

**Identification of the biologically relevant HLA DR-DQ amino acids in type 1 diabetes (T1D).** *G. Thomson[1], D. Karp[2], N. Marthandan[3], P. Guidry[3], S. Mack[5], R. Single[6], A. Valdes[7], R. Scheuermann[3,4], W. Helmberg[8], T1DGC.* 1) Dept Integrative Biol, Univ California, Berkeley, CA; 2) Dept Internal Med, U.T. Southwestern Medical Center, Dallas, TX; 3) Dept Pathology, U.T. Southwestern Medical Center, Dallas, TX; 4) Dept Clinical Sci, U.T. Southwestern Medical Center, Dallas, TX; 5) Children's Hospital Oakland Research Institute, Oakland, CA; 6) Dept Math and Statistics, University of Vermont, Burlington, VA; 7) Dept Twin Research and Gen Epi, King's College, London, UK; 8) Dept Blood Group Serology and Transfusion Medicine, University of Graz, Graz, Austria.

The HLA class II DRB1-DQB1 genes are the major type 1 diabetes (T1D) genetic susceptibility loci, with an extensive hierarchy of haplotype and genotype effects from very predisposing to very protective. HLA molecules present peptide fragments from self and foreign proteins to T cells; any specific HLA allele can only bind a subset of available peptides. Autoimmune disease results when there is a breakdown in tolerance to self peptides. With T1D, as well as the many other HLA associated diseases, it is difficult to identify the combinations of biologically relevant amino acids directly involved in disease given the high level of HLA polymorphism and the pattern of amino acid variability, including varying degrees of linkage disequilibrium (LD). Using a suite of complementary methods, we have analyzed ~1400 Caucasian pedigrees with high resolution HLA DR-DQ typing from the T1D genetics consortium (T1DGC), and some smaller T1D data sets from the literature to include ethnic variation. We applied standard analyses at the haplotype and single amino acid levels, as well as the unique combinations method which identifies sets of amino acids unique to e.g., predisposing versus protective alleles or haplotypes, and the overall conditional haplotype method to determine if additional amino acid sites are required to account for all disease risk. Additionally, we applied two new analytic tools which were very informative to our study: an asymmetric measure of LD which more accurately detects the correlation between amino acid sites than standard LD measures, and the sequence feature variant type (SFVT) method. With SFVT analyses, association tests are performed on variation at biologically relevant SFs based on structural (e.g., beta-strand 1) and functional (e.g., peptide binding) features of the protein, and combinations thereof. Groupings of variable amino acid sites shared by several alleles (shared epitopes) are likely better descriptors of the actual causative genetic variants than allele level variation. We demonstrated that amino acid variation within the peptide binding sites (PBSs) of the HLA DRB1 and DQB1 proteins explains the complex DRB1-DQB1 haplotype associations with T1D. Further, for DQB1, variation in pocket 9 of the PBS is the main contributor to T1D risk, whereas for DRB1 variation covering a number of pockets in the PBS is required. Our results can guide future functional studies of the role of HLA DR-DQ genes in T1D.

## 270

**HLA-DQ *trans* $\alpha\beta$ heterodimer formation and stability for disease association studies.** *H. Miyadera, K. Tokunaga.* Dept. Human Genetics, University of Tokyo, Tokyo, Japan.

*HLA-DQ* is a primary genetic factor of IDDM, celiac disease, narcolepsy and other autoimmune disorders, as has been revealed recently by genome wide association studies (GWAS). Unlike *HLA-DR*, extensive polymorphisms in both *DQA1* and *DQB1* allows the formation of $\alpha/\beta$ *trans* heterodimer in heterozygous individuals, and particular $\alpha/\beta$ *trans* combinations are known to be associated with certain autoimmune disorders. However, we are yet to have a comprehensive picture on the *DQA1* and *DQB1* allele combinations that do or do not generate additional heterodimers in *trans*. In the present study, we developed a quantitative assay for measuring the relative stability of HLA-DQ $\alpha/\beta$ heterodimer, and tested the dimer formation and stability of various *DQA1* and *DQB1* combinations, covering all major alleles in various populations.

**[Results]** We observed that HLA-DQ $\alpha/\beta$ heterodimer is formed mainly within evolutionally related subgroups (*DQA1*01-DQB1*05/*06* and *DQA1*02/*03/*04/*05/*06-DQB1*02/*03/*04*). However, certain *DQA1-DQB1* combinations of the distinct subgroups also form highly stable $\alpha/\beta$ heterodimers. Furthermore, the relative stability of $\alpha/\beta$ heterodimer differs significantly between alleles, ranging two orders of magnitude in the dimer stability index. Interestingly, highly stable dimer is formed between *DQA1* and *DQB1* alleles either in strong/weak linkage disequilibrium. Likewise, very weak stability was observed for some *DQA1-DQB1* combinations on the major haplotypes. Importantly, the level of dimer stability appears to be controlled by polymorphisms in the peptide binding groove.

These findings, together with the fact that the exceptionally high stability is observed for some of the disease associated *DQA1* and *DQB1* combinations, raise the possibility that the heterodimer stability *per se* might account for the association to some autoimmune disorders.

We also identified key amino acid residues that control the restricted dimer formation within subgroups. Based on these findings, the possible evolution of *DQA1* and *DQB1* loci, leading to the current *DQA1-DQB1* haplotypes, will be also discussed.

## 271

**A large-scale replication study identifies novel risk loci for Systemic Lupus Erythematosus.** *V. Gateva[1], J. Sandling[2], K. Taylor[3], S. Chung[3], X. Sun[1], W. Ortmann[1], R. Ferreira[1], B. Kimberly[4], E. Brown[4], S. Manzi[5], M. Petri[6], A. Lee[7], R. Kosoy[8], M. Seldin[8], P. Gregersen[7], L. Rönnblom[2], L. Criswell[3], A. Syvänen[2], T.W. Behrens[1], R.R. Graham[1].* 1) Genentech, Inc., South San Francisco, CA; 2) Department of Medical Sciences, Uppsala University, Uppsala,Sweden; 3) University of California- San Francisco, CA, USA; 4) University of Alabama-Birmingham, AL, USA; 5) University of Pittsburgh Medical Center, PA, USA; 6) Johns Hopkins University, MD, USA; 7) North Shore University Hospital, NY, USA; 8) University of California-Davis, CA, USA.

Systemic Lupus Erythematosus (SLE) is a chronic inflammatory autoimmune disease characterized by the presence of antibodies to nuclear self-antigens. Despite considerable clinical heterogeneity, SLE ranks among the most heritable of common autoimmune disease with a sibling risk ratio of ~30. Genome-wide association studies have recently identified at least 15 novel susceptibility loci for systemic lupus erythematosus (SLE). In order to confirm additional risk loci, we genotyped on a custom Illumina chip SNPs from 2,466 regions (3735 SNPs) that showed nominal evidence of association to SLE ($P < 0.05$) in a genome-wide study in an independent sample of 1,963 cases and 4,329 controls from European ancestry. Association was evaluated using an additive genetic model and over 7,000 ancestry informative markers were used to detect and adjust for population stratification. This replication effort identified 5 novel susceptibility loci that reached genome-wide significance ($P < 5 \times 10\text{-}8$). Another 21 loci showed association ($P \leq 1 \times 10\text{-}5$) suggesting that more loci might be involved in SLE pathogenesis. Finally, a screen of alleles previously associated with other autoimmune diseases identified 6 loci that may contribute to SLE. While each of the identified alleles accounts for only a fraction of the overall genetic risk, our results are providing new insights into the pathogenesis of lupus and are suggesting new targets and pathways for drug discovery and development.

## 272

**Genetic Interactions Reveal a Novel B-Cell Pathway in Systemic Lupus Erythematosus.** *A.M. Delgado-Vega[1], C. Castillejo-Lopez[1], S.V. Kozyrev[1], E. Sanchez[2], J.R. Lopez-Egido[1,3], J.A. James[4], J.T. Merril[5], J.A Kelly[3], K. Kaufman[6], K. Moser[3], G. Gilkeson[7], B.A. Pons-Estel[8], S. D'Alfonso[9], T. Witte[10], J.L. Callejas[11], J.B. Harley[3], P. Gaffney[3], J. Martin[2], J.M. Guthridge[4], M.E. Alarcon-Riquelme[1,3,12].* 1) Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Uppsala, Sweden; 2) Instituto de Biomedicina y Parasitología López-Neyra, CSIC, Granada, Spain; 3) Arthritis and Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma, USA; 4) Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma, USA; 5) Clinical Pharmacology Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA; 6) US Department of Veterans Affairs Medical Center, Oklahoma City, OK; 7) Department of Medicine, Division of Rheumatology, Medical University of South Carolina, Charleston, SC, USA; 8) Sanatorio Parque, Rosario, Argentina; 9) University of Eastern Piedmont, Novara, Italy; 10) Clinic of Immunology and Rheumatology, Medical School Hannover. University of Hannover, Germany; 11) Hospital Clínico San Cecilio, Granada, Spain; 12) GENyO, Andalucian Center for Genomics and Oncological Research Pfizer-University of Granada-Junta de Andalucía, Granada, Spain.

BACKGROUND: Epistasis or genetic interaction might explain larger genetic effects on the susceptibility to diseases than single-gene associations and help define functional pathways with potential therapeutic targets.

AIM: To identify genes that modify the susceptibility to SLE through their interaction with the B-cell scaffold protein with ankyrin repeats gene (*BANK1*).

METHODS: We searched for genetic interactions in the Affymetrix 100k genome-wide scan performed in 256 cases and 515 controls from Sweden. A subsequent replication study included two independent multicenter cohorts of European-Americans (n=676 cases and 850 controls) and Europeans (n= 1265 SLE cases and 1506 controls). We developed a genotypic interaction method based on contingency tables for all possible genotype combinations between pairs of SNPs with $r^2 < .80$ and calculated a Pearson *S* score of interaction association and its chi-squared *P* value. Each interacting combination was tested against the hypothesis of independence to derive an epistasis score ($S_e$) and a P value ($P_e$) was obtained through permutation.

RESULTS: *BANK1* showed genetic interactions with 29 genes, including the B-cell tyrosine kinase (*BLK*) and the inositol 1,4,5-triphosphate receptor 2 (*ITPR2*). One fifth of SLE patients (21% vs. 8 % of controls) were homozygous for the risk alleles of polymorphisms in these three genes with a significant epistatic effect ($P_e < 0.0002$). The interactions *BANK1xITPR2* and *BANK1xBLK* were replicated in two independent European-American ($P = 2.1 \times 10^{-6}$) and European sets ($P = 4.11 \times 10^{-9}$). The data was verified using multifactor dimensionality reduction (MDR). Moreover, BLK co-immunoprecipitated and co-localized with BANK1 in co-transfected HEK-293T. Exogenous expression of BANK1 in human Daudi B cells curbed BLK from reaching the plasma membrane with the subsequent accumulation in cytoplasmic compartments. Expression of BANK1 and BLK but not ITPR2 was modulated by IFN$\alpha$.

CONCLUSIONS: *BANK1*, *BLK* and *ITPR2* are genetically and functionally interacting partners and through their protein-protein interactions might results in a novel B-cell signaling pathway regulated by type I interferon $\alpha$. This pathway may affect B-cell responses to self-antigens in human lupus.

## 273

**A Comprehensive GWAS Follow-up in Multiple Sclerosis Identifies Novel Risk Loci.** *J.L. McCauley[1], R.L. Zuvich[2], A.H. Beecham[1], P.L. De Jager[3], A. Compston[4], D.A. Hafler[3], S.L. Hauser[5], J.R. Oksenberg[5], S.J. Sawcer[4], M.A. Pericak-Vance[1], J.L. Haines[2] for the International Multiple Sclerosis Consortium (IMSGC).* 1) Miami Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL; 2) Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN; 3) Division of Molecular Immunology, Center for Neurologic Diseases, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 4) Department of Clinical Neurosciences, Addenbrooke's Hospital, University of Cambridge, Cambridge, UK; 5) Department of Neurology, School of Medicine, University of California, San Francisco, CA.

Multiple sclerosis (MS [MIM 126200]) is a debilitating neuroimmunological and neurodegenerative disease affecting more than 400,000 individuals in the United States and more than 1.3 million worldwide. We previously published the first genome-wide association study (GWAS) for MS in a dataset of 931 Caucasian trios. This screen was performed using the Affymetrix GeneChip Human Mapping 500K array testing for association using both the transmission disequilibrium test (TDT) on the trios themselves as well as the Cochran-Mantel-Haenszel (CMH) test using the affected probands and 2,431 independent controls with stratification by country of origin. Our initial replication effort was focused on a small number of high priority single-nucleotide polymorphisms (SNPs) (n=110), and successfully confirmed two results (namely *IL7RA* and *IL2RA*). However, it is likely that additional true associations lie hidden within the top 10% of SNPs. We now present a comprehensive follow-up of this full set of SNPs. We genotyped 1,343 MS cases and 1,379 controls for 30,391 SNPs, demonstrating mild to moderate levels of significance ($p \leq 0.10$) in either the TDT or CMH analysis of our original GWAS screen. These 30,391 SNPs represented those that were able to be designed on the Illumina iSelect Custom BeadChip platform, and passed an extensive quality control process. 2,291 non-MHC SNPs demonstrated modest significance ($p < 0.05$). 150 SNPs exceeded a $p \leq 0.001$, for which we attempted a further replication of a small number of these using the Sequenom platform in an independent dataset of 2,164 MS cases and 2,016 controls. We find considerable evidence for several new susceptibility loci including *KIF21B* (rs12122721, combined $p=6.56 \times 10^{-10}$, Odds Ratio=1.22) and *TMEM39A* (rs1132200, combined $p=3.09 \times 10^{-8}$, Odds Ratio=1.24), both of which meet genome-wide significance. Our results clearly demonstrate that additional susceptibility loci may be hidden in the top 10% of association results from GWAS, and that a comprehensive follow-up is essential to the successful identification of genetic loci involved in complex disease.

## 274

**Novel Risk Variants in a Genome Wide Scan of Pediatric Onset Inflammatory Bowel Disease.** *M. Imielinski[1], R.N. Baldassano[2], A. Griffiths[3], R.K. Russell[4], V. Annese[5], M. Dubinsky[6], S. Kugathasan[7], J. Bradfield[1], T. Walters[3], P. Sleiman[1], C. Kim[1], J. Van Limbergen[11], S.L. Guthery[12], L. Denson[1], D.C. Wilson[8], S.F.A. Grant[1,2], M. Daly[9], M. Silverberg[10], J. Satsangi[11], H. Hakonarson[1,2].* 1) Ctr Applied Genomics, Children's Hosp Philadelphia, Philadelphia, PA; 2) Dept of Pediatrics, Children's Hosp Philadelphia, Philadelphia, PA; 3) The Hospital for Sick Children, University of Toronto, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada; 4) Paediatric Gastroenterology and Nutrition, Royal Hospital for Sick Children, Glasgow, UK; 5) Units of Gastroenterology and Endoscopy, IRCCS-CSS Hospital, S. Giovanni Rotondo, Italy; 6) Departments of Pediatrics and Common Disease Genetics, Cedars Sinai Medical Center, Los Angeles CA; 7) Department of Pediatrics, Emory University School of Medicine and Children's Health Care of Atlanta, Atlanta, GA, USA; 8) Paediatric Gastroenterology and Nutrition, Royal Hospital for Sick Children, Edinburgh and Child Life and Health, University of Edinburgh, UK; 9) Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, 185 Cambridge Street, Boston, Massachusetts 02114, USA; 10) Mount Sinai Hospital IBD Centre, University of Toronto, 441-600 University Avenue, Toronto, Ontario M5G 1X5, Canada; 11) Gastrointestinal Unit, Division of Medical Sciences, School of Molecular and Clinical Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK; 12) Department of Pediatrics, University of Utah School of Medicine and Primary Children's Medical Center, Salt Lake City, Utah.

Notwithstanding considerable recent success in gene identification, current understanding of the genetic basis of susceptibility to the inflammatory bowel diseases (IBD), Crohn's disease (CD) and ulcerative colitis (UC), remains incomplete. We report the results of a genome-wide association study on 2413 patients with pediatric onset IBD from the United States, Canada, Scotland and Italy and 6197 genetically matched controls to identify disease susceptibility genes. We performed high throughput genotyping on the Illumina HumanHap550 platform. For replication of positive signals, we used a large publically available CD meta-analysis dataset. We identified and replicated IL27 on 16p11 (rs1968752, P=1.27x 10-8, OR = 1.26 [1.16-1.36]) and LNPEP-LRAP on 5q15 (rs10044354, P=4.5 x 10-7, OR=1.22 [1.13-1.31]) as CD loci and SMAD3 on 15q22 (rs16950687, P=6.67 x 10-7, OR= 1.20 [1.12-1.29]) and HORMAD2 on 21q22 (rs2412973, P = 9.99 x 10-7, OR = 1.18 [1.10-1.26]) as IBD loci. We also discovered a Toll-like receptor gene cluster on 4p14 for UC with onset prior to 8 years of age (rs4833103, P = 1.805 x 10-8, OR=0.56 [0.46-0.69]). Our results further revealed that 22 of 32 previously implicated adult-onset CD loci and 10 of 17 previously implicated adult-onset UC loci contribute to the pathogenesis of the childhood-onset form of the disease. Functional followup of the 15p11 signal revealed robustly decreased IL-27 gene expression with additional copies of the rs1968752 risk allele. Measuring IL27 colonic gene expression in 37 CD and 13 control samples, we detected significantly reduced expression in CD when compared to normal tissue (P=0.028). Expression analysis of genes at the 4p14 locus revealed significant upregulation of TLR1, TLR6, and TLR10 transcripts in colonic biopsy samples obtained from UC patients relative to normal controls (P<0.05). The results of our current study substantially advance the knowledge of pathogenic mechanisms mediating early onset IBD. Our findings also suggest that molecular events in early-onset disease closely parallel molecular mechanisms in adult IBD. Loci discovered by our study further crystallize the links between inflammation and adaptive / innate immunity in the pathogenesis of IBD.

## 275

**A genome-wide search for rare variants underlying Crohn's disease risk identifies a potential susceptibility locus.** *C.A Anderson[1,3], J.C. Lee[2], J.C. Barrett[3], J.C. Mansfield[4], C.G. Mathew[5], J. Satsangi[6], M. Parkes[2], A.P. Morris[1].* 1) Genetic and Genomic Epidemiology Unit, Wellcome Trust Centre of Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Inflammatory Bowel Disease Research Group, Addenbrooke's Hospital, University of Cambridge, Cambridge, United Kingdom; 3) The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom; 4) Department of Gastroenterology and Hepatology, University of Newcastle upon Tyne, Royal Victoria Infirmary, Newcastle upon Tyne, United Kingdom; 5) Department of Medical and Molecular Genetics, King's College London School of Medicine, London, United Kingdom; 6) Gastrointestinal Unit, School of Molecular and Clinical Medicine, University of Edinburgh, Western General Hospital, Edinburgh, United Kingdom.

Genome-wide association (GWA) studies have identified over 30 risk loci for Crohn's disease (CD). Together, these variants are estimated to explain around 20% of the genetic variance in disease risk, clearly indicating that further susceptibility loci remain to be identified. To date GWA studies have only typically involved single SNP tests of association, and therefore (given current sample sizes) are only reasonably powered to identify relatively common variants of modest to large effect. In an effort to increase power to detect association to rarer variants we carried out a novel, gene-centric, multi-locus case-control association analysis of 1,748 CD cases from the UK and 2,938 unselected UK population controls (from the Wellcome Trust Case Control Consortium study) genotyped using the Affymetrix 500K SNP chip. Only gene-based variants with a minor allele frequency (MAF) ≤ 2% in controls were considered. Per gene, logistic regression analysis was undertaken where cases-control status was regressed against the proportion of rare variants at which an individual carried at least one minor allele. Five genes were associated with disease status at $P<1x10^{-4}$, though after checking cluster plots three were removed from replication efforts. Rare variants within a known CD risk gene (*IL23R*) were associated with disease-risk ($P<1x10^{-13}$) but replication was not deemed necessary. Our one remaining putative association ($P=2.8x10^{-5}$) is with an accumulation of minor alleles at six variants within a gene of strong biological candidacy. The gene encodes a ubiquitously expressed transcription factor which acts in the *IL-2* pathway. *IL-2* is critical for the activation of T cells and, in particular, the development and homeostasis of regulatory T cells which are, in turn, critical in down-regulating inflammatory responses by de-activating activated "effector" T cells. These effector cells are known to be important in organ-specific inflammation such as CD. Genotyping of the 6 variants with MAF ≤2% (plus 5 additional SNPs with a MAF between 2% and 5%) in approximately 2,000 independent CD cases and a further 2,024 UK population based controls is currently under way (and will be presented) in an attempt to replicate and thus confirm, this novel CD association.

## 276

**Deficiency of the Interleukin-1 Receptor Antagonist (DIRA) - A Systemic Autoinflammatory Disease of Skin and Bone.** *I. Aksentijevich[1], S.L. Masters[1], P.J. Ferguson[2], P. Dancey[3], J. Frenkel[4], A. van Royen-Kerkhoff[4], U. Tedgard[5], R. Laxer[6], T-H. Pham[1], M.G. Booty[1], B. Yaung[1], N. Sandler[7], J. Estes[8], E.F. Remmers[1], D.L. Kastner[1], R. Goldbach-Mansky[1].* 1) NIAMS/NIH, Bethesda, MD; 2) University of Iowa, Iowa City , IA; 3) Memorial University of St. Paul, Newfoundland , Canada; 4) University of Utrecht, Netherlands; 5) Lund University, Malmo, Sweden; 6) University of Toronto, Canada; 7) NIAID/NIH, Bethesda, MD; 8) NCI/NIH, Bethesda, MD.

Autoinflammatory diseases are systemic inflammatory diseases characterized by the absence of high-titer autoantibodies or antigen-specific T-cells. Genetic studies of autoinflammatory diseases have provided important insight into the regulation of innate immunity. An infant patient from Newfoundland presented with neonatal-onset sterile multifocal osteomyelitis, periostitis, joint swelling, and extensive cutaneous pustulosis. Clinical response to empiric treatment with the recombinant interleukin-1 receptor antagonist, anakinra, suggested IL-1 mediated pathology and prompted mutational analysis of IL1RN, the gene encoding interleukin-1 receptor antagonist (IL-1Ra). We identified a homozygous 2bp deletion (c.156_157delCA) that caused a frameshift mutation, N52Kfs25, leading to a premature stop codon. Both parents were heterozygous carriers and clinically asymptomatic. We then screened other patients presenting with a similar phenotype and identified 3 additional mutations in IL1RN. A nonsense mutation (E77X) was found in 5 patients from 3 unrelated families from the Netherlands. Two siblings from a consanguineous Lebanese family were homozygous for another nonsense mutation (Q54X). A patient from Puerto Rico was homozygous for a 175kb genomic deletion that includes IL1RN and five other genes from the IL-1 family. None of these mutations were found in a panel of 364 Caucasian control DNA samples, however, the 175 kb deletion was found in 3/119 unrelated samples from the geographically matched region of Puerto Rico and the N52Kfs25 mutation was found in 2/555 controls from Newfoundland. These mutations likely represent founder effects in genetically isolated populations. Functional studies assessing mRNA, ex vivo and in vitro protein expression, and cytokine assays were performed in patients, parents, and controls. IL1RN mutations result in absent or prematurely truncated protein that is not secreted, rendering patient cells hyper-responsive to IL-1 stimulation with increased production of proinflammatory cytokines and chemokines. Patients treated with anakinra exhibited a uniformly rapid clinical and immunological response. We define a new autosomal recessive autoinflammatory disease, DIRA, caused by Deficiency of the Interleukin-1 Receptor Antagonist. The absence of IL-1Ra leads to unopposed action of IL-1, which is associated with neonatal-onset, life-threatening autoinflammation highly responsive to anakinra.

