The American Society of Human Genetics
60th Annual Meeting
November 2-6, 2010 Washington, DC

PLATFORM ABSTRACTS

1

The 1000 Genomes Project: A comprehensive map of common and low frequency human genetic variation. *G.R. Abecasis, The 1000 Genomes Project.* Center for Statistical Genetics and Department of Biostatistics, University of Michigan, Ann Arbor, MI.

We present a map of common human genetic variation obtained by whole genome-sequencing on a population scale. We describe the location, allele frequency and local haplotype structure of over 20 million SNPs, 1 million short insertion-deletion polymorphisms and thousands of structural variants, many of them novel. In addition, we show that sequence data from many individuals can be used to discover segregating sequences that are absent from the current genome assembly, that are polymorhpic between individuals and that vary in frequency across populations. As of May 2010 >500 genomes have been sequenced, and we estimate that the 1000 Genome resource will ultimately contain ~95% of accessible variants with a minor allele frequency of 1% in each of five continental groups. Our results show that low-coverage sequencing of many individuals can highly accurate individual genotypes at shared sites. By examing two deeply sequenced nuclear families, we estimate the rate of de novo germ-line point mutation to be 10^{-8} per base pair per generation and characterise factors influencing the rate of structural mutation. We identify many novel fixed differences of strong functional candidacy between populations and show that the hitch-hiking effect has a marked effect on genetic variation around genes. Project data are already being used in a large number of other studies and we show that the resource can be used for accurate imputation of common variants into existing association studies.

2

Population genomics in the Americas: sub-continental ancestry and its implications for medical genomics. *A. Moreno Estrada[1