## 277

**Large differences in DNA methylation between monozygotic twin pairs.** *B. Herb[1], D. Undlien[2], H. Gjessing[3], R. Lyle[2], J. Harris[3], K. Gervin[2], A. Feinberg[1].* 1) Center for Epigenetics and Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 2) Department of Medical Genetics, Ullevål University Hospital, N-0407 Oslo, Norway; 3) Division of Epidemiology, Department of Genes and Environment, Norwegian Institute of Public Health, N-0403 Oslo, Norway.

Monozygotic twins possess identical genomic sequences, making them valuable models for studying the role of epigenetic modifications in determining phenotype. Here we have studied one epigenetic modification, DNA methylation, which assists in regulation of gene expression and differentiation of cells towards distinct lineages. We used the Comprehensive High-throughput Arrays for Relative Methylation (CHARM)[1] method to determine the genome-wide methylation profile of 15 monozygotic (MZ) and 21 dizygotic (DZ) twin pairs. Methylation levels within MZ twin pairs correlated more strongly than within DZ twins, suggesting a genetic contribution to epigenotype. Regions where there was a large consistent methylation difference between twins were defined as hypervariable regions. These regions show surprisingly large differences in methylation between twins, up to 30%, confirmed by bisulfite pyrosequencing, an independent highly quantitative method that interrogates individual CpG sites. This large difference in methylation at multiple loci is a new observation in twins and is consistent with the global magnitude of DNA methylation change observed within some individuals in the population as they age[2]. Variation in methylation at these hypervariable regions were roughly equally divided between three types of differences: (1) where MZ and DZ methylation variance was low, but differences between random pairings of samples was high, suggesting maternal effects; (2) where MZ pair differences varied less than did DZ pair differences, and these less than did random pairings, suggesting genetic sequence effects; and (3) where MZ, DZ, and random pairings were all highly variable, suggesting environmental effects. Thus regions of the epigenome may be affected unequally by genes and the environment. **1**. Irizarry R, Ladd-Acosta C, Carvalho B, et al. Comprehensive high throughput arrays for relative methylation (CHARM). Genome Research 2008; 18: 780-790 **2**. Bjornsson H, Sigurdsson M, Fallin MD, et al. Intra-individual change over time in DNA methylation with familial clustering. JAMA 2008; 299(24): 2877-2883.

## 278

**Genome-wide reprogramming of promoter DNA methylation in the derivation of human induced pluripotent stem cells.** *G. Fan[1], Y. Shen[1], Z. Xue[2], M. Bibikova[2], C. April[1], Z. Yue[1], A. Wang[1], Z. Ma[1], S.D. Fouse[3], W. Li[3], S. Ding[3], J.-B. Fan[2].* 1) Human Genetics, UCLA, Los Angeles, CA 90095; 2) Illumina, Inc., San Diego, CA 92121; 3) Chemistry, Scripps Research Institute, San Diego, CA 92037.

To understand DNA methylation regulation in direct reprogramming of somatic cells into induced pluripotent stem cells (iPSCs), we compared methylation patterns in 26,837 CpG sites across 14,152 gene promoter in human somatic cells, iPSCs, and embryonic stem cells (hESCs). Unsupervised clustering analysis showed that gene promoter methylation patterns between hESCs and iPSCs are highly similar to each other and are both distinctly different from somatic cells. Pair-wise comparisons between somatic cells and its corresponding iPSCs revealed that approximately 7-14% of gene promoters have altered methylation pattern. Interestingly, there are significantly more gene promoters that undergo de novo methylation than demethylation. Whereas genes with decreased methylation and increased expression in iPSCs are primarily developmental genes and transcription regulators (e.g. OCT4 and CDX1), genes that exhibit increased methylation and decreased expression in iPSCs encode proteins for specialized cellular function and cell differentiation pathways. Using supervised clustering analysis, we also identified unique methylation signatures for both iPSCs and hESCs that can be used as a biomarker to distinguish from each other. Collectively, our data demonstrated that the overall gene promoter methylation pattern of human iPSCs highly resembles that of hESCs and that both demethylation and de novo methylation are crucial events during direct reprogramming of somatic cells into iPSCs.

## 279

**The retinoblastoma (*RB1*) gene is imprinted.** *D. Kanber[1], T. Berulava[1], O. Ammerpohl[2], D. Mitter[1], J. Richter[2], R. Siebert[2], B. Horsthemke[1], D. Lohmann[1], K. Buiting[1].* 1) Institut für Humangenetik, Universitätsklinikum Essen, Essen, Germany; 2) Institut für Humangenetik, Christian-Albrechts Universität zu Kiel, Universitätsklinikum Schleswig-Holstein, Kiel, Germany.

Recently we observed a patient with complex phenotypic features and hypomethylation of all known imprinted loci tested (Caliebe et al., in preparation). Genome-wide CpG methylation analyses (Infinium® HumanMethylation27 BeadChip, Illumina) in the patient and appropriate controls confirmed hypomethylation of known imprinted loci and moreover, identified additional loci hypomethylated in the propositus. One of these loci is a 1.2 kb CpG island within intron 2 of the *RB1* gene. DNA methylation analysis by bisulfite cloning and sequencing of blood DNA from a normal control and patients with retinoblastoma and a maternally or paternally derived *RB1* deletion revealed that this CpG island is subject to parent-of-origin specific methylation: Clones obtained from the normal control showed methylated and unmethylated sequences. Almost all clones from the patients with a deletion on the maternal chromosome were derived from completely unmethylated sequences, whereas all of the clones obtained from the patients with a paternal deletion were derived from completely methylated sequences. To find out whether this CpG island serves as a promoter for a yet unknown transcript, we performed RT-PCR experiments and identified an alternative *RB1* transcript by connecting the CpG island with exon 3 of the original *RB1* gene, which contains an in-frame start codon. Furthermore, we could show that this alternative *RB1* transcript is expressed from the paternal chromosome only and thus subject to genomic imprinting. By primer extension analysis of mRNA from 14 individuals of seven families heterozygous for rare expressed variants in exons 3, 9, 12, 18, 21, and 23 of the *RB1* gene we found that all individuals tested showed a higher abundance of *RB1* mRNA from the maternal allele relative to the paternal allele. The mean of the normalized transcript ratios was 2.7±0.43 with low variation between different individuals. To find out whether the expression bias in favour of the maternal allele was due to interference of the two *RB1* promoters, we treated lymphoblastoid cell cultures with 5-aza-2'-deoxycytidine. This treatment resulted in reduced methylation of the alternative *RB1* promoter and reduced skewing of the allelic transcript levels. We propose that genomic imprinting of an alternative *RB1* promoter modifies the expression and function of the *RB1* gene.

## 280

**Deficiency in the histone methyltransferase Ash1l, a Drosophila homeotic selector homolog, causes blepharitis, hematopoietic defects, infertility and poor viability in mice.** *M.L. Brinkmeier[1], G.D. Gregory[2], A. Friedman[3], M. Jones[3], K. Geister[1], G.A. Blobel[2], I. Maillard[3], S.A. Camper[1].* 1) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 2) Department of Hematology, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Departments of Medicine & Cell and Developmental Biology, University of Michigan, Ann Arbor, MI.

*Ash1l*, (absent, small, or homeotic)-like 1, is an ortholog of the *Drosophila* gene *ash1*, which regulates expression of the *Hox* cluster through histone methylation. We identified Ash1l transcripts in the developing mouse pituitary gland through analysis of cDNA libraries generated from e12.5 wild type Rathke's pouch[a], and detected expression in many developing and adult tissues including the brain, lens, heart, and somites. We generated *Ash1l* knockout mice to test the role of *Ash1l* during development. Mice heterozygous for the *Ash1l* gene trap in intron 1 are viable but exhibit chronic inflammation of the eye or blepharitis, hematopoietic defects, and decreased male fertility. Homozygous mutants have poor viability and fail to thrive, with severe growth deficiency evident at P7. Mutant mice also display chronic blepharitis, female infertility, reduced male fertility, decreased abdominal fat, and abnormalities in the development of the hematopoietic and immunological systems. We examined expression of the *Hoxa* cluster in e16.5 mouse embryonic fibroblasts. Both *Hoxa11* and *Hoxa13* transcripts are reduced in *Ash1*l mutant fibroblasts, and the expected changes in histone methylation, increased H3K4me3 and decreased H3K27me3, are found at these loci by ChIP analysis. *Ash1l* mutants have hypoplastic uteri similar to *Hoxa11* mutants[b], supporting the hypothesis that ASH1L regulates *Hoxa11* in intact animals, and providing an explanation for the infertility. *Ash1l* mutants have phenotypic similarities to lens epithelial derived growth factor (Ledgf[c] or Psip1) and mixed-lineage leukemia 5[d] (Mll5) mutants suggesting an interaction between these proteins in epigenetic regulation of a set of common genes. These studies describe a developmental role for ASH1L in vivo and highlight its broad effects on gene transcription in many organ systems. Research supported by NIH grant R37-HD30428 a Brinkmeier et al. Genomics , 93: 449, 2009 b Gendron et al. Biol. of Reprod. 56:1097, 1997 c Sutherland et al. Mol. Cell. Biol. 26:7201, 2006 d Heuser et al. Blood. 113:1432-1443, 2009; Madan et al. Blood. 113: 1444-1454, 2009; Zhang et al. Blood 113: 1455-1463, 2009.

## 281

**Identification of imprinted genes contributing to specific brain regions using high throughput sequencing.** *C. Barr[1,2], K. Wigg[1], E. Dempster[1], L. Gomez[1], Y. Feng[1], P. Monnier[1], R. Logan[1], J. Eubanks[1].* 1) Genetics & Development, Toronto Western Hosp, UHN, Toronto, ON, Canada; 2) Program in Neurosciences and Mental Health, The Hospital for Sick Children.

Introduction: Current estimates suggest that there are around 100 imprinted genes, but this number may be as high as 600, with some imprinted in a tissue/cell type specific pattern and little is known of these genes. We followed up evidence that suggests imprinted genes influence the development of specific brain regions. We predicted that key genes expressed from the paternal genome will contribute to the development of the hypothalamus while key maternally expressed genes will influence the development of the neocortex. Methods: We developed a strategy using crosses and reverse crosses of mouse strains to determine the parent-of-origin of expressed genes. We used next generation sequencing of mRNA tags from the cortex and the thalamus/hypothalamus dissected from the F1 to identity tags differing in expression between the crosses. Results: We identified tags that were differentially expressed between the crosses showing the same pattern of parent-of-origin effect in both brain regions including known imprinted genes/transcripts (Meg3, Snrpn, Rian) and genes with no prior evidence for parent-of-origin effects (Pctk3, Klf10). We also observed genes that displayed a parent of origin effect in only one brain region with biallelic expression in the other region (Pon2, Itgb1bp1). Conclusions: Our results show the feasibility of using next generation sequencing of RNA as a genome-wide approach to identify novel imprinted genes as well as to identify tissue specific effects. The understanding of parent-of-origin effects in specific brain regions will make a significant contribution to our understanding of the role of imprinting in brain development and function.

## 282

**Somatic cell reprogramming involves differential methylation at tissue- and cancer- related CpG island shores.** *A. Doi[1], I. Park[2], B. Wen[1], P. Murakami[1], R. Irizarry[3], B. Herb[1], C. Ladd-Acosta[1], S. Loewer[2], T. Schlaeger[2], J. Miller[2], J. Rho[2], G.Q. Daley[2], A.P. Feinberg[1].* 1) Center for Epigenetics and Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Division of Pediatric Hematology/Oncology, Children's Hospital Boston and Howard Hughes Medical Institute, Boston, MA; 3) Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD.

Induced pluripotent stem (iPS) cells are cells that resemble embryonic stem (ES) cells and are derived from somatic cells by the ectopic expression of factors such as *Oct4*, *Sox2*, *Klf4*, and *c-Myc*. Although reprogramming of epigenetic marks in the parental somatic cells is fundamental to the generation of iPS cells, the DNA methylation patterns have not yet been fully characterized on a genome-wide scale. Here we used Comprehensive High-throughput Array-based Relative Methylation (CHARM) analysis to study the DNA methylation changes associated with somatic cell reprogramming. We observed differential DNA methylation in six human iPS cell lines compared to its parental fibroblast lines at 4401 regions (FDR≤5%), the majority of which were located at CpG island shores. Nine of these differentially methylated regions in the reprogrammed cells (R-DMRs) were validated using bisulfite pyrosequencing, a quantitative approach that measures DNA methylation at the single nucleotide resolution. Using GO annotation analysis, we found that R-DMRs were significantly enriched for genes shown to be important in development. We also observed a strong inverse correlation between differential gene expression and differential DNA methylation at R-DMRs that are within 1 kb of a transcriptional start site (TSS) of a gene. Interestingly, R-DMRs were significantly enriched in the previously identified tissue-specific T-DMRs that distinguish tissues representing the three germ cell lineages (i.e. brain, liver, and spleen) ($P < 10^{-4}$). R-DMRs were also enriched in cancer-specific C-DMRs which distinguish colon cancer from its matched normal counterpart ($P < 10^{-4}$). Moreover, the R-DMRs were able to differentiate brain from liver from spleen, as well as colon cancer from normal colon. These data suggest that DMRs involved in somatic cell reprogramming are also associated with normal tissue development as well as tumorigenesis.

## 283

**Genomewide ChIP-sequencing Reveals that the ATR-X Syndrome Protein Influences Allele-Specific Gene Expression Through its Interaction with Guanine-Rich VNTRs.** *M.J. Law[1], K.M. Lower[1], I. Dunham[2], J. Hughes[1], H. Ayyub[1], D.R. Higgs[1], R.J. Gibbons[1].* 1) MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, Oxford, Oxfordshire, United Kingdom; 2) European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom.

ATR-X is a syndromal form of X-linked mental retardation (XLMR) including severe mental retardation, facial, skeletal and genital abnormalities, short stature and varying degrees of alpha thalassaemia. It is due to constitutive mutations in the ATRX gene, which encodes a chromatin associated, SWI/SNF-like protein. It is known that alpha thalassaemia in ATR-X syndrome is due to down-regulation of alpha globin expression at the RNA level and this has led to the hypothesis that ATRX acts as a transcriptional regulator. To date, however, little is known about the mechanism of ATRX action at the globin genes, the identity of other ATRX targets and, therefore, how ATR-X pathology arises. To identify direct genomic targets of ATRX we performed whole genome Solexa sequencing of ATRX chromatin immunoprecipitated DNA from cultured primary human erythroid cells. This confirmed the alpha globin locus as a direct ATRX target, and identified 107 previously unidentified ATRX target genes. Sequence analysis of ATRX binding sites revealed that in approximately half of these sites the peak of binding coincided with guanine-rich variable number tandem repeats (VNTR). The alpha globin gene locus contains an ATRX binding VNTR the length of which is highly polymorphic. A clue to the function of ATRX has been found by looking at the effect of the VNTR length polymorphism. Although there is no obvious effect on globin expression in normal individuals, in the presence of an ATRX mutation the level of alpha globin expression is inversely correlated with VNTR length. A similar allele-specific effect was observed at another ATRX target, NME4, suggesting that this may be a genome-wide phenomenon. These data show that in the absence of ATRX, these G-rich VNTRs become cis-acting regulatory polymorphisms which act to down-regulate associated genes and that ATRX may be required to alleviate the negative effect of these features on transcription. Not only does this observation reveal a potential novel function for a SWI/SNF-like chromatin associated protein but this characterisation of the interaction between a rare trans-acting mutation and a cis-acting polymorphism provides a useful molecular paradigm to understand phenotypic variability and variable penetrance in human disease.

## 284

**The yeast ataxin-7 orthologue regulates lifespan through Sir2-dependent and histone deubiquitination chromatin modification pathways, linking aging and neurodegeneration in spinocerebellar ataxia type 7.** *S. Guyenet[1,2], B.K. Kennedy[3], A.R. La Spada[1,2].* 1) Depts. of Pediatrics and Cellular & Molecular Medicine, UCSD, San Diego, CA; 2) Dept. of Laboratory Medicine, Univ Washington, Seattle, WA; 3) Dept. of Biochemistry, Univ Washington, Seattle, WA.

Spinocerebellar ataxia type 7 (SCA7) is an inherited disorder characterized by cerebellar and retinal degeneration. SCA7 is caused by a CAG / polyglutamine repeat expansion in the ataxin-7 gene. We defined the normal function of mammalian ataxin-7 as a transcription factor, when we found that ataxin-7 is an integral subunit of the STAGA complex, a major transcriptional co-activator. Sgf73 is the Saccharomyces cerevisiae orthologue of ataxin-7, and functions in the orthologous yeast SAGA complex, influencing its Gcn5-dependent acetyl-transferase and Ubp8-dependent deubiquitinase activities. To characterize pathways of STAGA transcription regulation and to understand SCA7 transcription dysregulation, we studied the function of the yeast ataxin-7 orthologue, by characterizing the phenotype of SGF73 null yeast. Remarkably, deletion of SGF73 results in a replicative lifespan of ~42 divisions, more than two-thirds greater than WT strains on a comparable background. Epistasis analysis indicates that the replicative lifespan extension depends upon Sir2, a NAD-dependent deacetylase whose over-expression alone produces pronounced replicative lifespan extension, but at ~31 divisions the extension is significantly less than the SGF73 null. The genetic interaction between Sgf73 and Sir2 is paralleled by a functional relationship, as the SGF73 null exhibits reduced ribosomal DNA recombination. However, this does not result from increased Sir2 protein expression. As Sgf73 is required for histone H2B deubiquitination mediated by Ubp8, we examined the role of histone H2B deubiquitination in lifespan regulation, and found that deletion of Ubp8 is sufficient to produce marked lifespan extension. Furthermore, the histone ubiquitin-conjugating enzyme Rad6 is required for lifespan extension in the SGF73 null. Importantly, the combination of SIR2 over-expression and UBP8 deletion recapitulates the dramatic lifespan extension observed in SGF73 null yeast. Thus, the longevity phenotype may reflect a combination of increased histone ubiquitination and increased Sir2 activity. Our results indicate that Sgf73 coordinates a variety of chromatin modification pathways that are central to yeast lifespan regulation. We conclude that specific histone modification processes regulate lifespan in yeast, and likely affect neuron function in the mammalian CNS, linking the biology of aging to SCA7 neurodegeneration.

## 285

**Widespread translational repression by upstream open reading frames (uORFs): implications for human phenotypic variation and disease.** *S.E. Calvo[1,2,3], D.J. Pagliarini[1,2,3], V.K. Mootha[1,2,3].* 1) Broad Institute of Harvard/MIT, Cambridge, MA; 2) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 3) Department of Systems Biology, Harvard Medical School, Boston, MA.

Much phenotypic variation is due to changes in gene regulation rather than coding sequence. However, we are currently unable to interpret most forms of non-coding variation. Even within well-defined 5' UTR regions, which contain important post-transcriptional regulatory elements, we can predict few consequences of sequence variation. Here we focus on characterizing the widespread impact and variation of one 5' UTR regulatory element, the upstream open reading frame (uORF). uORFs are common mRNA elements defined by a start codon in the 5' UTR that is out-of-frame with the main coding sequence. In individual studies, uORFs have been shown to reduce mRNA stability and protein translation. However no study to date has investigated their global effect on protein expression. We report that uORFs correlate with significantly reduced protein expression of the downstream ORF, based on analysis of 11,649 matched mRNA and protein measurements from four published mammalian studies. We demonstrate that uORFs typically reduce protein expression by 30-80%, with a modest 0-30% decrease in mRNA levels. We identify uORF-altering polymorphisms in 509 human genes and demonstrate that these variants can alter protein levels. Additionally, we provide support for pathogenicity of uORF-altering mutations in 5 human diseases. Lastly, we present preliminary evidence of translational regulation by uORFs during stress response. Together, our results suggest that uORFs influence the protein expression of thousands of mammalian genes and that variation in these elements can influence human phenotype and disease.

## 286

**Human retrotransposon RNAs display divergent subcellular localization despite a common mechanism of genomic insertion.** *J.L. Goodier, P. Mandal, L. Zhang, H.H. Kazazian.* Dept Genetics, Univ Pennsylvania Sch Medicine, Philadelphia, PA.

Despite the immense significance retrotransposons have had for genome evolution, much about their biology is unknown, including little about the processes of forming their ribonucleoprotein (RNP) particles and transporting them about the cell. Suppression of retrotransposon expression, together with the presence of retrotransposon sequence within numerous mRNAs, means that tracking endogenous L1 RNP particles in cells is problematic. We overcome these difficulties by assaying in both living and fixed cells tagged-RNPs generated from constructs expressing retrotransposition-competent L1s. In this way, we directly visualize for the first time the subcellular colocalization of L1 RNA, ORF1p, and ORF2p, and show their targeting together to cytoplasmic foci. Foci are associated with markers of cytoplasmic stress granules. Furthermore, mutation analyses indicate that ORF1p can direct L1 RNP distribution within the cell. We also assayed RNA localization patterns of the non-autonomous retrotransposons Alu and SVA (a composite element comprised of SINE-R, VNTR, and Alu fragments). Despite a requirement for the L1 integration machinery, each manifests unique features of subcellular RNA distribution. SVA RNA patterning is distinctive, being cytoplasmic but without prominent foci, and concentrated in large nuclear foci that often ring nucleoli. Such variability suggests significant differences in the life cycles of these elements.

## 287

**Mutations in CCBE1 cause generalized lymph vessel dysplasia in humans.** *M. Alders[1], B.M. Hogan[2], F. Salehi[1], L. Al-Ghazali[3], E.A. Hennekam[4], E.E. Holmberg[5], M.M. Mannens[1], M.F. Mulder[6], T.E. Prescott[5], M. Witte[2], E.J. Schroor[7], P.J. Zwijnenburg[8,11], M. Vikkula[9], S. Schulte-Merker[2], R.C. Hennekam[10,11].* 1) Department of Clinical Genetics, Academic Medical Center Amsterdam, Amsterdam, Netherlands; 2) Hubrecht Institute - KNAW and University Medical Centre, Utrecht, The Netherlands; 3) Department of Paediatrics and Pathology, Faculty of Medicine and Health Sciences, UAE University, Al Ain, UAE; 4) Department of Clinical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 5) Department of Medical Genetics, Oslo University Hospital, Rikshospitalet, Oslo, Norway; 6) Department of Pediatrics, VU University Medical Center, Amsterdam, The Netherlands; 7) Department of Pediatrics Amalia, Isala Clinics, Zwolle, The Netherlands; 8) Department of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands; 9) Laboratory of Human Molecular Genetics, de Duve Institute, Université catholique de Louvain, Brussels, Belgium; 10) Clinical and Molecular Genetics Unit, Institute of Child Health, Great Ormond Street Hospital for Children, University College London, London, UK; 11) Department of Pediatrics, Academic Medical Center, Amsterdam, The Netherlands.

The lymphatic system is a general vascular system parallel to the cardiovascular system and important in many pathological processes. Both localized and generalized dysplasias exist. An unusually widespread malformation syndrome characterized by congenital limb lymphedema, lymphangiectasias, mental retardation and unusual facial characteristics is known as Hennekam syndrome (HS). HS is inherited in an autosomal recessive way. We employed homozygosity mapping in three distantly related patients and identified a 5.7 Mb homozygous region at chromosome 18q. The region was subsequently decreased to 0.5 Mb by homozygosity mapping in two additional patients born to consanguineous parents. It contained the *CCBE1* gene, the human orthologue of a zebrafish gene recently shown to be essential for lymphangiogenesis. Homozygous mutations in *CCBE1* were identified in these 5 patients, and mutation analysis in 19 additional families revealed compound heterozygous mutations in another 2 HS patients. Functional analysis of a subset of these mutations, conducted in a zebrafish in vivo model, showed that 2 of 3 tested mutations generated non-functional proteins. Our results demonstrate that a specific form of generalized lymph vessel dysplasia in humans can be caused by *CCBE1* mutation and that HS is genetically heterogeneous as mutations in *CCBE1* can be detected in about 30% (7/24) of families.

## 288

**VEGFR3 Mutations Can Cause Recessive Primary Congenital Lymphedema and Sporadic In Utero Generalized Edema.** *A. Ghalamkarpour[1], A. Mendola[1], C. Debauche[2], E. Haan[3], N. Van Regemorter[4], Y. Sznajer[4,5], D. Thomas[6], N. Revencu[1], Y. Gillerot[7], W. Holnthoner[8], P. Saharinen[8], L.M. Boon[1,9], J.B. Mulliken[10], K. Alitalo[8], M. Vikkula[1].* 1) Laboratory of Human Molecular Genetics, de Duve Institute, Université catholique de Louvain, Brussels, Belgium; 2) Department of Neonatology, Cliniques Universitaires St-Luc, Brussels, Belgium; 3) Deparment of Genetic Medicine, Women's and Children's Hospital, North Adelaide, Australia, and Dept. of Paediatrics, University of Adelaide, Adelaide, Australia; 4) Centre de Génétique ULB, Hôpital Erasme, Brussels, Belgium; 5) Unité de Génétique Clinique Pédiatrique, Université Libre de Bruxelles, Brussels, Belgium; 6) Unité diagnostic anténatal, Hôpitaux Iris Sud, Brussels, Belgium; 7) Center for Human Genetics, Cliniques Universitaires Saint-Luc, Brussels, Belgium; 8) Molecular/Cancer Biology Laboratory and Ludwig Institute for Cancer Research, Biomedicum Helsinki, Haartman Institute and Helsinki University Central Hospital, University of Helsinki, Helsinki, Finland; 9) Centre for Vascular Anomalies, Division of Plastic Surgery, Cliniques Universitaires Saint-Luc and Université catholique de Louvain, Brussels, Belgium; 10) Vascular Anomalies Center, Department of Plastic Surgery, Children's Hospital, Harvard Medical School, Boston, MA, USA.

Heterozygous mutations in VEGFR3 have been identified in some familial cases of primary congenital lymphedema with dominant inheritance, known as Nonne-Milroy disease. However, the majority of primary lymphedema is not dominantly inherited. Moreover, hydrops fetalis, a serious fetal condition with high mortality rate, is mostly sporadic and the etiology is unknown in about 25% of cases. In this study, we present the first case of isolated primary congenital lymphedema with recessive inheritance, caused by a homozygous mutation in VEGFR3. The novel mutation is a transition from alanine-to-threonine in amino acid 855, located in the ATP binding domain of the VEGFR3 receptor. Assessment of receptor function showed impaired ligand-induced internalization and ERK1/2 activity. Moreover, receptor phosphorylation was reduced, although, less so than for a kinase-dead VEGFR3 mutation, which causes Nonne-Milroy disease. In addition, we present a series of 12 patients, in whom in utero generalized skin edema or hydrops fetalis had been diagnosed, and in which the three lymphangiogenic genes, VEGFR3, FOXC2 and SOX18, were screened. In three of the patients, we identified mutations: two in VEGFR3 and one in FOXC2. Two of the mutations were de novo and one was either de novo or non-penetrant inherited. In these patients, the generalized edema resorbed spontaneously, either in utero or after birth. In the two individuals with a VEGFR3 mutation, edema remained limited to the lower limbs. In conclusion, a hypomorphic VEGFR3 mutation, with moderate effect on the receptor, in a homozygous state can result in insufficient lymphatic functioning. Thus, in addition to Nonne-Milroy disease with dominant inheritance, VEGFR3 alterations can cause isolated recessive primary congenital lymphedema. Moreover, de novo VEGFR3 mutations can cause hydrops fetalis. These data enlarge the phenotypic variability associated with VEGRF3 mutations, and expand our understanding of the aetiology of congenital lymphedema. The data suggests that large scale screening of VEGFR3 in all primary lymphedema and hydrops fetalis patients is necessary. (miikka.vikkula@uclouvain.be) (http://www.deduveinstitute.be/vikkula).

## †289

**A haplotype in the MSRA gene confers decreased risk of meconium ileus in cystic fibrosis.** *L.B. Henderson[1], V.K. Doshi[1], S.M. Blackman[1,2], K.M. Naughton[1], R.G. Pace[3], M.L. Drumm[4], M.R. Knowles[3], G.R. Cutting[1].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Pediatric Endocrinology, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Cystic Fibrosis-Pulmonary Research and Treatment Center, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC; 4) Departments of Pediatrics and Genetics, Case Western Reserve University, Cleveland, OH.

Meconium ileus (MI) is an intestinal obstruction that occurs in about 15 percent of cystic fibrosis (CF) neonates. Analysis of twins and siblings has demonstrated that MI is a highly heritable trait, indicating that genetic modifiers are largely responsible for this complication of CF. Genome-wide linkage analysis performed by our group revealed suggestive evidence for a modifier of MI at chromosome 8p23.1 (Blackman, *et al*, 2006). To conduct a regional association analysis, 3,058 SNPs were selected from a 9 Mb region within the 8p linkage peak where LOD score was >1.0. Genotypes were obtained from the North American Cystic Fibrosis Modifier Consortium GWAS using the Illumina 610 Quad array. Family-based association analysis of 136 families (269 patients) participating in the CF Twin and Sibling Study (CFTSS) in which at least one child had MI identified a cluster of SNPs within the methionine sulfoxide reductase A (MSRA) gene (highest association: rs11783705, $p=2.5x10^{-5}$). We then tested haplotypes composed of up to 8 consecutive SNPs for association with MI in a 1.85 Mb region containing the MSRA gene. A 3-SNP haplotype (rs10903323 T - rs4840475 G - rs17151637 A; frequency, 0.15) spanning a 3.5 kb region in intron 3 of *MSRA* was identified as being protective against MI ($p=3.1x10^{-7}$). In addition, phenotype and genotype data were obtained from the Gene Modifier Study Group (GMSG) for 137 CF patients with MI treated with surgery or enema and 141 CF patients verified as not having MI. Haplotypes were derived using an expectation-maximization method and case-control analysis showed that homozygosity for the T-G-A haplotype was protective against MI (p=0.040). Odds ratios were constructed to assess the difference in risk between subjects with two copies of the protective haplotype and those with no copies. In the CFTSS and GMSG cohorts, respectively, subjects homozygous for the T-G-A haplotype had 7.2-fold (95% CI, 1.4-36; p=0.016) and 9.3-fold (95% CI, 1.2-74; p=0.036) reduced odds of MI than subjects not carrying this haplotype. Interestingly, the SNPs comprising the haplotype show little correlation with each other, suggesting that the association is not due to ancestral linkage disequilibrium, but more likely due to a functional variant being harbored on this haplotype. In summary, variation in the MSRA gene modifies risk of a congenital intestinal complication in two independent collections of CF patients. *Supported by NHLBI, NIDDK, and CFF..*

## †290

**Mutations in polarin (PLRN) cause an ARC syndrome phenotype and defects in epithelial polarisation and apical junction complex formation.** *A.R. Cullinane[1], A. Straatman-Iwanowska[1], A. Zaucker[1], Y. Wakabayashi[2], C. Bruce[1], F. Rahman[1], J. Rappoport[3], I.M. Arias[2], H. Wolburg[4], A.S. Knisely[5], D.A. Kelly[6], F. Mueller[1], E.R. Maher[1,7], P. Gissen[1,8].* 1) Medical & Molecular Genetics, University of Birmingham, Birmingham, United Kingdom; 2) Cell Biology and Metabolism Program, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD, USA; 3) School of Biosciences, University of Birmingham, Birmingham, UK; 4) Institute of Pathology, University of Tübingen, Tübingen, Germany; 5) Institute of Liver Studies, King's College Hospital, London, UK; 6) The Liver Unit, Birmingham Children's Hospital, Birmingham, UK; 7) West Midlands Regional Genetics Service, Birmingham Women's Hospital, Birmingham, UK; 8) The Metabolic Unit, Birmingham Children's Hospital, Birmingham, UK.

Arthrogryposis, Renal dysfunction, and Cholestasis (ARC) syndrome is a multisystem disorder associated with abnormal localisation of some polarised membrane transporter proteins. Distinct apical and basolateral poles are essential for epithelial function and organ development but the molecular pathways determining the biogenesis of polarised membranes are not fully characterised. Mutations in VPS33B (a homologue of yeast Vps33, a SM protein) account for the majority of ARC patients but the role of vps33b in cell polarity is not clear. We identified a novel protein polarin (PLRN ) that interacts with VPS33B, which is crucial for VPS33B function and that pathogenic mutations in PLRN occur in ARC patients without VPS33B mutations. Decreased Plrn or Vps33b expression in mouse renal collecting duct cells led to abnormal localisation of specific polarised membrane proteins and to disordered apical junction complex formation. In an in vivo model, knockdown of polarin in zebrafish resulted in defects in biliary tract development. Our findings establish that a vps33b/polarin/rab11a intracellular trafficking pathway is functionally distinct from another vps33-related pathway (vps33a/vps16) and is required for (a) normal epithelial polarisation and apical junction complex formation, and (b) normal liver and kidney development and function.

## 291

**Autophagy Impairment Contributes to Dysfunction and Cell Death in Spinal and Bulbar Muscular Atrophy.** *J. Young[1], N. Ivanov[2], G.A. Garden[3], A.R. La Spada[1].* 1) Depts. of Pediatrics and Cellular & Molecular Medicine, UCSD, San Diego, CA; 2) Allen Institute for Brain Science, Seattle, WA; 3) Dept. of Neurology, Univ Washington, Seattle, WA.

Macroautophagy is involved in the pathogenesis of many human diseases, including neurodegenerative disorders. Autophagy typically is a pro-survival response in neurons affected by aggregate-prone proteins. However, accumulations of autophagosomes in neurons may signify autophagy impairment, turning a pro-survival pathway into a key contributor to cell dysfunction and death. Previously, we demonstrated that induction of autophagy by selective nutrient deprivation can protect against polyglutamine (polyQ) protein stress in a primary neuron model of spinal and bulbar muscular atrophy (SBMA). We now provide evidence that the autophagy-lysosome fusion process is necessary for degrading both non-pathogenic and pathogenic length androgen receptor (AR) protein, as inhibition of lysosomal degradation increases toxicity of both AR-19Q and AR-112Q in primary neurons. In primary neurons expressing polyQ-expanded AR, LAMP1-positive autolysosomes decrease in fluorescent intensity and change in their subcellular localization over time, becoming increasingly associated with polyQ aggregates. Ultrastructural analysis similarly revealed time-dependent differences in the frequencies of autophagic vacuoles (AVs) and autolysosomes (ALs) in the motor neurons of SBMA AR YAC transgenic mice. Interestingly, we noted increased AV and AL formation in AR20 mice, underscoring the importance of autophagy as a degradative program for many "normal" proteins, including polyQ proteins of non-pathogenic repeat lengths. Autophagy pathway impairment was also found to be a key factor in SBMA motor neuron degeneration in vivo. Prior to disease onset, AR YAC CAG100 mice intensified autophagy pathway activity, as they displayed a significantly increased number of ALs compared to AR YAC CAG20 and non-transgenic mice. However, failure of autophagy pathway progression coincided with manifestation of disease, as symptomatic AR100 mice showed diminished numbers of ALs, despite a significant increase in AVs. Our results indicate that initially autophagy is a beneficial pathway in SBMA, and autophagy likely remains neuroprotective, so long as it is functioning properly. Yet, once autophagy pathway impairment develops, autophagy will contribute to neuron dysfunction and cell death, and this is exacerbated when a proteotoxic stress persists.

## 292

**Autosomal Recessive Hypophosphatemic Rickets is associated with inactivation mutation in the *ENPP1* gene.** *V. Levy-Litan[1,2], E. Hershkovitz[3], L. Avizov[1,2], N. Levental[3], V. Chalifa-Caspi[2], E. Manor[1,4], D. Berkovitz[5], J. Goding[6], R. Parvari[1,2].* 1) Genetics & Virology, Ben Gurion University of the Negev, Beer Sheva , Israel; 2) National Institute of Biotechnology Negev, Ben Gurion Univ, Beer Sheva, Israel; 3) Soroka Medical Center, Beer Sheva, Israel; 4) Genetics Institute, Soroka Medical center, Beer Sheva, Israel; 5) Migal- Galilee Technology Center, Israel; 6) Dept. of Pathology and Immunology, Monash Univ, Victoria, Australia.

We have characterized an enlarged consanguineous Bedouin family displaying Autosomal Recessive Hypophosphatemic Rickets. The main clinical manifestations of the disease were short stature and bowing legs. Laboratory findings were hypophosphatemia, hyperphosphaturia, elevated plasma alkaline phosphatase and normal levels of FGF23, serum calcium, calcium excretion, PTH and vitamin D metabolites. The structure and size of the family were suitable for positional cloning of the causing gene through linkage study. After excluding linkage to the three known genes causing this disease, we carried out a genome-wide linkage analysis using the 250K Affymetrix microarrays (SNPs), and located the chromosomal region containing the mutated gene. We identified the mutation, located in the ecto-nucleotide pyrophosphatase/phosphodiesterase 1 (*ENPP1*) gene. ENPP1 generates inorganic pyrophosphate (PPi), an essential physiologic inhibitor of calcification. The mutation is a single nucleotide change that causes substitution of a strictly conserved amino acid. The population of 236 healthy Bedouins was tested by DHPLC and polymorphism was excluded. The function of the mutation was tested by transfection of a pSVT7 expression vector harboring the mutated full-length ENPP1 coding sequence into COS cells and measuring the NPP activity in comparison to the activity in cells transfected with the normal sequence vector. We found that the mutation completely abolished NPP activity. Previously described inactivating mutations in this gene caused aberrant ectopic calcification disorders. Our results are therefore surprising and cannot be explained in terms of known functions of this enzyme.

† Trainee Award Finalist

## 293

**Founder Mutation Causing Recessive Type VIII Osteogenesis Imperfecta in West Africans and African-Americans is Contemporaneous with Atlantic Slave Trade.** *W.A. Cabral[1], A.M. Barnes[1], A. Adeyemo[2], K. Cushing[3], D. Chitayat[4], F.D. Porter[5], S.R. Panny[6], F. Majid[6], J. Bailey-Wilson[7], L. Brody[3], C.N. Rotimi[2], J.C. Marini[1].* 1) BEMB, NICHD, NIH, Bethesda, MD; 2) CGHD, NHGRI, NIH, Bethesda, MD; 3) GTB, NHGRI, NIH, Bethesda, MD; 4) Hospital for Sick Children, Mount Sinai Hospital, Toronto, Canada; 5) HDB, NICHD, NIH, Bethesda, MD; 6) MD Dept of Health, Baltimore, MD; 7) SGS, NHGRI, NIH, Baltimore, MD.

Recessive osteogenesis imperfecta (OI) is caused by defects in the genes encoding cartilage-associated protein (*CRTAP*) or prolyl 3-hydroxylase 1 (P3H1/*LEPRE1*), which form the ER-resident collagen prolyl 3-hydroxylation complex together with cyclophilin B (*PPIB*). Deficiency of P3H1 causes the severe to lethal bone dysplasia type VIII OI (OMIM #610915). We have previously identified a *LEPRE1* mutation, IVS5+1G>T, in unrelated African Americans (AA) and contemporary West African (WA) immigrants (Nat Genet 39:359-365, 2007). Our screening of gDNA from 3 groups of Mid-Atlantic AAs (Pennsylvania, Maryland, and Washington, DC) determined a carrier incidence of 1/200-300 for this mutation, predicting occurrence of homozygous lethal type VIII OI in about 1/250,000 births in this population. To trace the origin of this mutation, we screened gDNA from more than 1200 contemporary WAs. Nineteen of 1284 unrelated individuals (1.48%) from Nigeria and Ghana were heterozygous carriers, half of whom were from the Yoruba or Ibo ethnic groups of Nigeria. The high carrier frequency for this founder mutation among WAs predicts that this mutation alone would cause recessive OI in 1/18,250 births in WAs, which is equal to the incidence of *de novo* dominant OI in North America. To estimate the age of the mutation, we examined microsatellites and short tandem repeats spanning a 4.2 MB region surrounding the *LEPRE1* gene on chromosome 1p. Disease allele haplotypes were determined for probands and carriers from 10 contemporary WA and 3 AA families. All carriers shared a haplotype of 175-425Kb. Using the linkage disequilibrium analysis method of Rannala & Slatkin (2000), the mutation was estimated to have originated between 385 and 630 years before present (1379-1624 C.E.). This timing is consistent with the model that this West African allele was transported to the Americas during the peak of the Atlantic slave trade (1450-1860 C.E.). Furthermore, expansion of this allele in the West African population was approximately concurrent with transport to the New World, making it difficult to estimate the number of potential carriers transported to the Americas. We are currently exploring several hypotheses to explain the existence of the high carrier frequency for the *LEPRE1* IVS5+1G>T mutation in WAs, including genetic drift, positive selection for heterozygotes and as a “hitchhiker” linked to a selected allele of a neighboring gene on chromosome 1.

## 294

**Selection for blue eyes in Europe and light skin pigmentation in East Asia at OCA2/HERC2.** *M.P. Donnelly, W.C. Speed, J.R. Kidd, A.J. Pakstis, K.K. Kidd.* Dept Gen, Yale Univ Sch Med, New Haven, CT.

OCA2 and HERC2 are two genes on chromosome 15 separated by less than 10 kb. Mutations in this region have been shown to have an effect on pigmentation including causing oculocutaneous albinism type 2. In Europeans, a three SNP haplotype (rs4778138, rs4778241, rs7495174) and three individual SNPs (rs12913832, rs916977, rs1667394) have been associated with blue eyes. We have labeled the three SNP haplotype BEH1. We found that the first individual SNP, rs12913832, was in near complete LD with another SNP (rs1129038). We treat these two SNPs together as a haplotype, BEH2. We also found that the other two individual SNPs were actually in near complete LD with each other and decided to label them BEH3. In East Asians, a SNP (rs1800414) has been identified that is associated with a light skin pigmentation phenotype. We typed these eight SNPs in 64-70 population samples. We then examined worldwide distribution of the four pigmentation alleles. We saw that the light skin allele was at its highest frequency in eastern East Asia, at midrange frequencies in Southeast Asia, and at lower frequencies in western East Asia. It is virtually absent from the rest of the world. BEH1 and BEH3 show very similar global patterns, low frequencies to midrange frequencies in Africa and East Asia, midrange frequencies in India and Eastern Siberia, and midrange to high frequencies in Southwest Asia, Europe, Western Siberia, the Pacific Islands, and the Americas. BEH2 shows a different pattern from the other two. It shows low frequencies in East Africa, India, Eastern Siberia, and the Americas, midrange frequencies in Southwest Asians and Southern Europeans, and high frequencies in Eastern and Northwestern Europe and Western Siberia. We then typed additional SNPs and test each pigmentation allele for selection using the Relative Extended Haplotype Homozygosity (REHH) test. We found that the light skin allele of rs1800414 is under selection in East Asia and that the blue eye allele of BEH2 is under selection in Europe and Southwest Asia. We show light skin pigmentation has been selected for in East Asia. This is likely due to lower UV exposure at the higher latitudes (compared to equatorial Africa) and the need for lighter skin for vitamin D production. We also show that blue eyes are selected for in Europe. This is most likely due to sexual selection, though another unknown effect of this particular allele could be selected for and the blues eyes are a side effect.

## 295

**A Bayesian Hierarchical Clustering algorithm for CNV data : application to the WTCCC CNV experiment.** *N. Cardin[1,2], D. Vukcevic[2], R. Pearson[2], P. Donnelly[1,2], J. Marchini[1,2], The Wellcome Trust Case Control Consortium.* 1) Dept Statistics, Univ Oxford, Oxford, Oxfordshire, United Kingdom; 2) The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN, United Kingdom.

There has been growing consideration of the possibility that structural variation in the form of copy number variants (CNVs) contribute to human phenotypic variation and there is considerable interest in developing genome-wide strategies for CNV association. The WTCCC has designed an Agilent array that aims to assay ~10,000 CNVs across the genome and the array has been used to assay 19,000 samples from 8 diseases. A central challenge in the analysis of this complex dataset is the determination of polymorphic CNVs in the samples as well as accurate and robust CNV genotype calling and testing. We have developed a Bayesian hierarchical clustering algorithm designed to work with multi-probe CNV assays that meets these challenges and has several novel features. Firstly, our model consists of hierarchical mixtures of t-distributions, together with outlier classes, and allows variation between cohorts that protects against differential bias when it occurs, but at the same time pools information across cohorts. Secondly, estimation of the number of genotype classes in each CNV is facilitated by our use of Bayesian methods and a number of novel metrics that protect against model mispecification of intensity variation. Finally, our model produces posterior estimates of copy number class in each sample and we have developed Bayesian testing methods that account for the inherent uncertainty in these calls. We illustrate the utility of these methods using real data from the WTCCC study.

## 296

**Targeted Genotyping of Human Copy Number Polymorphisms.** *C.D. Campbell[1], J.M. Kidd[1], N. Sampas[2], P. Tsang[2], C. Baker[1], E.E. Eichler[1,3].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Agilent Technologies, Santa Clara, CA; 3) Howard Hughes Medical Institute, Seattle, WA.

Copy number polymorphisms (CNPs) are common in human populations (frequency > 1%), and a few CNPs have been strongly associated to human diseases including lupus, psoriasis, and Crohn's disease. Furthermore, genome-wide association studies of SNPs have, thus far, explained a small fraction of the genetic risk to most complex diseases suggesting that other forms of genetic variation, including CNPs, may be important. Therefore, we sought to develop a platform to genotype known CNPs. We assembled a comprehensive list of 5654 non-overlapping CNP regions defined at a high-resolution by sequencing or high-density SNP microarrays. These include 2865 polymorphic novel insertions that are not in the human genome reference sequence. Although high-density SNP microarrays have proved to be an effective and high-throughput approach to genotype CNPs, 44% of our targeted regions (1234 of the 2789 non-novel sequences loci) have little probe coverage (<5 probes) on the highest density commercial SNP arrays. In addition, the 2865 polymorphic novel insertions are not targeted by any platform based on the human genome reference sequence, thus 72% of the targeted 5654 CNP regions are not adequately captured by existing SNP array platforms. We have designed a comparative genomic hybridization (CGH) microarray targeting these CNPs and were able to cover 81% (>5 probes) of the CNP regions. We have tested this microarray on 25 HapMap individuals. Based on copy number estimates from orthologous technologies for 1024 CNPs, our microarray hybridization data are consistent with the previous data for 767 sites (75%). In addition, we have developed an approach to determine diploid copy number from microarray hybridization data. We have tested this method on a subset of the polymorphic novel insertions, and we could clearly distinguish copy-number classes for 475 out of 1148 polymorphic sites tested (41%). For the remaining 673 sites, we were able to generate lower confidence copy-number estimates. In our modestly sized pilot study, we have identified 11 CNPs with potential large frequency differences between populations. We are in the process of assaying many of the Phase I and Phase III HapMap samples to identify further population-specific CNPs between populations and to potentially identify recurrent CNPs. These results suggest that our platform for CNPs will be complementary to high-density SNP arrays for interrogating common genetic variation for association to human disease.

## 297

**The Contribution of Copy Number Variants and SNPs to Osteoporotic Fractures and Osteoporosis Related Phenotypes. The Framingham Study.** *YH. Hsu[1,2], DE. Karasik[1], H. Zhou[1], CL. Cheung[1], Y. Zhou[3], S. Demissie[3], LA. Cupples[3], DP. Kiel[1].* 1) Inst Aging Res, Hebrew SeniorLife and Harvard Medical School, Boston, MA; 2) Molecular and Integrative Physiological Sciences Program, Harvard School of Public Health, Boston, MA; 3) Biostat. Dept., Sch Public Health, Boston University, Boston, MA.

Previously, we have reported results from a genome-wide association study of osteoporosis related phenotypes (bone mineral density and hip geometry). Despite novel candidate genes were identified, these newly discovered genetic variants together only explained a small proportion of the phenotypic variation. Copy number variations (CNVs) has recently been identified as a major source of genetic diversity, but a comprehensive understanding of the phenotypic effect of CNVs is beginning to emerge. We estimated CNVs from the Affymetrix 550K SNP chips in the ~9200 individuals from the Framingham Study. Three distinct algorithms were used parallel to segment CNV regions from diploid DNA intensity data. We excluded individuals with either duplication or loss of the entire chromosome for at least one chromosome due to the artificial synthesis on cell immortalization. We identified more than 114,780 CNVs, and the sizes of these CNVs were ranged from 1.8 Kb to 25Mb. Only 1% of the identified CNVs were common (frequency > 1%)in this Caucasian population. The most common CNV was found to be on chr17 with frequency > 25%. Majority of the common CNVs can be found in both offsprings and parents. We performed genome-wide association analyses using 2.5 million imputed SNPs and common CNVs to localize susceptible genes for incident osteoporotic fractures during an average of 20-year follow-up in the Framingham Osteoporosis Study. We conducted both population- and family-based analyses adjusted for several covariates. For population-based analyses, a Cox-proportional hazard model with a random effect that accounted for individual correlations within families was applied. The principal components (PCs) estimated by the EIGENSTRAT were adjusted in the models. Family-based association tests were used to further confirm results. Several genome-wide ($p < 5x10^{-8}$) and strong associations were observed ($p < 5x10^{-7}$), i.e. CNVs on Chr1p36, Chr20q33; and SNPs on FBN1 gene, Chr1p34 and Chr8p23. Chromosome 1p36 has been reported by linkage studies, but the genes responded for this QTL has not been found. Of note, none of the SNPs in these top associated CNVs were also found to be associated with osteoporotic fractures. The top associated CNVs and SNPs are being further evaluated in a replication study. In conclusion, our results by utilizing sequence variants from both CNVs and SNPs reveal novel candidate genes and pathways to further elucidate the etiology of osteoporotic fractures.

## 298

**Genome-wide CNV association study identifies a Rare Deletion on Chromosome 6p25 Associated with Osteoporotic Fractures.** *K. Estrada[1], M.J. Peters[1], Y-H. Hsu[2,3], H. Eussen[4], A. de Klein[4], H.A. Pols[1,5], J.B. van Meurs[1], T.A. Knoch[6], D. Karasik[2], D.P. Kiel[2], A.G. Uitterlinden[1,5], F. Rivadeneira[1,5], The GEFOS consortium.* 1) Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands; 2) Institute for Aging Research, Hebrew SeniorLife and Harvard Medical School, Boston, MA, and the Framingham Study, Framingham, MA, USA; 3) Molecular and Integrative Physiological Sciences Program, Harvard School of Public Health, Boston, 02115 MA, USA; 4) Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands; 5) Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 6) Biophysical Genomics & Erasmus Computing Grid, Erasmus Medical Center, Rotterdam, The Netherlands.

Background: Osteoporosis is a systemic skeletal disease characterized by reduced bone mineral density (BMD) and increased susceptibility to fracture. Copy Number Variants (CNVs) are a form of structural variation where individuals have gains or losses of DNA. Both common and rare CNVs have been observed in the human genome, they may affect the function of genes and potentially influence a number of traits and diseases. Aim: To evaluate the association of CNVs with the risk for OF using large population-based studies with genome-wide CNV data and clinical records of incident OF. Methods: First, we obtained the normalized intensity data of 5,824 individuals of the Rotterdam Study (RS-I), a population based cohort of individuals of age 55 years and over, genotyped with the Illumina 550K Array. After quality control 49,229 CNVs (mapping to 26,162 genomic locations) were identified using computational methods designed to segment CNV regions from diploid DNA intensity data. We then performed a genome-wide CNV association study in 809 OF cases assessed on 5178 individuals. Secondly, after correction for multiple testing we replicated the only significantly associated locus in two additional cohort studies: the Rotterdam Study extension (RS-II, n=2157, 181 cases) and the Framingham Osteoporosis Study (FOS, n=3513, 367 cases) genotyped with the Illumina 550K Array and the Affymetrix 500K Array, respectively. Results: A 210 kb deletion located on chromosome 6p25 was found in seven individuals of the RS-I population (0.1%). This deletion was associated with the risk of OF and 6 of the 7 carriers were women who suffered from OF during follow-up (OR:32.6 [95%CI 3.9-271], p=0.001). The prevalence of this deletion and the association with fractures was replicated in two follow-up studies: RS-II (OR:5.5 [95%CI 1.0-30.3], p=0.04) and FOS (OR:2.9 [95%CI 0.9-9.0], p=0.06). Meta-analysis using RS-I, RS-II and FOS yielded a combined 5-fold increase in the risk of fracture (OR: 5.1 [95%CI 2.1-12.1, $p=3x10^{-6}$). Conclusions: These results suggest that deletions in the 6p25 locus predispose to higher risk of OF. Interestingly, case reports with 1-13 Mb deletions in this region include skeletal phenotypes such as epiphysial displasia and craniofasial deformations. This is a first step towards the evaluation of the role of rare variants in a complex disease such as osteoporosis. Replication of these findings in additional independent populations is underway.

## 299

**A large scale study of copy number variation implicates the genes *DMXL1* and *TULP3* in the etiology of primary open angle glaucoma.** *LK. Davis[1], KJ. Meyer[1], EI. Schindler[1], JS. Beck[2], DS. Rudd[1], AJ. Grundstad[3], TE. Scheetz[3,4], TA. Braun[3,4], JH. Fingert[1,4], JC. Folk[4], SR. Russell[4], TH. Wassink[1,5], VC. Sheffield[1,2,6], EM. Stone[1,4,6].* 1) Interdisciplinary Genetics Program, Univ Iowa, Iowa City, IA; 2) Department of Pediatrics, Univ Iowa, Iowa City, IA; 3) Center for Bioinformatics and Computational Biology, Univ Iowa, Iowa City, IA; 4) Ophthalmology and Visual Sciences, Univ Iowa, Iowa City, IA; 5) Department of Psychiatry, Univ Iowa, Iowa City, IA; 6) Howard Hughes Medical Institute, Univ Iowa, Iowa City, IA.

This study sought to investigate the role of rare copy number variation (CNV) in age-related disorders of blindness, with a focus on primary open angle glaucoma (POAG). Here we present data from a whole-genome copy number screen in a large cohort of 400 individuals with (POAG), 400 individuals with age related macular degeneration (AMD), and 100 age-matched controls (AMCs). DNA samples from patients and controls were tested for copy number variation using a combination of 500K Affymetrix SNP GeneChip Microarrays and 5.0 Affymetrix SNP GeneChip Microarrays. The signal intensity data generated from these arrays was then analyzed with multiple CNV detection programs including CNAG v2.0, PennCNV, and dChip. After sample-based and CNV-based quality control measures were implemented, a complete set of over 11,000 CNVs was generated consisting of CNVs detected with all programs on all platforms. From this data, a "stringent criteria set" of 1,105 CNVs was developed, consisting of CNVs called by two programs or present on two platforms in a single individual. A set of "POAG enriched" CNVs was then generated by comparison of the POAG, AMD and AMC data. "POAG enriched" CNVs were drawn from both the stringent criteria POAG set and the total POAG data set and were required to be absent completely from the AMD and AMC data sets. A total of nine POAG enriched variants met criteria for qPCR or cGH array validation. Six of these variants were confirmed and met criteria for follow-up. Of these, *DMXL1*, *TULP3* and *PAK7* each provide independent lines of evidence suggesting roles for these genes in POAG. The CNVs presented here are exceedingly rare and are not found in the Database of Genomic Variants or a sample of 862 controls run on Affymetrix 6.0 arrays (personal communication, C. Marshall). All three genes show high expression in the eye, including the retina, ocular nerve and ocular nerve head. Additionally, CNV locations of *DMXL1* and *PAK7* overlap previously identified linkage signals for glaucoma on chromosomes 5p23.1 and 20p12, respectively (Wiggs, et al., 2004; Pang,et al., 2006). Here we present data consistent with the hypothesis that rare copy number variation plays a role in the development of POAG.

## 300

**Correlation of Abnormal DNA Copy Number Analyses and Cardiac Disease in a Pediatric Population.** *S.R. Lalani[1], J.L. Jefferies[2,3], X. Wang[1], F.M. Boricha[1], L.W. Patterson[2], C. Shaw[1], A.L. Beaudet[1], S.W. Cheung[1], J.W. Belmont[1].* 1) Dept of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Section of Pediatric Cardiology, Texas Children's Hospital, Houston, TX; 3) Division of Adult Cardiovascular Diseases, the Texas Heart Institute at St. Luke's Episcopal Hospital, Houston, TX.

**Background:** Congenital cardiovascular malformations (CVM) are the most common birth defects and occur in 5-8/1000 live births. The causative factors contributing to most sporadic CVM are not known. Limited case series have suggested a large role for genomic disorders in the pathogenesis of CVM, but there have been no extensive clinic and hospital-based studies to gauge the clinical utility of screening patients with CVM accompanied by other birth defects. We designed a study to delineate *de novo* and inherited segmental aneusomies contributing to CVM and define critical genes within these regions that are important for cardiac patterning and myocardial function. **Methods:** We reviewed the clinical data on 4951 individuals at Texas Children's Hospital, between February 2004 and February March 2009, who had an array-CGH analysis. All echocardiograms performed for suspected CVM in this cohort were reviewed. BAC arrays were employed in 1541 cases and the remaining were studied with custom oligonucleotide arrays with either 44K or 105K features. A total of 7993 transthoracic echocardiograms were performed on 1657 patients with CVM. The array-CGH data were compared to 3294 age-matched controls without cardiac defects. **Results:** Of all the 1657 patients with CVM, 398 (24%) had a cytogenetic aberration, including aneuploidies and copy number variations (CNV). About 122 (30%) of these individuals either had trisomy 21, trisomy 18, Turner syndrome, DiGeorge syndrome, Williams syndrome or 1p- syndrome. Parental studies were completed in 100 cases of submicroscopic CNVs (25%); of these, 38% were found to be *de novo* events. Recurrent, often *de novo* CNVs were identified at certain loci including 15q14 with *ACTC1* deletion, 17q21.31 with *MAPT* deletion, 2q22.3 with *ZEB2* deletion, 20p12.3 with loss of *BMP2,* and 16q24.3. The range of cardiac defects was highly diverse, encompassing all major classes of malformations. The comparison of CNVs in this cohort with age matched control of over 3000 cases, has delineated several relevant CNVs associated with CVM. **Conclusions:** Cytogenetic aberration is a common finding in individuals with significant cardiovascular defects, occurring in nearly 24% of our cohort. Patients with complex cardiovascular malformations and those with accompanying birth defects or mental retardation are likely to have an identifiable genomic imbalance. DNA copy number analysis is indicated in such patients both for diagnostic and counseling purposes.

## 301

**Whole-genome array CGH in 500 patients with epilepsy identifies novel susceptibility loci.** *H. Mefford[1,2], I. Helbig[3], K. Buysse[4], C. Baker[2], S. von Spcizak[3], H. Muhle[3], P. Ostertag[3], M. Guipponi[5], U. Stephani[3], E. Eichler[2,6].* 1) Pediatrics, University of Washington, Seattle, WA; 2) Genome Sciences, University of Washington, Seattle, WA; 3) Neurogenetics, University Medical Center Schleswig-Holstein, Kiel, Germany; 4) Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium; 5) Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 6) Howard Hughes Medical Institute, University of Washington, Seattle, WA.

Epilepsy is one of the most common neurological disorders in humans, with a prevalence of 1% and a lifetime incidence of 3%. Several genes have been identified in rare autosomal dominant and severe sporadic forms of epilepsy, but the genetic cause is unknown in the vast majority of cases. We recently identified deletions of 15q13.3, previously associated with intellectual disability (ID), autism and schizophrenia, as the first recurrent copy number variant (CNV) and most prevalent known genetic risk factor associated with epilepsy. Although CNVs are known to play an important role in the genetic etiology of many neurodevelopmental disorders, including ID, autism and schizophrenia, the role of CNVs other than 15q13.3 deletions in epilepsy is unknown. We hypothesize that, as in other neurodevelopmental disorders, CNVs play an important role in the genetic basis of epilepsy. We have performed whole-genome oligonucleotide array CGH with an array containing 135,000 probes on a cohort of 500 individuals with various idiopathic epilepsies. We targeted ~100 regions that are known or predicted rearrangement "hotspots" at high density (1 probe/2.5-5 kb), with remaining probes distributed throughout the genome at an average spacing of 1 probe per 38 kb. We detected one or more rare genic CNVs in 7.6% of affected individuals that are not present in 2493 controls; four individuals had two or more rare CNVs. We identified CNVs in genes previously implicated in other neurodevelopmental disorders, including two deletions in *AUTS2*, and one deletion in *CNTNAP2*. Importantly, we find that 2.2% of patients carry deletions of rearrangement hotspots previously associated with ID, autism or schizophrenia. These include deletions of 15q13.3 in 0.8% of patients, validating our previous studies. Five individuals (1%) have deletions of 16p13.11, a significant enrichment compared to controls (5/500 vs 0/2493, p=0.0001, Fisher's exact). 16p13.11 deletions appear to be at least as frequent as deletions of 15q13.3 and represent a new susceptibility locus for epilepsy. Two individuals have deletions of 1q21.1, a locus associated with ID and schizophrenia. In summary, we find rare CNVs in 7.6% of 500 epilepsy patients, confirm the role of 15q13.3 deletions in generalized epilepsy, and identify deletions of 16p13.11 as a new susceptibility locus for epilepsy. Our findings suggest a common etiology for seemingly diverse disease such as ID, autism, schizophrenia and epilepsy.

## 302

**Which CNV analysis strategies can identify a well-known disease causing CNV?** *N. Pankratz[1], A. Dumitriu[2], J.C. Latourelle[2], J.B. Wilk[2], A.L. DeStefano[2], K.N. Hetrick[3], R.H. Myers[2], T. Foroud[1], the PSG-PROGENI and GenePD Investigators, Coordinators and Molecular Genetic Laboratories.* 1) Indiana University, Indianapolis, IN; 2) Boston University, Boston, MA; 3) Johns Hopkins, Baltimore, MD.

Copy number variants (CNVs) are known to cause Mendelian forms of Parkinson disease (PD) - most notably in SNCA and PARK2. PARK2 has a recessive mode of inheritance; however, recent evidence demonstrates that a single CNV in PARK2 (but not a single missense mutation) may increase the risk for PD. We recently performed a GWAS for PD that excluded individuals known to have causative mutations, including anyone known to have two PARK2 mutations (those with one were still included). Data were re-clustered using only samples with high quality intensity data. Markers that appeared to co-hybridize to the X or Y chromosome were removed; all other markers were included. Data were analyzed using PennCNV, and CNVs with a confidence score >10, spanning at least 5 SNPs, and outside regions of known instability (telomeres, centromeres and immunoglobin regions) were analyzed. The final sample included 816 cases and 856 controls. Two cases and no controls harbored a CNV in SNCA (both duplications of the entire gene), while 22 cases and 8 controls harbored CNVs in PARK2. Of the 12 CNVs tested using quantitative PCR, 12 were validated (100%). When CNVs were analyzed in PLINK, using either a single position analysis or a 400kb window analysis, PARK2 was not significant ($p=0.19$; all p-values reported are empirically derived and genome-wide). Similar to previous publications, we analyzed large, rare CNVs (spanning ≥20 SNPs, ≥100kb and not present in the Database for Genomic Variants (DGV)) and common copy number polymorphisms (CNPs) separately. The large, rare analysis failed to identify PARK2 ($p=0.21$), since deletions near exons 3 and 4 are found in HapMap samples and therefore the 12 cases and 1 control overlapping these CNPs were removed from analysis. The CNP analysis also failed ($p=0.19$), because those same variants were the only CNVs in the region included. Regions within PARK2 did reach genome-wide significance when analyses were limited to large CNVs, independent of their frequency in DGV ($p=0.007$) or when limited to CNVs overlapping RefSeq genes ($p=0.03$). Using the gene-based approach, genome-wide significance was also noted for two novel genes: DOCK5 ($p=0.0001$) and USP32 ($p=0.0006$). Like PARK2, USP32 is involved in the ubiquitin proteasome system. Since the rare versus common approach would not have identified DOCK5, USP32, or PARK2, alternative analysis strategies, such as this gene-centric approach, may be useful in identifying certain deleterious CNVs.

## 303

**Host Haplotype associations with HBV Vaccine-Induced Immunity.** *S.M. Williams[1], K.K. Ryckman[1], K. Fielding[2], A.V. Hill[3], M. Mendy[4], P. Rayco-Solon[2,5], G. Sirugo[5,6], M.A. van der Sande[5,7], P. Waight[5], H.C. Whittle[5], A.J. Hall[2], B.J. Hennig[2].* 1) Ctr Human Gen Res, Vanderbilt Univ Med Ctr, Nashville, TN; 2) London School of Hygiene & Tropical Medicine, London, UK; 3) Wellcome Trust Ctr Hum Genet, Oxford, UK; 4) The Jenner Institute, Oxford University, Oxford, UK; 5) Medical Research Council Laboratories, The Gambia; 6) Medical Genetics Unit, Ospedale S. Pietro FBF, Rome, Italy; 7) National Institute for Public Health and the Environment (RIVM), Bilthoven, Netherlands.

Introduction: Hepatitis B virus (HBV) infection remains a significant health burden world-wide, but vaccines can decrease this problem. Evidence from family and twin studies indicate that host genetic factors play a role in modulating HBV vaccine-induced immunity. We previously tested for correlation with vaccine-induced peak antibody levels (anti-HBs) that are predictive of long-term vaccine efficacy as well as protection against infection and persistent carriage. Our earlier study (PLoS ONE. 2008 Mar 26;3(3):e1898) concentrated on single nucleotide polymorphism (SNP) analysis and we reported associations with IFNG, MAPK8, and CD58. Methods: A total of 688 SNPs from 117 genes were included for a detailed 2, 3 and 4 sliding window haplotype analysis in our Gambian dataset. Analysis was performed on two separate samples: one consisted of 197 unrelated individuals and another consisted of 454 individuals from 174 families. Additionally, these two samples were analyzed for haplotype associations together, for a total of 651 individuals. We carried out both global and individual haplotype test, with adjustments for covariates. Generated haplotypes were tested for association with peak anti-HBs level and these findings were compared with previous single SNP results. Results: There were five genes that had at least one significant haplotype in the unrelated and family analysis with a p-value < 0.1 and significance in the combined analysis at a p-value < 0.01. These genes were: CD44, CD58, CDC42, IL19 and IL1R1. The most significant results in the unrelated and family data combined was for a 3-SNP haplotype (rs353644-rs353630-rs7937602) in CD44 (adjusted p = 9.1x10-5). A 3-SNP haplotype (rs1414275-rs11588376-rs1016140) in CD58 was also associated with anti-HBs level (adjusted p = 0.008). Earlier single SNP analysis did not exhibit significant association with CDC42, IL19 and IL1R1. Conclusions: We have identified strong haplotype effects in five genes and have replicated these effects in an independent dataset. It is of note that the haplotype analyses identified genes found in our earlier study and several new ones.

## 304

**Development of a pharmacogenetic test that predicts response to inhaled corticosteroids in asthma patients.** *A.C. Wu[1,2], B.E. Himes[3,4,5,6], K. Tantisira[6], S.T. Weiss[5,6].* 1) Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim HealthCare, Boston, MA; 2) Children's Hospital, Boston Division of General Pediatrics, Boston, MA; 3) Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA; 4) Children's Hospital Informatics Program, Boston, MA; 5) Harvard Partners Center for Genetics and Genomics, Boston, MA; 6) Channing Laboratory, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

Inhaled corticosteroids (ICS) are the most commonly used controller medications for asthma, yet up to 20% of asthma patients do not benefit from ICS therapy and are at risk for potential side effects. Our goal was to develop a predictive test of response to ICS using genetic data from 169 Caucasian subjects who were randomized to the ICS arm of the Childhood Asthma Management Program (CAMP) drug trial. Subjects were considered to be non-responsive to ICS if they experienced one or more hospitalizations or emergency department visits during the four years of the drug trial. Whole-genome genotype data was collected for subjects using Illumina's HumanHap550 Genotyping BeadChip. Bayesian networks, multivariate models that are able to account for simultaneous associations and interactions among variables, were used to create the predictive model. Because of the large number of single nucleotide polymorphisms (SNPs) available, the subset of those with a Bayes factor greater than 500 (n=14) were selected to learn the Bayesian network. This Bayes factor threshold indicated that each of the 14 SNPs was over 500 times more likely to be related to ICS than to be independent from it. Predictive accuracy of the model was assessed using 10-fold cross-validation in which all subjects were randomly split into 10 groups, and each group was used as an independent dataset to predict response to ICS while the remaining 9 groups were used to quantify model parameters. Predicted and observed response to ICS were compared using areas under receiver operating characteristic curve (AUROC). The AUROC for 10-fold cross-validation was 0.97, indicating that our model may generalize to independent populations and contain SNPs that are relevant to the biology of response to ICS. Testing in an independent population in ongoing.

## 305

**A prospective study of HLA-B*1502 genotyping in preventing carbamazepine-induced Stevens-Johnson syndrome and toxic epidermal necrolysis.** *P. Chen[1], C.Y. Shen[1], J.J. Lin[2], C.T. Ong[3], S.L. Wu[4], P.J. Tsai[1], W.H. Chen[1], W.H. Chung[5], S.I. Hung[1,6], C.F. Chang[7], C.H. Chen[1], S.Y. Lin[1], H.P. Chuang[1], L.C. Huang[1], Y.J. Chen[1], Y.T. Chen[1], Taiwan SJS consortium.* 1) Inst Biomedical Sci, Academia Sinica, Taipei, Taiwan; 2) Chu Shang Show Chwan Hospital, Nan-tou, Taiwan; 3) Chia-Yi Christian Hospital, Chia-Yi, Taiwan; 4) Changhua Christian Hospital, Chang-Hua, Taiwan; 5) Department of Dermatology, Chang Gung Memorial Hospital, College of Medicine, Chang Gung University; 6) Institute of Pharmacology, School of Medicine, National Yang-Ming University; 7) PharmiGene Inc., Taiwan.

Stevens-Johnson syndrome (SJS) and its related disease toxic epidermal necrolysis (TEN) are two of the most serious adverse reactions caused by drugs. Carbamazepine (CBZ), used primarily in the treatment of epilepsy and bipolar disorder, is the commonest cause of SJS/TEN in Southeast Asian countries. We previously reported that carbamazepine (CBZ)-SJS/TEN was strongly associated with the human leukocyte antigen (HLA) B*1502 in Han Chinese. This association was subsequently confirmed in Thailand, Hong Kong, and descendants of Southeastern Asians where HLA-B*1502 is prevalent. The risk of CBZ-SJS/TEN was significantly higher in patients carrying the HLA-B*1502 allele than those without the allele (odds ratio: 1357, 95% CI: 193- 8838; $p = 1.6 \times 10$ -41). A prospective study aimed at preventing carbamazepine induced SJS/TEN by identifying individuals at risk using HLA-B*1502 genotyping was carried out in Taiwan. For ethic consideration, the study was designed as a nonrandomized trial and used historical incidence as control. From January 2007 up to May 2009, a total of 3110 subjects have been recruited from 25 participating hospitals in Taiwan. 7.9% of patients were found to carry the HLA-B*1502 allele and no CBZ were prescribed; instead, they were given either alternative medications or remained the same prior-study medications. Of all enrolled patients who completed in 2-months follow-up, 5.3% developed mild and transient skin rashes; five however, had more wide-spread rashes and were hospitalized (final diagnosis, 3 maculopapular eruption, 1 hypersensitivity syndrome, 1 urticaria). None of the 3110 participating patients developed SJS/TEN; this is in contrast to the estimated 8 cases of CBZ-SJS/TEN in the historical control (0.25% of CBZ users) (p value=0.0013; Fisher's two-tailed exact tests). Our data suggested that application of HLA-B*1502 genotyping as a screening tool for patients taking CBZ can effectively reduce the incidence of these life-threatening adverse drug reactions.

## 306

**Genetic Variation at the NFATC2 Locus Increases Edema in the Diabetes REduction Assessment with Ramipril and Rosiglitazone Medication (DREAM) Study.** *S.D. Bailey[1], C. Xie[2], R. Do[1], A. Montpetit[3], B. Keating[4], S. Yusuf[2], H. Gerstein[2], J. Engert[1,5], S. Anand[2], DREAM Investigators.* 1) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 2) Population Health Research Institute, McMaster University, Hamilton, Ontario, Canada; 3) McGill Univerisity and Genome Quebec Innovation Centre, Montreal, Quebec, Canada; 4) The Institute for Translational Medicine and Therapeutics, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; 5) Department of Medicine, McGill University, Montreal, Quebec, Canada.

Thiazolinediones (TZDs) are a class of drugs used to treat type-2 diabetes (T2D) that derive their insulin sensitizing effects from the activation of the peroxisome proliferator-activated receptor (PPAR)-$\gamma$. TZDs have a proven therapeutic effect and have been shown to reduce the incidence of T2D and to increase the reversion to normal glycemia in glucose intolerant individuals. However, the use of TZDs has been associated with several adverse effects, namely: edema, weight gain and congestive heart failure (CHF). We hypothesized that some of these adverse outcomes are genetic in origin and that they are the result of an interaction between a genetic variant(s) and TZD treatment. We have genotyped the entire DREAM study sample (n=5269) with a customized vascular disease 50k single nucleotide polymorphism (SNP) array, the IBC (ITMAT/Broad/CARE) chip. The DREAM trial is a multi-ethnic study of individuals at high risk of developing T2D, recruited from 16 countries randomized into four groups following a 2x2 design: placebo only, rosiglitazone and placebo, ramipril and placebo, and both drugs combined. The IBC chip assays 50,000 SNPs in approximately 2100 candidate genes selected for their involvement in cardiovascular disease and metabolism and thus gives us fine scale coverage of the genetic variation present in these genes. One SNP in the nuclear factor of activated T-cells calcineurin dependent 2 gene (NFATC2) was significantly associated with edema in the rosiglitazone treatment arm of the DREAM trial (n=1969, combined p= $3.973 \times 10^{-5}$). The observed association was driven primarily by the European part of the study (n=1139) (OR = 1.80, CI=1.43-2.25, p= $4.36 \times 10^{-7}$). European individuals heterozygous or homozygous for the risk allele had a higher rate of edema (adjusted HR=1.65, p=$5.31 \times 10^{-5}$ and adjusted HR=2.62, p= $2.36 \times 10^{-4}$ respectively). We believe this is the first demonstrated genetic interaction with rosiglitazone resulting in edema described in a randomized controlled drug trial. Identifying genetic variants that interact with TZDs to produce the associated complications may elucidate the pathways involved and lead to preventive strategies among diabetics.

## 307

**Genome-wide association study identifies novel genomic regions associated with drug-induced Long QT Syndrome.** *D.M. Roden[1], S. Kaab[2], M. Sinner[2], P. Kannankeril[11], A. Wilde[3], C. Bezzina[4], E. Schulze-Bahr[5], P. Guicheney[6], N. Bishopric[7], R. Myerberg[8], J.J. Schott[9], A. Pfeufer[10], Y. Nakamura[12], T. Tanaka[12], C.R. Ingram[1], S. Carter[1], Y. Bradford[13], A.L. George Jr[1], M.D. Ritchie[13], D.C. Crawford[13].* 1) Department of Medicine and Pharmacology, Vanderbilt University, Nashville, TN; 2) Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-University, Munich, Germany; 3) Department of Clinical and Experimental Cardiology Academic Medical Centre, Amsterdam/Interuniversity Cardiology Institute of the Netherlands, Utrecht, the Netherlands; 4) Experimental and Molecular Cardiology Group, Department of Experimental Cardiology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; 5) Department of Cardiology and Angiology, Hospital of the University of Münster, Münster Germany; 6) INSERM, Institut de Myologie, Université Pierre et Marie Curie, Institut Fédératif de Recherche 14 and Groupe Hospitalier Pitié-Salpêtrière, Service de Cardiologie, Hôpital Lariboisièr, and Service de Biochimie; Departments of Medicine, Division of Cardio; 7) Pediatrics, Miller School of Medicine, University of Miami, Miami, Florida; 8) INSERM, UMR915, l'institut du thorax; Université de Nantes; CNRS ERL3147; and CHU Nantes, l'institut du thorax, Service de Cardiologie, Nantes, France; Institut für Humangenetik, Technical University of Munich, Munich, Germany;; 9) Helmholtz Zentrum Munich, German Research Center for Environmental Health, Neuherberg, Germany; 10) Department of Medicine, Cardiology, Ludwig-Maximilians-Universität, Munich, Germany; 11) Department of Pediatrics, Vanderbilt University, Nashville, TN; 12) Center for Genomic Medicine, RIKEN Yokohama , Yokohama City, Kanagawa, Japan; 13) Department of Molecular Physiology & Biophysics and Center for Human Genetics Research, Vanderbilt University, Nashville, TN.

A serious side-effect of anti-arrhythmic drugs is drug-induced long QT syndrome (diQTS), a marked prolongation of the QT interval with torsades de pointes. Incidence of diQTS is as high as 5% of patients exposed to anti-arrhythmic drugs. Yet, despite this relatively high incidence, large collections of meticulously phenotyped cases and controls necessary for genome-wide association studies (GWAS) have been lacking for diLTQS. To fill this void, seven centers from the United States and Europe have formed a collaboration to collect cases ascertained by arrhythmia specialists. For this study, cases were defined as patients with diLQTS and controls were defined as patients exposed to ibutilide (excluding patients with 2 hour deltaQTc>50) and drug-exposed patients (excluding patients with Bazett QTc>500). A total of 1,025 DNA samples from this collaboration were genotyped for the Illumina Infinium 610Quad BeadChip in collaboration with RIKEN. Nine DNA samples were of insufficient quality and were omitted from subsequent analyses. Of the 592,532 SNPs targeted for genotyping, 28,137 SNPs were omitted for low genotyping efficiency or call rates (≤90%), 3,484 SNPs were omitted for being monomorphic in the dataset, and 38,242 SNPs were omitted for having a minor allele frequency ≤1%. Hardy Weinberg Equilibrium (HWE) was calculated for each SNP, and SNPs that deviated from HWE were noted but not excluded from the analysis. For the final set of 554,6261 SNPs, single SNP tests of association assuming an additive genetic model were conducted using PLINK in 183 cases and 519 controls of European ancestry. Tests of association were also performed adjusting for age and sex. We identified one SNP associated with diLQTS at p<7.0x10-8 on chromosome 9 (ORadj = 2.28; 95% CI: 1.69-3.08). We also identified four other associations at a significance level of p<9.0x10-7 on chromosomes 9, 11, and 19 and ten other associations at a significance level of p<9.0x10-6 on chromosomes 7, 9, 13, 19, 21. Of the 15 SNPs most significantly associated with diLQTS, only three were located within genes, none were near known miRNAs, and none were located in a candidate gene or region for diLQTS. In summary, we have identified potential novel genomic regions associated with diLQTS. Further studies are needed to replicate these findings and to determine the function of these new candidate regions to better define the role individual genetic variants have on the development of diLQTS.

## 308

**Identification of differentially expressed miRNAs in mouse cortical neurons with or without lithium treatment.** *H. Chen[1], N. Wang[2], M. Yasuda[2], H. Umemori[2], M. Burmeister[1,2], M. McInnis[1].* 1) Department of Psychiatry, University of Michigan Medical School, Ann Arbor, MI; 2) Molecular and Behavioral Neuroscience Institute, University of Michigan Medical School, Ann Arbor, MI.

Lithium (Li) is commonly used for the treatment of bipolar disorder (BPD). The molecular basis of its treatment action has not been completely understood. Recently, miRNAs have been recognized as important players in brain development and in regulation of synaptic plasticity. Li has been shown to affect a number of miRNAs in adult rat hippocampus (Zhou, et al., 2009). However, little is known about time course of Li-induced changes in miRNA expression. To address this issue, we employed an in vitro study of miRNA expression in cortical neurons derived from embryonic day 16.5 (E16.5) mouse brains. We obtained the expression profiles of 380 miRNA in neurons cultured at 12h (hour), 24h, 36h, 48, 96h, and 192h with or without Li treatment. Overall, 70% (266) of the miRNAs examined were expressed in the cultured neurons. Eighty-five of the 266 miRNAs showed two-fold or greater changes in treated compared to untreated neurons at any of the time points. We identified 55 miRNAs that show an early (within 48h) response to lithium treatment, and 30 with a more delayed (4 day or above) response. There were 14 miRNAs that significantly and consistently changed expression patterns over time in untreated neurons (false discovery rate, FDR < 0.15), and 48 miRNAs (FDR < 0.15) in neurons treated with Li. Six miRNAs were differentially expressed between treated and un-treated samples over the entire time course (FDR < 0.15). Our results indicate that Li significantly influences the expression of miRNAs in cultured neurons, and suggest a novel mechanism of Li's therapeutic action in BPD.

## 309

**To randomize or not to randomize - how "quasi-experimental" trials in naturalistic settings will change medical practice.** *F.W. Frueh, B.C. Agatep, B. Dechairo, C. Ewel, C.L. Sanders, E.J. Stanek, R.S. Epstein.* Medco Health Solutions, Bethesda, MD.

Genetic and pharmacogenetic research is becoming integrated into medical practice at variable pace depending on therapeutic, provider, and patient segmentation. Consumers and providers are becoming aware of the importance of weighing genetic factors when making medical decisions, and the recent market entry of several "consumer genetics" companies promises an unprecedented increase in the amount of genetic information to draw associations between genetic variations and phenotypes. However, this information is collected in uncontrolled, non-randomized, retrospective fashion, leading to significant and often justified criticism of the validity of these associations. This is particularly problematic in situations where this information may be used to make critical clinical decisions. While the prospective randomized controlled trial (RCT) remains the gold standard to demonstrate that a particular intervention exhibits a desired effect, alternative epidemiological and comparative effectiveness study designs may be more suitable to investigate the validity and clinical utility of genetic markers. Importantly, such alternative approaches can be executed in naturalistic, real-world settings, thereby providing more useful information about how a particular test affects clinical outcomes. We have developed a system to conduct "quasi-experimental" trials, i.e. studies that exhibit all the characteristics of a typical RCT, except that we do not randomize the participants. This approach greatly facilitates the conduct of pharmacogenetic research, particularly in situations where randomization may prove difficult because of ethical or social concerns. As a large pharmacy benefit provider, we have access to a nationally representative member population of approximately 65 million in the United States. We have utilized this resource to design and execute a series of non-randomized, controlled prospective studies to assess whether pharmacogenetic testing for warfarin (CYP2C9 and VKORC1) and tamoxifen (CYP2D6) impacts clinical outcomes and clinical decision making, respectively. Early data indeed suggests that the design of these studies is reflective of other prospective and retrospective studies and, therefore, allows us to extrapolate our findings to a broader, real-world population. Early data from this research and a description of the operational approach to conduct quasi-experimental trials in a naturalistic setting will be presented.

**310**

**Nine years of systematic DNA collection in clinical pharmacology trials - what have we learned?** *M.G. Ehm[1], L.K. Hosking[2], A.J. Yeo[2], L.R. Cardon[3], J.F. Hoke[4], J.H. Kim[4], J.W. Polli[5], V.D. Schmith[4], Z. Xue[1], C.F. Spraggs[2].* 1) Genetics, GlaxoSmithKline, Res Triangle Park, NC; 2) Genetics, GlaxoSmithKline, Harlow, UK; 3) Genetics, GlaxoSmithKline, Collegeville, PA; 4) CPMS, GlaxoSmithKline, Res Triangle Park, NC; 5) DMPK, GlaxoSmithKline, Res Triangle Park, NC.

Drug disposition variability is a concern in drug development, and may be partly under genetic control, by functional variants in genes involved in metabolism. Understanding of the genetic basis of inter-individual variability may assist in dose selection, drug safety prediction, drug-drug interaction characterization and in explaining pharmacokinetic (PK) study outliers. Clopidogrel and tamoxifen have illustrated important clinical consequences due to PK variability related to pharmacogenetics (PGx). In light of growing evidence in the 1990s of the impact PGx could have on drug disposition, all clinical pharmacology studies in GSK had systematic DNA collection since 1998. We evaluated, retrospectively, the impact that this sample collection had on drug development at GSK. A total of 29 early-phase PGx experiments were included in the analysis, using 1990 samples collected from 73 clinical pharmacology studies. These experiments investigated if functional variation in drug metabolism (ADME) genes was associated with safety endpoints (9 experiments, including analysis of UGT1A1 and drug-induced hyperbilirubinemia), PK variability or outliers (18 experiments focusing on CYP2D6, 2C9, and 2C19, and occasionally 3A5, 1A2, 2B6), or pharmacodynamic endpoints (2). Twenty of these experiments provided a definitive result: a positive or negative association. Ten experiments showed that genetic variation could explain variability in liver enzymes (3), PK parameters (4), and other enzymes (1) and clarified metabolic routes (2). Three of these experiments influenced subsequent study designs. The other 10 studies showed that ADME candidate genes were not associated with PK variability (8) or liver enzyme variability (2) suggesting either other genes or non-genetic mechanisms were involved. Our analysis shows that systematic DNA collection in clinical pharmacology studies had limited effects on decision making in early clinical development, a time of great product attrition. However, important information regarding pharmacokinetic variability was gained in 20 of 29 experiments. Thus, systematic DNA collection in late stage clinical pharmacology studies has the potential to provide important information about genetic factors affecting pharmacokinetic variability, relevant to drug labeling, and drug interactions. This retrospective review of 9 years of systematic DNA collection illustrates the strength and limitations of PGx studies in early phase clinical trials.

## 311

**Common genetic variation in UGT8 is associated with musical ability.** *S. Lee[1,4], H. Park[1], J.I. Kim[1,2,3], Y.S. Ju[1,2], H.J. Kim[1,4], S.I. Cho[5], J. Sung[6], J.S. Seo[1,2,5,6].* 1) ILCHUN Genomic Medicine Institute, Medical Research Center, Seoul National University, Seoul, Korea; 2) Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, Seoul, Korea; 3) Psoma Therapeutics Inc., Seoul, Korea; 4) Department of Biomedical Sciences, Seoul National University Graduate School, Seoul, Korea; 5) Seoul National University School of Public Health, Seoul, Korea; 6) Macrogen Inc., Seoul, Korea.

Musical abilities like recognizing music and singing performance are specific traits of human. Specific regions of the brain were found to be activated by musical stimuli, but genes and molecules associated with musical ability are not identified. To identify genetic determinants that influence musical ability, we performed both genome-wide linkage analysis using 1,099 microsatellite markers and genome-wide association (GWA) study with Infinium 610k genotyping array in large extended families. A total of 1026 individuals (54% females) from 73 families were enrolled and pitch production accuracy (PPA) test was applied to determine musical ability. We found significant evidence of linkage (LOD score = 3.25) for musical ability on chromosome 4q25. In chromosome 4, we further investigated sequence variants associated with musical ability that reached a genome-wide significance threshold (P < 1.5 x 10-6). SNP rs4148255 (P = 1.6 x 10-10), located within the gene encoding UDP-galactose ceramide galactosyltransferase (UGT8) near to significant linkage locus, showed strong association with musical ability. UGT8 is highly expressed and known to act in the brain organization. Here, we represent first genetic variants for musical ability, evolutionary distinct faculty from groups of individuals.

## 312

**Chromosome 13q13-q14 locus overlaps mood and psychotic disorders: the relevance for redefining phenotype.** *M. Maziade[1,2], Y.C. Chagnon[1,2], M.A. Roy[1,2], A. Bureau[2,3], A. Fournier[2], C. Mérette[1,2].* 1) Dept Psychiatry, Laval University, Québec, Canada; 2) Centre de recherche Université Laval Robert-Giffard, Québec, Canada; 3) Dept of Social and Preventive Medicine, Laval University, Québec, Canada.

The nosology of major psychoses is challenged by the findings that schizophrenia (SZ) and bipolar disorder (BP) share several neurobiological, neuropsychological and clinical phenotypic characteristics. Moreover, several vulnerability loci or genes may be common to the two DSM disorders. We previously reported, in a sample of 21 kindreds (sample 1), a genome-wide suggestive linkage in 13q13-q14 with a common locus (CL) phenotype that crossed the diagnostic boundaries by combining SZ, BP and schizoaffective disorders. Our objectives were to test phenotype specificity in a separate sample (sample 2) of 27 kindreds from Eastern Quebec and to also analyze the combined sample of 48 kindreds (1274 family members). We performed nonparametric and parametric analyses and tested as phenotypes: SZ alone, BP alone, and a CL phenotype. We replicated in sample 2 our initial finding with CL with a maximum $NPL_{pair}$ score of 3.36 at D13S1272 (44 Mb), only 2.1 Mb telomeric to our previous maximum result. In the combined sample, the peak with CL was at marker D13S1297 (42.1 Mb) with a $NPL_{pair}$ score reaching 5.21, exceeding that obtained in each sample and indicating consistency across the two samples. Our data suggest a susceptibility locus in 13q13-q14 that is shared by schizophrenia and mood disorder. That locus would be additional to another well documented and more distal 13q locus where the G72/G30 gene is mapped.

## 313

**Inherent Confounding in Association Mapping Studies.** *A. Platt[1,2], M. Nordborg[1,2].* 1) Computational and Molecular Biology, University of Southern California, Los Angeles, CA; 2) Gregor Mendel Institute, Vienna, Austria.

Almost every phenotypic trait important enough to be studied is the product of multiple factors, yet the vast majority of association studies employed to find the genetic underpinnings of those traits are built on what are essentially single-factor statistical models. We introduce a general three-locus model that illustrates the pitfalls of these methods.

Considering binary variables representing a neutral locus, a causative locus, and a latent variable, we show that when a neutral allele is in disequilibrium with one contributing factor of the phenotype it will have a non-zero correlation with the trait but never a stronger correlation than the actual contributing factor does. However, when a neutral allele is in disequilibrium with two or more contributing factors, it can easily have a higher correlation with the trait than either of the contributing factors do, even without any form of epistatic interactions. Under these circumstances a study will produce false positive results that look exactly like true positives *and no amount of additional data will resolve the issue*. Increased sampling will only further confirm the false results.

Disequilibrium among markers and factors can come about in many ways, but is most commonly the result of genetic linkage, natural selection, or population structure. The presence of linkage is an inherent fact of human genetics. Over short genetic distances alleles across many loci will be in disequilibrium, and if more than one causative locus is within this region false positive results are likely to occur nearby. Natural selection tends to create disequilibrium among all factors contributing to the trait under selection. This generates pockets of disequilibrium dispersed genome-wide and allows for factors far apart from each other to contribute to false positives within linkage distance of any or all of the causative factors. Population structure creates disequilibrium across entire genomes. In structured populations false positives can occur at great distances from any of the true causative factors.

Great care need be taken in interpreting association studies. When multiple factors are involved in a trait fine-mapping may be impossible. Even when clean, clear results present themselves they can be in the wrong places.

## 314

**Identification of a second major locus for neurodegeneration with brain iron accumulation.** *M.B. Hartig[1,2], A. Iuso[2], M. Hempel[1,2], K. Heim[2], E. Jurkiewicz[3], T. Kmiec[3], T. Meitinger[1,2], T. Strom[1,2], H. Prokisch[1,2].* 1) Institute of Human Genetics, Technische Universität München, Munich, Germany; 2) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany; 3) Memorial Children's Health Institute, Warsaw, Poland.

Neurodegeneration with brain iron accumulation (NBIA) is a rare disorder defined by iron deposits in basal ganglia. We have clinically and genetically characterized a cohort of 54 NBIA patients from Poland. Expression profiles in whole blood allowed the differentiation of two groups of patients. A clinically and transcriptionally homogeneous group of 27 patients with generalized dystonia, pyramidal signs, cognitive deficits and retinal degeneration carries mutations in the mitochondrial pantothenate kinase gene PANK2. A founder mutation in PANK2 (~37 % frequency) prompted us to perform homozygosity mapping in the clinically and transcriptionally distinct group of patients affected by spastic paraparesis, dystonia, psychiatric symptoms, neuropathy and optic atrophy. The linkage study in 2 families led to the identification of a new NBIA gene harbouring 4 different mutations in 21 patients. The gene codes for a protein with a transmembrane domain and so far unknown function. Gene identification in index patients of the third group is ongoing by whole genome exon sequencing.

## 315

**Identification of Multiple Functional PLTP locus SNPs and their effects on linkage evidence for other chromosomes.** *E.A. Rosenthal[1], R. Rajagopalan[1], G. Wolfbauer[2], J.J. Albers[2], E.M. Wijsman[1,3], G.P. Jarvik[1,4].* 1) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 2) Department of Medicine, Division of Metabolism, Endocrinology, and Nutrition, University of Washington, Seattle, WA; 3) Department of Biostatistics, University of Washington, Seattle, WA; 4) Department of Genome Sciences, University of Washington, Seattle, WA.

Atherosclerosis is a common complex trait that is associated with phospholipid transfer protein (PLTP) activity in mice. Identification of genetic underpinnings of PLTP activity in humans may help to detect new genetic pathways for atherosclerosis. To that end, we performed Bayesian joint linkage and segregation analysis on PLTP activity on four large families(n=210 with phenotype and genotype, n=462 total) ascertained for familial combined hyperlipidemia. This data set contains genotypes for 486 multiallelic markers and the Illumina 48K HumanCVD SNP BeadChip. We use the Bayes factor (BF), the ratio of posterior to prior odds of linkage, over 2 cM intervals, to evaluate evidence for linkage, with a BF>10 indicative of possible linkage. We found evidence for linkage on chr 2 (BF=34.7) and 11 (BF=33.7), but not for the structural locus on chr 20 (BF=3.5). However, 6 PLTP promoter SNPs (rs553359, rs6065904, rs6073952, rs394643, rs4810479, and rs7679) are associated with PLTP activity in both the families and in a separate unrelated case control cohort (all p<=0.01). Incorporating rs6065904 as a covariate in the segregation model results in BF=57.5 and 6.5 for chr 2 and 11, respectively. The increase in the BF on chr 2 and the decrease on chr 11 suggest that using information from SNPs associated with the phenotype of interest may help differentiate between true and false positive linkage signals in other regions of the genome.

## 316

**Visualizing chromosome mosaicism and detecting ethnic outliers by the method of "rare" heterozygotes and homozygotes (RHH).** *R. McGinnis, P. Deloukas, W. McLaren, M. Inouye.* Wellcome Trust Sanger Institute, Cambridge, UK.

Ethnically matched disease cases and controls are being genotyped at hundreds of thousands of SNPs across the human genome to conduct genome-wide association (GWA) scans for loci that cause complex human disease. Here we describe a novel approach for evaluating genotypes of a GWA scan to both visualize the mosaicism of ethnically admixed chromosomes and to identify outlier subjects whose ethnic ancestry is different or admixed compared to most other subjects in the genotyped sample set. Each "ethnic outlier" is detected by counting a genomic excess of "rare" heterozygotes and/or homozygotes whose frequencies are low (<1%) within genotypes of the sample set being evaluated. This method also enables simple and striking visualization of non-Caucasian chromosomal DNA segments interspersed within the chromosomes of ethnically admixed individuals, thereby delineating the mosaic structure of human chromosomes derived from admixture of ethnically distinct groups. We show that our visualization of admixed chromosome mosaicism gives results similar to other visualization methods such as those based on hidden Markov or related models (e.g. SABER) but with much less computational time and burden. We also show how to enhance detection of ethnic outliers by focusing our own and other outlier detection methods on regions of admixture rather than diluting the evidence for outlier ancestry by evaluating GWA genotypes from the entire genome considered in aggregate. We have validated our method in the Wellcome Trust Case Control Consortium (WTCCC) study of 17,000 subjects as well as in HapMap subjects and simulated outliers of known ethnicity and admixture. The method's ability to precisely delineate chromosomal segments of non-Caucasian ethnicity has enabled us to clearly demonstrate previously unreported non-Caucasian admixture in two HapMap Caucasian parents (NA12872, NA11993) and in a number of WTCCC subjects. We will illustrate both visualization of mosaicism and detection of ethnic outlier status with examples from WTCCC, HapMap and simulated admixture. Its simple visual discrimination of discrete chromosomal segments of different ethnicity implies that this method of rare heterozygotes and homozygotes (RHH) is likely to have diverse and important applications in humans and in other species.

## 317

**Using High density 1meg SNP genotyping arrays to determine the lower size for anomalous contiguous homozygosity as a methodology for generating candidate genes in the NIH Undiagnosed Disease Program.** *T. Markello[1], D. Adams[1], W.A. Gahl[1,2,3].* 1) Med Gen Branch, NIH/NHGRI, Bethesda, MD; 2) Head, Intramural Program Office of Rare Diseases, NIH, Bethesda MD; 3) Clinical Director, NIH/NHGRI, Bethesda MD.

Recent recognition of regions of contiguous homozygosity seen by SNP array genotyping has given rise to a potential way to uncover identity by descent and segmental uniparental disomy. This allows determination of submicroscopic variations in the individual human genome undetectable by Giemsa staining, FISH studies or comparative genomic hybridization. To define the lower size limit when using high density SNP arrays for single proband homozygosity analysis, we measured the correlation lengths of homozygosity runs of consecutive snps within every one megabase interval over all chromosomes in 54 samples . We used the Illumina 1MDUO array as the single platform for genotyping. Critical to this determination is the truncating effect on the correlation lengths caused by spurious heterozygosity from low signal quality SNPs. Using group scores and individual scores as filter parameters, we found that removing less than 10% of the total snps removed 95% of the spurious heterozygous samples and made quantitative homozygosity length correlation analysis feasible for 81% of the genome. We analyzed 27 controls and 27probands in the Undiagnosed Disease Program(UDP )recently initiated at the NIH. We found each chromosome interval has a different normal homozygosity length depending on snp density and recombination frequency that roughly follows the HapMap frequencies. The occurrence of contiguous snps declines rapidly for most intervals at 50 to 300 contiguous snps, and at 200kb to 750kb for the total homozygous length. Regions that exceed these sizes are seen in 67% (18/27) of individuals in the Undiagnosed Disease Cohort versus only 18% (5/27) for parents and unaffected sibling controls. The sum of these regions, normalized to the total genomic length visible in the snp array, gives good correlation ($r^2$=0.91) in 26 known consanguinity controls from 13 different pedigrees. Defining the minimum limits for declaring excess homozygosity, on a region by region basis, will facilitate the ability to call very small regions anomalous, and improve the ability to generate candidate loci for genetic recessive disorders. Combined with available haplotyping tools, this approach allows computationally simplified linkage in favorable cases. We describe one case of linkage found using this method and a second case of cryptic homozygosity not known from the medical history.

**318**

**Identification of two mutations for Mendelian Susceptibility to Mycobacterial Diseases (MSMD) using a novel homozygosity mapping approach accounting for heterogeneity.** *A.V. Grant[1], S. Boisson-Dupuis[2], E. Herquelot[1], D.K. Nolan[1], J. Feinberg[1], J. Bustamante[1], A. Boland[3], O. Sanal[4], Y. Camcioglu[5], A. Palanduz[6], S. Kilic[7], J.L. Casanova[1,2], L. Abel[1,2].* 1) Laboratory of Human Genetics of Infectious Diseases, Necker Faculty, Institut National de la Santé et de la Recherche Médicale, Unité 550, Paris, France; 2) Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, New York, USA; 3) Commissariat à l'Energie Atomique, Institut de Génomique, Centre National de Génotypage, Evry, France; 4) Immunology Division, Hacettepe University Children's Hospital, Ankara, Turkey; 5) Department of Pediatrics, Infectious Diseases, Clinical Immunology and Allergy Division, Cerrahpasa Medical School, Istanbul University, Istanbul, Turkey; 6) Department of Pediatrics, Sisli Etfal Hospital, Istanbul, Turkey; 7) Division of Internal Medicine, Gulhane Military School of Medicine, Ankara, Turkey.

Mendelian susceptibility to mycobacterial diseases (MSMD: MIM 209950) is a rare syndrome that develops in predisposed individuals in response to innocuous mycobacteria including BCG vaccine. A total of 7 genes have been implicated in MSMD and functional tests allow for screening of the corresponding defects in new patients, but over half of patients have no identified defects. We hypothesized that new mutations in several genes can explain these cases, and assembled a sample of 17 MSMD patients belonging to consanguineous families. Genome-wide (GW) homozygosity mapping (HM) is a powerful method to locate rare recessive Mendelian mutations. However, statistical power can decrease dramatically in the presence of genetic heterogeneity which is likely to increase with sample size. We propose a novel approach to test for linkage accounting for genetic heterogeneity in the context of GWHM. After having computed GW multipoint LOD scores, the principle is, at each map position i,: a) to rank individual family multipoint LOD scores, b) to compute the sum of LOD scores for the first to the kith family, where $k_i$ is the last family at the ith position with a positive LOD score. In order to evaluate the significance of these sums in the context of GW results from the entire sample, empirical p-values are calculated using permutations. The method was applied to a GW scan conducted using MERLIN in the 17 MSMD families genotyped with a 250K SNP array. No positive LOD scores were obtained when considering the entire sample. Using our method, we identified 3 regions providing evidence for linkage at $p < 10^{-3}$. The most significant result was obtained in the chr19 region (P=0.00037) with 6 linked families. The two most compelling candidate genes of that region, *IL12RB1* and *TYK2*, were sequenced in the 6 corresponding patients. One patient was found to be homozygous for a D101Y mutation in *IL12RB1* and another patient for a 9 bp deletion in exon16 of *TYK2*. While *IL12RB1* is a known MSMD gene, the D101Y patient could not be investigated by functional tests due to absence of fresh cells. Conversely, *TYK2* is a novel MSMD gene, and only a single patient with TYK2 deficiency and complex clinical features has been reported so far. Identification of two novel mutations using our novel permutation-based GWHM approach validates our overall strategy to account for genetic heterogeneity that could be applied to cohorts of patients with other suspected recessive Mendelian disorders.

## 319

**Genome-wide significant predictors of metabolites in the one-carbon metabolism pathway.** *A. Hazra[1,2], P. Kraft[1], R. Lazarus[2,3], C. Chen[1], S. Chanock[4], P. Jacques[5], J. Selhub[5], D. Hunter[1,2].* 1) Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA; 2) Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 181 Longwood Avenue, Boston, MA 02115, USA; 3) Department of Ambulatory Care and Prevention, Harvard Medical School, Boston, MA 02115, USA; 4) Division of Cancer Epidemiology and Genetics, National Cancer Institute(NCI), US National Institutes of Health (NIH), Department of Health and Human Services (DHHS), Bethesda, Maryland 20892, USA; 5) Vitamin Metabolism and Aging Laboratory and Nutritional Epidemiology Program, Jean Mayer US Department of Agriculture Human Nutrition Research Center on Aging, Tufts University, Boston, MA 02111, USA.

Diminished plasma B-vitamin levels and increased homocysteine have been associated with cancer, cardiovascular disease, and neurodegenerative disorders. Common variants in $\alpha$*,1,2-fucosyltransferase (FUT2)* on chromosome 19q13 were associated with plasma vitamin $B_{12}$ levels among women in a genome-wide association study (GWAS) in the Nurses' Health Study (NHS) NCI-Cancer Genetic Markers of Susceptibility (CGEMS) project. To identify additional loci associated with plasma vitamin $B_{12}$, homocysteine, folate and vitamin $B_6$ (active form pyridoxal 5'-phosphate, PLP), we conducted a meta-analysis of three GWA scans (total n=4,763, consisting of 1,658 women in NHS-CGEMS, 1,647 women in Framingham-SNP-Health Association Resource (SHARe), 1,458 men in SHARe). On chromosome 19q13, we confirm the association for plasma vitamin $B_{12}$ with rs602662 and rs492602 (p-value=$1.83x10^{-15}$ and $1.30x10^{-14}$, respectively) in strong linkage disequilibrium (LD) with rs601338 (P=$6.92x10^{-15}$), the *FUT2 W143X* nonsense mutation. We identified additional genome-wide significant loci for plasma vitamin $B_{12}$ on chromosomes 6p12 (P=$4.05x10^{-8}$), 10p12 (p-value=$2.87x10^{-9}$), and 11q11 (p-value=$2.25x10^{-10}$) in genes with biological relevance. We confirm the association of the well-studied functional candidate SNP *5,10-methylene tetrahydrofolate reductase (MTHFR) Ala222Val* (dbSNP ID: rs1801133; p-value=$1.27x10^{-8}$), on chromosome 1p36 with plasma homocysteine and identify another locus (P=$5.81x10^{-10}$, composite LD with *MTHFR 222* in our data is $r^2$=0.53) 102kb away from *MTHFR*, and an additional genome-wide significant locus on chromosome 9q22 (p-value=$2.06x10^{-8}$) associated with plasma homocysteine. We also identified genome-wide associations with variants on chromosome 1p36 with plasma PLP (p-value=$1.40x10^{-15}$). Genome-wide significant loci were not observed for plasma folate. These data reveal new biological candidates and confirm prior candidate genes for plasma homocysteine, plasma vitamin $B_{12}$, and plasma PLP.

## 320

**Genome-wide association study identifies loci involved in regulation of human plasma N-glycome.** *I. Rudan[1,2,3], C. Hayward[4], A. Essafi[4], J.E. Huffman[4], A. Knezevic[5], O. Polasek[3,6], O. Gornik[5,7], V. Vitart[4], L. Zgaga[6], M. Pucic[5], I. Redzic[5], F. Borovecki[8], N.D. Hastie[4], J.F. Wilson[1], A.F. Wright[4], H. Campbell[1,4], P.M. Rudd[9], G. Lauc[5,7].* 1) Public Health Sciences, The University of Edinburgh, Edinburgh, Scotland, United Kingdom; 2) Croatian Centre for Global Health, University of Split Medical School, Split, Croatia; 3) Gen-Info Ltd, Zagreb, Croatia; 4) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Western General Hospital, Crewe Road, Edinburgh, UK; 5) Department of Biochemistry and Molecular Biology, University of Zagreb Faculty of Pharmacy and Biochemistry, Zagreb, Croatia; 6) Andrija tampar School of Public Health, University of Zagreb Medical School, Zagreb, Croatia; 7) Glycobiology Laboratory, Genos Ltd, Planinska 1, Zagreb, Croatia; 8) Centre for Functional Genomics, University of Zagreb Medical School, Zagreb, Croatia; 9) NIBRT, Dublin-Oxford Glycobiology Lab., Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland.

Statement of purpose: Glycosylation is the most complex and abundant post-translational modification of proteins. Human glycome is several orders of magnitude more complex than human proteome because glycosylation is not template driven. Genetic defects that affect protein glycosylation are a known cause of more than 30 human diseases. Population variability in terminal glycans contributes to protein heterogeneity that can be advantageous for evading pathogens and adapting to changing environments. Understanding of the genetic regulation of glycosylation is still very limited due to experimental limitations in quantifying glycans in complex biological samples, but new high-throughput methods have now been developed. This study reports on the first attempt to combine high-throughput glycomics with high-throughput genomics. Methods used: High performance liquid chromatography (HPLC) analysis was used to quantify levels of 16 groups of N-linked glycans on the human plasma proteins in about 1,000 island isolates from Vis, Croatia and in further 1,800 islanders from Korcula, Croatia and Orkney, UK as two independent replication cohorts. As proof of principle, we selected bi-antennary N-linked glycan A2 (GlcNAc2Man3GlcNAc2), the only glycan that did not co-elute with others in a single HILIC-HPLC run. Subsequent genome-wide association study (GWAS) using 317,503 single nucleotide polymorphisms (SNP) was conducted to identify genetic loci influencing variation in glycan A2. Summary of results: The study identified fucosyltransferase 8 (FUT8) on chromosome 14 as the key regulator of the levels of glycan A2, reaching genome-wide significance in discovery cohort and replication cohorts. The meta-analysis showed that the SNP most strongly associated with glycan A2 levels in human plasma was rs11621121 (p=7.59x10-18). This finding is biologically highly plausible, because glycan A2 is a known substrate for FUT8, which is alpha (1,6) fucosyltransferase. The effect was much stronger in females than in males, and the breakdown of females into premenopausal and postmenopausal groups and further association analysis showed that FUT8 explains up to 20% of the trait variance (adjusted for age) in premenopausal women and up to 5% in postmenopausal women. Conclusion: We demonstrated that human plasma N-glycans are amenable to genome-wide association studies and that their genetic regulation shows similar characteristics to other biological quantitative traits.

## 321

**Common variants in the TSLP/WDR36 region are associated with Eosinophilic Esophagitis.** *K. Annaiah[1], J. Spergel[1], J. Sherrill[2], L. Martin[2], C. Blanchard[2], R. Chiavacci[1], C. Kim[1], J. Flory[1], W. Glaberson[1], A. Thomas[1], P. Sleiman[1], M. Rothenberg[2], H. Hakonarson[1].* 1) Children's Hosp Philadelphia, Philadelphia, PA; 2) Cincinnati Children's Hospital Medical Center, OH.

Eosinophilic esophagitis (EE) is a polygenic disorder characterized by the accumulation of eosinophils in the esophagus. We performed a genome-wide association study on clinically confirmed EE patients to identify common variants associated with disease risk. Two hundred and seventy two EE samples from Cincinnati Children's Hospital (CCHMC) and 180 EE samples from Children's Hospital of Philadelphia (CHOP) were genotyped on the illumina® 550K BeadChip. Control samples were obtained from CHOP, all were genotyped on the illumina® 550K BeadChip and were negative for EE. All patients and controls were of Caucasian descent. Following standard quality control filtering of the genotype data we carried out Chi-squared analysis at each SNP using the CCHMC samples (181 cases, 1974 controls) as a discovery cohort. We detected genome-wide association with variants on chr5q22 that mapped to a single LD block encompassing the TSLP and WDR36 genes. The most significantly associated SNP rs3806932 which maps upstream of the TSLP gene remained genome-wide significant after Bonferroni correction for multiple testing (uncorrected P-value 8.81x10-8 OR 0.54). Eleven other SNPs in LD with rs3806932 were also significantly associated with EE and mapped to the same LD block on 5q22. We subsequently replicated the association in the independent CHOP cohort (170 cases, 1130 controls) with rs3806932 (P-value 9x10-3; OR 0.73; combined P-value for rs3806932 across the CCHMC and CHOP cohorts 4.6x10-9). In summary, we have identified the first genetic association with EE predisposition and implicate TSLP and/or WDR36 as potential genes involved in disease susceptibility.

## †322

**The largest genome-wide association analysis in US Hispanics identifies novel asthma susceptibility variant and gene-gene interactions.** *C.R. Gignoux[1], D. Hu[2], C. Eng[2], M. Via[2,3], S. Huntsman[2], J. Rodriguez-Santana[4], W. Rodriguez-Cintron[5], R. Chapela[6], P. Avila[7], M. Aldrich[2,3], S. Choudhry[2,3], E. Ziv[2,3,8], E.G. Burchard[2,3,9].* 1) Program in Pharmaceutical Sciences and Pharmacogenetics, University of California, San Francisco, CA; 2) Department of Medicine, University of California, San Francisco, CA; 3) Institute for Human Genetics, University of California, San Francisco, CA; 4) Pediatric Pulmonary Program of San Juan, San Juan, Puerto Rico; 5) Veterans Affairs Medical Center, University of Puerto Rico School of Medicine, San Juan, Puerto Rico; 6) Instituto Nacional de Enfermedades Respiratorias, Mexico City, Mexico; 7) Division of Allergy-Immunology, Northwestern University Feinberg School of Medicine, Chicago, IL; 8) Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, CA; 9) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA.

Asthma is a common but complex respiratory ailment; current data indicate that interaction of genetic and environmental factors lead to its clinical expression. In the U.S., asthma prevalence, morbidity and mortality are highest in Puerto Ricans and lowest in Mexicans. This is paradoxical since both groups are considered “Hispanic or Latino”, are admixed and share varying proportions of African, Native American and European ancestry. Although there were many potential explanations for this observation, including social and environmental factors, the rich mixed ancestry of Latinos provides the intrinsic variability needed to untangle complex gene-environment interactions in disease susceptibility and severity. To better understand the etiology of asthma and its high variance across populations, we organized the largest multi-center family-based study of U.S. Latinos: over 700 parent-child trios and carefully matched controls for childhood asthma in both Puerto Ricans and Mexicans. Using the Affymetrix 6.0 platform we found a novel, genomewide significant association among Puerto Ricans in the Myosin Phosphatase gene PPP1R12B with a p value < $10^{-8}$ in an evolutionary conserved intronic region, with an OR of 3.7 (95% CI 2.4-5.9). Myosin Phosphatase is expressed in the lung smooth muscle. Other SNPs offer suggestive evidence (p value < $10^{-6}$) of association. Incorporation of admixture proportions with the top GWAS hits estimates case/control status better than either ancestry or genotype alone. We furthered our investigation by incorporating multiple methods of detecting higher-order interactions: SNP-SNP comparisons and machine learning. SNPs near a chemokine-like gene CMTM7 and a nitric oxide synthetase trafficking isoform (NOSTRIN) appear to have epistatic effects in cases at a high significance level (<$10^{-15}$). NOSTRIN has high haplotype homozygosity in Native Americans consistent with it being under positive selection. Further tests of higher-order machine learning approaches are discussed along with the need to measure both genotype and ancestral origin for each position in the genome. To investigate patterns of admixture that may contribute to disease we show how differential ancestry across Puerto Rico can confound association studies, where historical settlement patterns resulted in significant ancestry clines seen today.

## 323

**Twenty bone mineral density loci identified by large-scale meta-analysis of genome-wide association studies.** *F. Rivadeneira[1,2], U. Styrkarsdottir[3], K. Estrada[1], B.V. Halldorsson[3,4], Y.H. Hsu[5], J.B. Richards[6,7,8], M.C. Zillikens[1], F.K. Kavvoura[9], S.H. Ralston[10], U. Thorsteinsdottir[3,11], C.M. van Duijn[2], T. Spector[8], D.P. Kiel[5], K. Stefansson[3,11], J.P.A. Ioannidis[9,12], A.G. Uitterlinden[1,2], The GEFOS Consortium.* 1) Dept Internal Medicine, Erasmus MC, Rotterdam, 3015GE Netherlands; 2) Dept Epidemiology, Erasmus MC, Rotterdam, 3015GE, The Netherlands; 3) deCODE Genetics, 101 Reykjavík, Iceland; 4) Reykjavik University, 103 Reykjavik, Iceland; 5) Hebrew Senior Life, Harvard Medical School, Boston, MA, 02131 USA; 6) Dept Medicine, McGill University, Montréal, H3G 1Y6 Canada; 7) Dept Human Genetics, McGill University, Montréal, H3G 1Y6 Canada; 8) Dept Twin Research and Genetic Epidemiology, Kings College London, London, SE1 7EH, United Kingdom; 9) Dept Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina 45110, Greece; 10) Rheumatic Diseases Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, United Kingdom; 11) Faculty of Medicine, University of Iceland, 101 Reykjavík, Iceland; 12) Center for Genetic Epidemiology and Modeling, Tufts Medical Center, Tufts University School of Medicine, Boston, MA 02111, USA.

Bone mineral density (BMD) is a heritable complex trait used in the clinical diagnosis of osteoporosis and the assessment of fracture risk. We performed meta-analysis of five genome-wide association studies of femoral neck (FN) and lumbar spine (LS) BMD in 19,195 Caucasian subjects. We identified 20 loci that reached genome-wide significance (GWS P<5x10-8), of which 13 map to new regions including 1p31.3 (GPR177), 2p21 (SPTBN1), 3p22 (CTNNB1), 4q21.1 (MEPE), 5q14 (MEF2C), 7p14 (STARD3NL), 7q21.3 (FLJ42280), 11p11.2 (LRP4; ARHGAP1; F2), 11p14.1 (DCDC5), 11p15 (SOX6), 16q24 (FOXL1), 17q21 (HDAC5) and 17q12 (CRHR1). The meta-analysis also confirmed at the GWS level, seven known BMD loci on 1p36 (ZBTB40), 6q25 (ESR1), 8q24 (TNFRS11B), 11q13.4 (LRP5), 12q13 (SP7), 13q14 (TNSF11), and 18q21 (TNFRS11A). In one of the population-based cohorts the 15 SNPs associated with LS-BMD combined explained ~2.9% of the trait variance and the 10 SNPs associated with FN-BMD combined explained ~1.9% of the trait variance. A compound allelic score of SNPs associated with FN-BMD was associated consistently with the risk for both types of fracture. The risk of vertebral fracture increased 6% by each increase in the number of “low FN-BMD” risk alleles (OR=1.057, 95%CI[1.010,1.107]; P=0.02), while the risk of incident non-vertebral fracture increased 3% by each increase in “low FN-BMD” risk alleles (HR=1.033, 95%CI[1.004,1.063]; P=0.03). Adjustment for FN-BMD showed that at least 46% of the genetic effect on vertebral fracture can be explained by FN-BMD (ORadjusted= 1.031, 95%CI[0.983,1.080]; P=0.21), and 54% of the genetic effect on incident non-vertebral fracture is through FN-BMD (HRadjusted= 1.015, 95%CI [0.986,1.044]; P=0.32). The novel SNPs associated with BMD variation map to genes in signaling pathways of potential relevance to bone metabolism. Our findings highlight the complex genetic architecture underlying osteoporosis and identify novel genetic variants that are associated with BMD.

† Trainee Award Finalist

## 324

**Web-based, participant-driven studies yield novel genetic associations for common traits.** *N. Eriksson[1], J.M. Macpherson[1], J. Tung[1], L. Hon[1], B. Naughton[1], S. Saxonov[1], L. Avey[1], A. Wojcicki[1], I. Pe'er[2], J. Mountain[1,3].* 1) 23andMe, Inc., Mountain View, CA; 2) Department of Computer Science, Columbia University, New York, NY; 3) Department of Anthropology, Stanford University, Stanford, CA.

Despite the recent, rapid growth in genome-wide data, much of human variation remains entirely unexplained. A significant challenge in the pursuit of the genetic basis for variation in common human traits is the efficient, coordinated collection of genotype and phenotype data.

We report on initial results from a participant-driven study of 22 common traits based on a novel research framework that facilitates the parallel study of a wide assortment of traits within a single cohort. The approach takes advantage of the interactivity of the web both to gather data and to present genetic information to research participants, while taking care to correct for the population structure inherent to this study design.

We present novel associations for hair curl, "asparagus anosmia" (the inability to smell the methanethiol produced after eating asparagus), and photic sneeze reflex. For hair curl, we identify two independent SNPs: rs17646946 (p-value less than $10^{-28}$, near *TCHH*) and rs7349332 (p-value less than $10^{-8.4}$, in *WNT10A*). For asparagus anosmia, we identify one SNP in a region of olfactory receptors, rs4481887, with a p-value less than $10^{-16}$. For photic sneeze reflex, we identify one SNP, rs1040173, with a p-value less than $10^{-9.7}$. In order to validate the web-based, self-reporting design, we have in addition replicated associations in the genes *OCA2, HERC2, SLC45A2, SLC24A4, IRF4, MC1R, TYR, TYRP1* and *ASIP* for hair color, eye color, and freckling.

The other traits analyzed in this study include laterality preferences (handedness, footedness, ocular dominance, and hand-clasp), simple physical characteristics (whether participants have had cavities, have worn braces, have had wisdom teeth removed, have astigmatism, wear glasses, have attached earlobes, and suffer from motion sickness while riding in a car), and personality traits and preferences (optimism, a preference for sweet versus salty food, and preference for night-time versus morning-time activity).

## 325

**A genome wide association study identifies novel loci influencing telomere length in the Amish.** *O.T. Njajou[1], C.M. Damcott[2], E.H. Blackburn[1], N. Bendjilali[1], T.I. Pollin[2], J. Lin[1], D.S. Evans[1], A.R. Shuldiner[2], B.D. Mitchell[2], W-C. Hsueh[1].* 1) University of California San Francisco, San Francisco, CA; 2) University of Maryland, Baltimore, MD.

Telomere length (TL) is associated with aging, is moderately heritable, and may be influenced through sex chromosome or parent-specific effects. To better understand its genetic basis, we conducted a genome wide association study of TL using the Affymetrix 500K SNP array in a homogeneous population. Our sample included 803 healthy Lancaster County, PA Old Order Amish individuals participating in the Heredity and Phenotype Intervention (HAPI) Heart Study who were connected within a 14-generation pedigree (46% women, mean age: 44 ± 14 yrs, ranging from 20 - 81 yrs). TL in leukocytes was measured by quantitative PCR (value range: 1,679 - 13,000 bp, mean: 5,497 ± 2,074 bp). An additive genetic model was used after adjusting for the effects of age, sex and plate assay. The variance component analysis was used conditioning on the pedigree structure. The significance was evaluated using the likelihood ratio test. The heritability estimate of TL in this sample was 0.56 ± 0.06 ($p < 0.001$). The table below shows the only SNPs with genome wide significance ($p < 10E-7$).

| RS NUMBER | CHR | MAF | P VALUE | LOCATION/ GENE |
|---|---|---|---|---|
| rs2547254 | 19q13 | 0.05 | 3.23E-11 | Intergenic |
| rs816606 | 10p15 | 0.05 | 8.10E-08 | Intron/ DIPC2 |

In silico fine mapping analysis using imputed SNPs did not identify any stronger signals suggesting that these two were the most significantly associated SNPs within their respective regions. We have identified two novel quantitative traits loci influencing TL in the Amish. These variants have not been previously reported to be associated with TL. Our observations might provide new insights into the mechanisms of TL homeostasis in humans and pave the way to functional studies of these novel genes.

## 326

**Identifying the Genetic Underpinnings Associated With Healthspan.** *S.S. Murray[1], M. Shaw[1], S. Topol[1], P.C. Ng[2], N.J. Schork[1], K. Wang[3], H. Hakonarson[3,4], S.R. Cummings[5], E.S. Orwoll[6], E.J. Topol[1].* 1) Scripps Genomic Medicine, Scripps Translational Science Institute, La Jolla, CA 92037; 2) Genomic Medicine, J. Craig Venter Institute, San Diego, CA 92121; 3) Center for Applied Genomics, Children's Hospital of Philadelphia, PA, 19104; 4) Department of Pediatrics, Children's Hospital of Philadelphia, PA, 19104; 5) San Francisco Coordinating Center, California Pacific Medical Center Research Institute, San Francisco, CA 94107; 6) Oregon Health Sciences University, Portland, OR 97239.

Current genome-wide association studies (GWAS) have largely focused on understanding the causes of specific diseases. It is now possible, and indeed essential, to study the genetic underpinnings of human health. As with any complex trait a critical step in its investigation is defining the exact phenotype to be investigated and the study designs to be used to identify genetic associations. We define the "healthy aging" phenotype as individuals who have reached 80 years of age or greater, having suffered no chronic illnesses, are not taking any significant chronic medications, and maintaining high cognitive function with all activities of daily living preserved.

We have collected DNA samples from over 650 such individuals which we have termed "wellderly" who are over age 80 (range 80-105), without any significant chronic illness, and thus fulfill our definition of healthy aging. We have completed a pilot GWAS by genotyping 200 wellderly samples in the Illumina Human1M and comparing to over 1200 young, population-based controls (mean age = 7 years, range = 3 weeks-18 years) genotyped on the Illumina Human610. We analyzed 498,520 overlapping SNPs that had a minor allele frequency ≥ 0.05. The most significant association identified in this pilot study is at rs10746192 ($1.85 \times 10^{-6}$), located in the intron of PPFIA2, a protein tyrosine phosphatase. Other associated regions of interest include SNPs near the genes MSRA which functions as a repair enzyme for proteins that have been inactivated by oxidation, and PARK2, where defects in PARK2 are associated with Parkinson's Disease. We are currently completing the GWAS by expanding genotyping to over 400 wellderly samples and genotyping a second control sample in the Illumina Human1M. This second control sample, termed the "illderly," is comprised of deceased members of two large population-based cohorts (MrOS and SOF, respectively) that have been tracked for more than two decades. The causes of death are those most common in the US, such as cardiovascular disease, stroke, and cancer. We will present results from both GWA studies comparing both assayed and imputed genotypes in the >400 wellderly cohort to the >400 illderly cohort and the >1200 healthy young cohort, respectively. The findings from this research initiative are expected to identify key gene modifiers that afford protection across the spectrum of late-onset diseases.

## 327

**Modeling Cornelia de Lange syndrome and investigating Nipbl function using conditional Nipbl knockout mice.** *T. Strachan, T.G. Smith, S.H. Laval, F. Chen, M. Rock, S. Jones, H. Peters.* Institute of Human Genetics, Newcastle University, Newcastle Tyne, United Kingdom.

Cornelia de Lange syndrome (CdLS) is a multiple malformation disorder characterized by growth retardation, cognitive impairment, limb abnormalities and distinctive craniofacial abnormalities. We formerly reported identification and characterization of a previously unstudied gene, NIPBL, as the major CdLS locus (Tonkin, E et al. Nature Genet. 36, 636-641; 2004). We also showed that the NIPBL protein works in a complex with the previously unstudied KIAA0892 protein to load cohesins onto chromatin (Seitan, V et al, PLoS Biol. 4, e242; 2006). Approximately half of CdLS individuals have causative NIPBL mutations. Possession of a single mutant NIPBL allele is sufficient to generate severe developmental phenotypes, but sister chromatid cohesion is not overtly affected. The NIPBL and KIAA0892 proteins are respectively orthologs of S. cerevisiae Scc2 and Scc4, that form a cohesin-loading complex to regulate sister chromatid cohesion. Metazoan homologs of Scc2 and Scc4 work as cohesin-loading complexes and are also important in various processes in developmental gene regulation, including in axon and cell migration. In order to model CdLS and study Nipbl function we used Cre-loxP gene targeting to generate a conditional Nipbl allele. Inactivation of both Nipbl alleles using Pgk-Cre resulted in early embryonic lethality. Heterozygotes showed a high degree of perinatal mortality, with clear evidence of prenatal growth retardation and also postnatal growth retardation in the case of surviving newborns. Heterozygotes also showed evidence of skeletal defects that could be attributed to developmental delay effects and in some cases to patterning defects. When we used a Wnt1-cre strategy to inactivate both Nipbl alleles in cranial neural crest derivatives, we found a remarkable array of craniofacial abnormalities that phenocopied hallmark features of CdLS with high penetrance. These include microcephaly, micrognathia, cleft palate and low-set ears. In addition, neural crest-derived cranial bones were largely absent or severely stunted.

## 328

**The biochemistry and evolution of the Williams-Beuren syndrome-associated gene *GTF2IRD1*.** *S.J. Palmer[1], N. Santucci[2], J. Widagdo[1], S. Bontempo[2], K.M. Taylor[1], E.S.E. Tay[2], P.W. Gunning[3], E.C. Hardeman[1].* 1) Department of Anatomy, University of New South Wales, Sydney, NSW, Australia; 2) Muscle Development Unit, Children's Medical Research Institute, Westmead, NSW, Australia; 3) School of Medical Sciences, Department of Pharmacology, University of New South Wales, Sydney, Australia.

Williams-Beuren syndrome (WBS) results from a hemizygous microdeletion within chromosome 7q11.23 involving 28 genes. Its features typically involve characteristic physical abnormalities and a set of cognitive and behavioural symptoms that are collectively called the Williams syndrome cognitive profile (WSCP). Haploinsufficiency of elastin (encoded by *ELN*) causes the supravalvular aortic stenosis. The remaining physical and neurological symptoms have not yet been attributed to specific genes, but it is assumed that they are due to haploinsufficiency for a subset of the remaining 27. Genotype/phenotype correlations in patients with smaller deletions have mapped the typical craniofacial dysmorphology and WSCP to a pair of evolutionarily-related DNA-binding proteins, GTF2IRD1 and GTF2I. We have generated *Gtf2ird1* knockout mouse lines by deletion of exon 2 and replacement with a LacZ reporter allowing us to map expression and examine the consequences of gene inactivation. These mice show developmental abnormalities and cognitive impairments that reflect aspects of the disease. In an effort to understand the underlying biochemistry of these features, we have developed studies that focus on GTF2IRD1 function. Using phylogenetic footprinting analysis of the *GTF2IRD1* upstream region, we have found a conserved sequence adjacent to the transcription start that contains a cluster of canonical GTF2IRD1 binding sites. GTF2IRD1 binds to this region with high affinity and negatively regulates its own transcription. Binding is contingent upon the presence of multiple DNA recognition sites and this has permitted us to model the DNA binding properties of GTF2IRD1 and predict its potential range of target genes. We have used phylogenetic comparisons of GTF2IRD1 protein and its relatives to develop an understanding of the key functional domains. A series of internal duplications has created multiple conserved repeat domains (RDs) with a novel peptide fold unique to this small group of proteins. We are analysing the function of the RDs and other conserved domains in a modular fashion using yeast 2-hybrid interactions. These studies have demonstrated that GTF2IRD1 is specifically targeted for sumoylation and analysis in cell culture has revealed a mechanism for regulation of GTF2IRD1 protein levels. In combination, these data indicate that the levels of GTF2IRD1 are tightly regulated at transcription and after translation.

## 329

**The First Identified Mutations in *VANGL2* Associated with Human Neural-Tube Defects.** *H. WANG[1], T. ZHANG[2], Y. LEI[1], L. WANG[2], H. LI[3], Y. SHI[1], Z. JIANG[1], Y. CHEN[3], Q. YANG[1], B-L. WU[1,4], L. JIN[1].* 1) School of Life Sciences, Institute of Biomedical Sciences, Fudan University, Shanghai, 200433, China; 2) Capital Institute of Pediatrics, Beijing, China; 3) Maternal-Child Medical Center, Suzhou General Hospital, Suzhou, China; 4) Children's Hospital Boston and Harvard Medical School, Boston MA, USA.

Neural-tube defects (NTDs) are most frequent congenital abnormalities in humans ranging from benign dysraphisms such as dermal sinus to severe dysraphisms such as craniorachischisis. NTDs behave as a complex trait with genetic heterogeneity, incomplete penetrance, variable expressivity, which further modulated by environmental factors like pericomceptional folic acid supplementation. To date VANGL1 mutations have been associated with human NTDs. No mutation in VANGL2 has been identified yet by a few studies, although Vangl2 mutations in loop-tailed mice have been linked to severe NTDs. We have identified three mutations in the VANGL2 gene (S84F, R353C, F437S) and two in the VANGL1 gene (G39S, N313S) in 172 miscarriage or stillborn samples of Chinese subjects with severe NTDs (familial or sporadic). All identified mutations are disease-specific, which have not been found in 600 matched controls. In a protein-protein interaction assay, R353C partially and F437S completely abrogated the interaction of the VANGL2 protein with its binding partners, disheveled-1, -2, and -3. The question of whether VANGL2 mutant alleles have preferred expression between the proband and carrier in a family was probed. Our hypothesis was substantiated by the finding of monoallelic expression pattern in the VANGL2 gene using a few different tissue samples from 5 NTDs cases and 1 control individual. For an example, the NTDs proband (R353C, 1543C>T) expressed the mutated VANGL2 allele (1543T) in the meninges tissue and the wild allele (1543C) expressed in both kidney and lung tissues. Furthermore we found differing methylation statuses in the VANGL2 promoter of both alleles, of which the promoter with hypermethylation correlated with the inactivated allele and the promoter with hypomethylation correlated with the activated allele. To our knowledge this is the first instance that disease-specific mutations in the VANGL2 gene have been identified in human NTDs evidenced by genetic and functional analysis. These findings confirm genetic defects of VANGL2 associated with human neural-tube defects, and its tissue-specific monoallelic expression pattern correlated with hemimethylation promoter, which may contribute to the different NTDs expressivity and penetrance.

## 330

**Preaxial polydactyly caused by Gli3 haploinsufficiency is rescued by Zic3 loss of function.** *M. Quinn, J. Purnell, S.M. Ware.* Cincinnati Children's Hospital and University of Cincinnati College of Medicine, Cincinnati, OH.

Gli3 and Zic3 are two members of the Gli superfamily of zinc finger transcription factors required for normal development. Mutations in ZIC3 cause X-linked heterotaxy whereas mutations in GLI3, an important mediator of hedgehog signaling, cause Pallister-Hall syndrome, Greig cephalopolysyndactyly, and isolated limb malformations including pre- and postaxial polydactyly. To examine the expression of Zic3 during limb development, we generated transgenic Zic3-LacZ bacterial artificial chromosome lines as well as performing whole mount in situ hybridization. The results indicate that Zic3 is expressed in the limb bud in a domain that overlaps with Gli3. Zic3 and Gli3 have previously been shown to physically interact via their zinc finger binding domains. To investigate the effect of this interaction in vitro, we performed transactivation assays. Surprisingly, the results indicate that Zic3 converts Gli3 from a repressor to an activator. During limb development, post-translational processing of Gli3 from activator to repressor is important to antagonize the potent morphogen sonic hedgehog and limit it to the posterior domain, the zone of polarizing activity. To investigate the interaction of Zic3 and Gli3 in vivo, we intercrossed Zic3 knock out mice with the XtJ polydactylous mouse mutant that lacks Gli3 function. Analysis of fetuses at E17.5 demonstrates preaxial polydactyly in XtJ heterozygous fetuses and severe polydactyly in null (homozygous XtJ / XtJ ) fetuses. Zic3 null; Gli3 heterozygous fetuses show a complete rescue of the polyldactylous phenotype whereas a partial rescue of severe polydactyly is identified in Zic3 null; Gli3 null fetuses. Analysis of molecular markers by in situ hybridization indicate that ectopic sonic hedgehog expression in Gli3 null embryos at E12.5 is normalized in Zic3 null;Gli3 null embryos indicating that absence of Zic3 functions to suppress ectopic hedgehog signaling. These results indicate that two Gli superfamily members that cause disparate human congenital malformation syndromes interact genetically and further suggest that Zic3 plays a previously unrecognized role in regulating limb digit number via its modifying effect on Gli3 activator versus repressor levels.

## 331

**A mouse model for heterotaxy spectrum heart defects.** *A. Haaning, M. Quinn, S.M. Ware.* Cincinnati Children's Hospital and the University of Cinncinnati College of Medicine, Cincinnati, OH.

Mutation of Zinc finger of the cerebellum 3 (Zic3) in humans leads to X-linked heterotaxy syndrome, which is characterized by left-right patterning abnormalities and complex congenital heart defects. Zic3 encodes a C2H2-type zinc finger transcription factor that is necessary for early patterning during gastrulation and left-right axis determination. Zic3 null mice have been well characterized and often die during embryogenesis from gastrulation or heart defects, or soon after birth from exencephaly or neural tube closure defects. In order to generate a mouse line for tissue-specific deletion of Zic3, a neomycin cassette flanked by FRT sites was inserted into intron 1 of the murine Zic3 locus. Introduction of this cassette resulted in embryonic lethality of conditional hemizygous males and conditional homozygous females. Zic3 expression was quantified by real-time amplification from E12.5 total RNA. In conditional embryos, Zic3 expression was approximately three to four times lower than in wild-type siblings, indicating that the neomycin cassette results in a hypomorphic allele. Analysis of the embryonic phenotype indicated a much higher penetrance of heart defects, neural tube closure defects, and exencephaly than Zic3 null embryos; however, there is no evidence of early embryonic lethality due to gastrulation defects as with Zic3 null embryos. More than 50% of E12.5 Zic3 conditional embryos have heart defects as compared to less than 20% of E12.5 Zic3 null embryos; however, the specific types of cardiovascular malformations identified in both mouse lines are similar. Furthermore, these defects, which are caused by reversed heart looping, aberrant structural patterning, and/or delayed development, recapitulate those seen in the human heterotaxy population. Because of these similarities and a higher penetrance of heart defects than Zic3 null mutants, Zic3 conditional mice will be an excellent model for analysis of heart defects resulting from heterotaxy syndrome.

## 332

**Developmental Underpinnings of Apparently Acquired Aortic Aneurysm in Marfan Syndrome Revealed by Analysis of Mutant Embryonic Stem Cells.** *M.E. Lindsay[1,2,3,4], I.J. Domian[4], K.R. Chien[4], H.C. Dietz[1,3].* 1) Institute of Genetic Medicine; 2) Division of Pediatric Cardiology, Department of Pediatrics; 3) HHMI, Johns Hopkins Hospital, Baltimore, MD; 4) Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA.

Marfan Syndrome (MFS), caused by deficiency of fibrillin-1, shows early death due to aortic aneurysm. Increased TGFβ signaling contributes to aneurysm progression. Cells derived from the ascending aorta of patients and mice with MFS show many phenotypic abnormalities (high expression of AT1, MMPs, and collagens) that are reminiscent of myofibroblasts (Myo-Fib) that typically derive from epithelial-to-mesenchymal transition (EMT), a process driven by TGFβ. Using lineage-tracing studies, we show that many cells within the aortic root of Marfan mice derive from the endothelium through endothelial-to-mesenchymal transition (EnMT), a specialized form of EMT, while all VSMCs in this vascular segment in wild-type animals derive from the anterior heart field (AHF). Pathologic EnMT in the Marfan mice begins postnatally, and extends from the valve annulus to the sinotubular junction, perfectly describing the zone of predisposition for aneurysm formation in MFS. Using lineage tracing we show that endothelial cells in this region uniquely derive from the AHF, suggesting that cellular ontogeny dictates regional variation in the predisposition for EnMT and aneurysm formation. To gain further insight into the basis for this predisposition, we have undertaken detailed phenotypic analysis of *Fbn1$^{C1039G}$* homozygote mice that show profound fibrillin-1 deficiency. These animals demonstrate failed morphogenesis in structures populated by the AHF including right ventricular cardiomyopathy, aortic-mitral valve discontinuity, and proximal great arterial hypercellularity with longitudinal overgrowth. To investigate the performance of AHF cells in the context of severe fibrillin-1 deficiency, we have generated ES cell lines harboring fluorescent reporter alleles that allow for the specific isolation of the AHF population. *Fbn1$^{C1039G}$* homozygote cell lines show a reproducible expansion of the AHF progenitor population during in vitro differentiation assays versus wild type cells. Consistent with our in vitro data, lineage- tracing analyses using newborn *Fbn1$^{C1039G}$* heterozygous and homozygous animals demonstrate in vivo expansion of this cell lineage in the proximal aorta in terms of both cell number and transmural distribution. We are currently using human iPS (induced pluripotent stem cells)to further interrogate the hypothesis that altered performance of AHF-derived progenitors drives EnMT and disease progression, a finding that would have immediate therapeutic implications.

## 333

**Calpain as a Therapeutic Target in Inherited Aortic Aneurysm: Lessons from Rare Mendelian Disorders.** *D. Kim[1], N. Patel[1], M. Lindsay[1,2], E. Goldmuntz[3], A. John[3], J. Garbarini[3], H. Dietz[1,2].* 1) Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 2) Howard Hughes Medical Instutute, Chevy Chase, MD; 3) Department of Genetics, Children's Hospital of Philadelphia, Philadelphia, PA.

We identified a complex mutant allele (R527C/P2423S) of the filamin A gene (FLNA) on the X chromosome in a male patient with a syndromic presentation of tetralogy of Fallot with ascending aortic aneurysm (ToF/AAA). Subsequent screening of individuals with nonsyndromic ToF/AAA revealed a recurrent missense mutation (A1764T) in 3 unrelated patients. These mutations were not observed in over 500 control chromosomes. A1764T occurs at the calpain cleavage site of filamin A. Patient-derived cells harboring the R527C/P2423S allele showed abnormally small cleavage products that decreased in abundance in the presence of the calpain inhibitor MDL28170, but increased upon treatment with ionomycin, a cacium ionophore that increases intracellular calcium and promotes calpain activity. Both the R527C/P2423S and A1764T alleles were associated with a marked reduction in full-length filamin A (9.3 and 19.3% of wildtype, respectively), with restoration of levels after MDL28170 treatment. Taken together, these data suggest that reduced filamin A levels and/or accentuated calpain cleavage contributes to the pathogenesis of disease, including aortic aneurysm. Filamins contribute to the organization and stability of the actin cytoskeleton, integrate cellular signaling cascades, and regulate diverse cellular functions including adhesion and motility. In other aneurysm disease states, our lab has shown that increased TGFβ activity induces cells within the aortic media to undergo endothelial-to-mesenchymal transition (EMT), resulting in myofibroblasts with many deleterious behaviors including expression of matrix-degrading enzymes. TGFβ has also been shown to upregulate calpains. We hypothesized that cleavage of filamin A by calpain may be required for TGFβ-induced EMT. In this scenario, mutations that promote filamin A cleavage, as in ToF/AAA, might sensitize cells to TGFβ-induced EMT, culminating in aneurysm formation. To test this hypothesis, NMuMG epithelial cells were treated with TGFβ to induce EMT in the presence or absence of MDL28170. The addition of calpain inhibitor blocked TGFβ-induced EMT in a dose-dependent manner (reduced expression of α-smooth muscle actin and vimentin). Calpain inhibition also preserved epithelial cell morphology after TGFβ treatment. These data inform the pathogenesis of inherited aortic aneurysm and suggest that calpain inhibitors may find broad application in other diseases driven by pathologic EMT including cancer and many fibrotic states.

## 334

**DSCAM and COL6A2 from the Down syndrome heart critical region interact to generate congenital heart defects.** *T.R. Grossman[1], A. Gamliel[2], R.J. Wessells[3], O. Taghli-Lamallem[4], K. Jepsen[2], K.L. Peterson[5], J.R Korenberg[6], M.G Rosenfeld[2], R. Bodmer[4], E. Bier[1].* 1) Sec Cell & Dev Biol, Univ California, San Diego, La Jolla, CA; 2) HHMI, School of Medicine, University of California, San Diego, La Jolla, CA; 3) University of Michigan, Ann Arbor, MI; 4) Burnham Institute of Medical Research, La Jolla, CA; 5) School of Medicine, UCSD, La Jolla, CA; 6) School of Medicine, University of Utah.

Down syndrome (DS) is caused by trisomy or partial duplication of chromosome 21 and is the leading cause of congenital heart defects (CHD) in humans. Molecular studies of DS individuals with small duplications of chromosome 21 have defined a candidate region for CHD, which includes the SH3BGR, DSCAM, COL6A1, COL6A2, COL18, WRB, and HES1 genes. However, the roles of these candidate genes in CHD as well as neurological defects, remains poorly understood. We have employed both Drosophila and murine models to identify key genes contributing to DS CHD. As core developmental and physiological functions of the heart have been well conserved during the evolution of vertebrates and Drosophila, we systematically expressed mammalian candidate genes individually and in all possible pairwise combinations in the fly heart. This comprehensive combinatorial analysis revealed that DSCAM and COL6A2 collaborate to disrupt heart function in flies. Based on these findings in the fly, we generated a mouse model in which we individually expressed or co-expressed DSCAM and COL6A2 in the heart. Double transgenic mice exhibited several hallmark defects of DS CHD including atrial-septal defects and ventricular hypertrophy. This analysis illustrates the combinatorial potential of the fly and its use a multicellular screening system for identifying key interacting candidate genes in mice, and identifies DSCAM and COL6A2 as important multigenic contributors to signature congenital heart defects associated with DS.

## 335

**Why are people interested in personalized genomic risk information?** *B. Bernhardt[1], E. Gordon[2], S. Gollust[3], S. Keddem[1], C. Zayac[1], R. Pyeritz[1].* 1) Penn Center for the Integration of Genetic Health Care Technologies, Univ Pennsylvania, Philadelphia, PA; 2) Coriell Institute, Camden, NJ; 3) Robert Wood Johnson Foundation Health & Society Scholar Program, University of Pennsylvania.

Little is known about motivations and perceptions of people who are interested in obtaining personalized genomic risk information for common complex conditions. To address this gap, we surveyed people who signed up to attend an enrollment event/informed consent session for the Coriell Personalized Medicine Collaborative (CPMC), a research study examining the clinical utility of personalized genomic risk information. The survey assessed knowledge, perceptions, and motivations of people who registered to attend a CPMC enrollment event. Of the 210 people surveyed to date, 109 have responded (52% response rate). Primary reasons for participation in the CPMC included desire for information about disease risk (100%), finding out how to improve their health (96%), helping others through research participation (93%), and obtaining information at no cost (89%). Although most people correctly knew that their genomic testing would provide information about diabetes risk (85%) or side effects from drugs (67%), some incorrectly thought they would learn about their risk for all genetic diseases (28%), or their risk for having a child with a birth defect (7%), and 9% thought that gene therapy would be available to change their disease risk. Only 8% of respondents believed that their genetic information would definitively predict disease onset, and 97% agreed that common health conditions are caused by genes in combination with lifestyle and environment. Although respondents chose to participate in the CPMC, some thought there were risks to participating, including concern about obtaining insurance (25%), and learning about an increased risk for a disease they don't want to know about (38%). Nearly all (96%) participants planned to share their results with their doctor because they expected their doctor to tell them what to do to improve their health, or to prescribe them the best medicine based on their results. These findings indicate that people want genomic risk information so they can take action to improve their health in consultation with their doctor. Although most respondents do not support a deterministic understanding of genetics, some fail to understand the limitations of genomic testing and have unrealistic expectations about the benefits of genomic medicine. Additional public and professional education is needed to prepare for the future integration of genomics into health care.

## 336

**Demographic characteristics and perceptions of genetic testing among consumers of DTC Personal Genomics Services: The Scripps Genomic Health Initiative.** *C.S. Bloss[1,2], L. Ornowski[1,2], C. Lin[3], E. Silver[3], E. Levin[3], S. Kieran[3], A. DuRoss[3], S.S. Murray[1,2,4], M. Cargill[3], D.A. Stephan[3], V. Vanier[3], N.J. Schork[1,2,4], E.J. Topol[1,2,4].* 1) Scripps Genomic Medicine, Scripps Translational Science Institute, La Jolla, CA; 2) Scripps Health, La Jolla, CA; 3) Division of Applied Genomics, Navigenics Inc, Redwood Shores, CA; 4) Dept of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA.

The Scripps Genomic Health Initiative is a large longitudinal cohort study designed, in part, to assess response to testing among consumers of direct-to-consumer (DTC) personal genomics services. Participants purchase the Navigenics Health Compass product at a discounted rate and are administered baseline (pre-risk disclosure), as well as 3- and 12-month follow-up (post-risk disclosure) web-based behavioral health assessments. Here we present demographic characteristics and baseline data on perceptions of DTC genetic testing for a subset of N = 2,779 individuals in our study. A total of 81.6% of the sample self-reported Caucasian, 56.1% are female, and the mean age was 46.5 (SD 12.7). The modal education level was 16 years, and the modal household income was 100k-149k/year. The vast majority of participants reported having health insurance (98.8%) and most reported visiting their physician once (37.9%) or twice (18.9%) per year. Participants were asked about their perceptions of DTC genetic testing and to report any concerns they had about participating in the study. An individual could endorse multiple answers, and responses were as follows: 12.2% endorsed concerns related to learning about their disease risk; 15.5% concerns related to the quality and reliability of the results; 15.9% concerns related to not knowing how they would feel about their results; 34.7% concerns related to privacy issues about their data; and 43.8% indicated no concerns. Separate logistic regressions of each of these variables on gender, age, income, and education were conducted. Gender and age were significant predictors of concerns related to learning about disease risk and concerns related to not knowing how they would feel about the results; specifically, younger individuals and women were more likely to endorse these concerns. Younger age was associated with endorsing concerns related to the quality and reliability of the results, and younger age and higher education were both associated with concerns related to privacy issues. Finally, older age, higher income, and lower education were all associated with endorsing no concerns. Although these effects were statistically significant, effect sizes were small. These findings suggest that concerns among consumers of DTC personalized genomics services may vary as a function of demographics. For companies offering these services, provision of tailored education and counseling services may be warranted.

## 337

**Genomic counseling in the setting of an institutional initiative for Direct-to-Consumer personal genomic testing.** *E. Edelman[1], R.R. Sharp[1,2], C. Eng[1].* 1) Genomic Medicine Institute, Cleveland Clinic, Cleveland, OH; 2) Department of Bioethics, Cleveland Clinic, Cleveland, OH.

In January 2009, the Cleveland Clinic (CC) launched a genomic education initiative for physicians that included an opportunity for CC physicians to elect direct-to-consumer (DTC) SNP genotyping through 23andMe at no personal cost. Since the release of DTC SNP testing, medical and scientific experts have raised concerns that such testing is not yet ready for routine clinical application due to unknown clinical validity, utility of analyses, effect on behavior, and patient privacy. As genetics professionals we actively participated in this CC education campaign and created a system for providing accurate and accessible information regarding the risks, benefits, and limitations of DTC SNP profiling to interested staff physicians. This was a unique opportunity for physicians to explore the overlap of recreational and medical aspects of predictive SNP genotyping. Approximately 2800 staff and trainee physicians were given the opportunity to obtain 23andMe testing. The Genomic Medicine Institute (GMI) was charged with providing pre- and post-test education for individuals who chose to pursue DTC SNP genotyping. Speakers from GMI and the CC Bioethics Department (BD) presented a series of CME lectures to provide CC clinicians with education about genome-wide association studies, predictive SNP-based testing, ethical and legal implications of DTC/predictive genomic testing, the 23andMe product, and genetic counseling (GC) options. To ensure that CC's offer of genetic testing to employees was compliant with the Genetic Information Nondiscrimination Act, GMI met with CC Legal Counsel (LC) to identify legal requirements and provide recommendations to CC Executive Administration (EA). Through collaborative efforts between GMI, BD, LC, and EA, we were successful in developing a mutually agreeable and responsible model for post-test genomic counseling. Due to the complex nature of results interpretation, it is essential for genetics professionals to be available for pre- and post-test education. We offered 2 post-test options: an opportunity to discuss 23andMe results with a genetic counselor but no personal risk assessment or a comprehensive genomic counseling appointment with discussion of both 23andMe results and family history assessment. We hope that our experiences can be used as an example of a viable model for both education and GC as other hospitals and institutions develop proactive genomic education programs that target their own institutional needs.

## 338

**The Gene Partnership Project (GPP): Implementing the "Informed Cohort" model for the ethical recruitment and engagement of participants into genomic research.** *I.A. Holm[1,2,3], K.D. Mandl[2,3,4,5], B. Adida[2,3,5], C.M. Clinton[1,2], D.J. Nigrin[3,4,5,6], P.L. Taylor[3,7], L.M. Kunkel[1,2,3,8], I.S. Kohane[2,3,4,5].* 1) Division of Genetics and the Program in Genomics, Children's Hospital Boston, Boston, MA; 2) The Manton Center for Orphan Disease Research, Children's Hospital Boston, Boston, MA; 3) Department of Pediatrics, Harvard Medical School, Boston, MA; 4) Harvard Medical School Center for Biomedical Informatics, Boston, MA; 5) Children's Hospital Informatics Program at the Harvard-MIT Division of Health Sciences and Technology, Boston, MA; 6) Information Services Department, Children's Hospital Boston, Boston, MA; 7) Office of the General Counsel, Children's Hospital Boston, Boston, MA; 8) Howard Hughes Medical Institute at Children's Hospital Boston, Boston, MA.

The mapping of the human genome has allowed researchers to discover relationships between genotype and phenotype, and has the potential to allow genome-informed medical decision-making leading to diagnoses and therapies that are targeted, have reduced variability in outcome, maximize efficacy, and minimize adverse effects. As these studies generate information of greater health significance, there is emerging consensus that mechanisms for the ethical return of genomic information will be needed. To this end we have initiated a large genotype-phenotype project at Children's Hospital Boston (CHB), the "Gene Partnership Project" (GPP), which is a new model for the ethical recruitment of patients into genomic research protocols, and has been approved by the CHB Institutional Review Board. GPP is a longitudinal genotype-phenotype registry where data on de-identified participants will be collected and stored for use by researchers. However GPP is novel since, unlike other registries, we have developed a system to disclosure meaningful research results or incidental findings to participants in an ethical manner and without identifying the participant. GPP uses the CHB Indivo personally-controlled health record (PCHR) and a PCHR-based messaging system whereby participants can select which genetic results to receive. The seemingly paradoxical reconciliation of maintaining participant privacy, yet providing results, is achieved through targeting messages to specific participants based on their genetic data. Oversight of the communication of results to participants is crucial and will be implemented by the "Informed Cohort Oversight Board" (ICOB) comprised of ethicists, geneticists, lay people, and communications experts. We have termed this new research paradigm the "Informed Cohort" to reflect that participants are engaged in, and benefit directly from, genomics research while maintaining their privacy. While the enrollment of children presents unique ethical challenges, as the participant is a child and the receiver of information is the parent, the enhanced communication and the natural participation of the "family unit" in pediatrics is an advantage for the Informed Cohort, and we have developed PCHR models that parse parental and child autonomy. Our vision is to implement the Informed Cohort broadly to other medical practices and institutions. We believe that the Informed Cohort represents the future of genomic research and personalized genomic medicine.

## 339

**Ethical and Social Challenges in Whole Genome Resequencing: A Resequencing Study of Two Families with a Rare Disease.** *H. Tabor.* Ctr Pediatric Bioethics, Seattle Children's Hosp, Seattle, WA and Dept of Pediatrics, University of Washington.

Only a small number of individual genomes have been sequenced to date, including those of geneticists James Watson and Craig Venter. These individuals have been largely healthy and knowledgeable about the analysis and interpretation of genome sequence data. As whole genome sequencing becomes cheaper and faster, it is beginning to be applied to different kinds of individuals and populations. One target application is the study of rare Mendelian disorders whose causative genes have not been identified with other techniques. The ethical issues and challenges may be different in this context. Families may have different concerns about consent and return of results, and individuals and families with rare diseases may have different expectations and concerns about the kinds of variants that may be identified. We collaborated on a whole genome resequencing study of two families with a rare limb defect syndrome. To our knowledge, this is the first time both that an entire family has been resequenced and that this whole genome resequencing research has been disease-focused. This context raises new questions about consent, return of results, perceptions of risk and benefits, and concerns about privacy, confidentiality and identifiability. We will describe the approaches to consent and the return of results used in this study. We will specifically address the return of unexpected and clinically significant results to research participants. We will present qualitative analysis of data from interviews with affected and unaffected family members, and with research team members about the risk and benefits of the research and their experiences of the consent process and of the return of results. We believe this is the first empirical data analysis reported on research participants' expectations and experiences of participating in this kind of research. The results from this study can inform researchers and policy makers about the ethical dimensions of the design and conduct of whole genome sequencing studies in families and in disease population. They can also help inform future ethical empirical research in this area.

## 340

**Re-consent for deposition of research data in dbGaP: opinions of participants in a longitudinal cohort study.** *E.J. Ludman[1], L. Spangler[1], M. Fujii[1], S.B. Trinidad[2], S.M. Fullerton[2], W. Burke[2].* 1) Center for Health Studies, Group Health Cooperative, Seattle, WA; 2) Department of Bioethics and Humanities, University of Washington School of Medicine, Seattle, WA.

Existing cohorts represent an important source of data for genetic epidemiological investigation; however, ethical concerns have prevented clear consensus as to best practices for wide data-sharing. Past studies have mostly used hypothetical scenarios to examine participant preferences, particularly with regard to the nature and/or necessity of consent to address data sharing practices. As part of the multi-institution electronic MEdical Records and GEnomics (eMERGE) research study, one site (based at Group Health Cooperative, in Seattle, WA) decided that living participants in a longitudinal aging study should be asked for their consent to submit existing study data to the national database of Genotype and Phenotype (dbGaP). This provided a unique opportunity to study participant preferences with regard to re-consent. Of 1340 cognitively intact study participants contacted for re-consent, 1159 (88%) re-consented;152 (9.5%) refused consent, and 29 were ineligible. A subsequent telephone survey with 365 people who agreed to consent by mail (of 400 approached) examined the reasons for and against signing the consent and their satisfaction with the re-consent process. Respondents saw both benefits and risks in allowing submission of their data to dbGaP, although they spontaneously suggested a greater number and range of benefits of agreeing to such sharing. Trust in, and respect for, the researchers of the original study and the health care organization were important reasons for allowing consent. Most said it was very (69%) or somewhat (21%) important that they were asked their permission to send their health and genetic information to a databank. It would have been completely (27%) to somewhat (28%) unacceptable if they had been sent a letter that asked them to contact the researchers to opt out. A notification-only model would have been completely (47%) to somewhat (20%) unacceptable to the majority of participants. Similarly, respondents would find it completely (54%) or somewhat (16%) unacceptable if their research information had been added to the national databank without notifying or asking them for their permission. In summary, for this study sample, even though respondents were willing to allow their data to be shared, the majority thought it was important that researchers asked for their active consent. Further investigation of the opinions of other participant cohorts will better inform current data-sharing research policy and practice.

## 341

**Researcher Practices on Returning Genetic Research Results.** *S.B. Haga, G.M. Tindall, C. Heaney.* Inst Genome Sci & Policy, Duke Univ, Durham, NC.

As genetic and genomic research proliferates, debate has ensued about whether and how to return results to participants. However, little data exist of actual researcher practices. We surveyed genetic/genomic researchers to assess the frequency of which they considered, offered to, or actually returned research results, what factors influenced these decisions, and the method of communicating results. The sample group was comprised of corresponding authors of 488 articles reporting genetic variations associated with a disease phenotype in a U.S. population published in 2006-07. We invited the corresponding authors by email to participate in an online survey; 105 individuals completed the survey (24%). Based on likelihood ratio tests, we identified several significant correlations ($p<0.01$) between various independent variables and researchers' choice to offer to return or the actual return of individual results. When researchers offered to return individual results, the inclusion of minors in the study was significantly related with several disease types (diseases of the nervous system, the blood and blood forming organs, and endocrine, nutritional/metabolic and immune disorder diseases), funding source, research institution type, author's highest degree, and method used to report summary results. With respect to the actual return of results, we also found significant interaction effects between the inclusion of minors when paired with disease type (diseases of the blood and blood forming organs, and endocrine, nutritional/metabolic and immune disorder diseases) and author's highest degree. These results together suggest that the inclusion of minors in a study has a strong effect on whether results will be returned. Additionally, returning results had significant relationships with PhD-degreed researchers when paired with all other variables as compared to MD or MD/PhD-degreed researchers, indicating that authors with a PhD may be more likely to consider returning results than those with medical training. Disease type (blood and blood forming organs and endocrine, nutritional/metabolic disease, and immunity disorders) also showed a significant result when paired with degree, funding type and research institution type and tested in combination against outcomes in which researchers offered to or actually returned results. In summary, our data identify several factors important in the consideration and return of research results that warrant further study.

**342**

**Institutional Review Board Attitudes toward Human Research Protection in Genetic Research: Results from the Genetics Research Review and Issues Project (GRRIP).** *A.A. Lemke[1], H. Starks[2], G.L. Wiesner[3], K.L. Edwards[2], GRRIP.* 1) Northwestern University, Chicago, IL; 2) University of Washington, Seattle, WA; 3) Case Western Reserve University, Cleveland, OH.

A number of concerns and potential participant risks have been raised in the conduct of human genetic research; however, there have been no systematic or parallel studies to identify concerns about human genetic research among both researchers and individuals with Institutional Review Board (IRB) and research ethics expertise. The objective of this study was to assess IRB professionals' experiences with and attitudes toward review of genetic research. A parallel study was conducted to assess ASHG member views on the same issues and results of that component of the GRRIP will be reported separately. This investigation involved multiple collaborators including the University of Washington Center for Genomics and Healthcare Equality, Case Western Reserve University Center for Genetics Research Ethics and Law, the American Society of Human Genetics (ASHG), Public Responsibility in Medicine and Research (PRIM&R), and the GRRIP investigators. A cross-sectional study design was employed using an anonymous, web-based survey that was sent to PRIM&R members, a national organization representing IRB professionals. An initial study phase included focus groups and interviews with IRB professionals to help inform the development of the survey. The survey was sent to PRIM&R members on April 28, 2009, and as of June 1, 2009, 204/2,777 individuals had responded to the survey. Seventy-seven percent of the respondents were female, 59% had over 5 years of experience serving on an IRB, and 70% reported a biomedical focus of their IRB panel. Compared to other types of research, 75% of respondents felt that different guidance is needed for writing genetic research consent forms and half indicated that genetic research studies take more time to review. The majority (80%) of participants strongly or somewhat agreed that researchers have an ethical obligation to return individual research results that would affect a person's health or healthcare. Thirty-five percent of respondents indicated that they felt it would be very or somewhat likely that a participant would be personally identified in a study involving coded genetic data and 34% felt it would be very or somewhat likely for a research participant to be harmed as a result of this identification. With regard to broad consent, 58% strongly or somewhat agreed that this is acceptable to their IRB. Findings from this survey study will be used to facilitate consensus recommendations in human research protections.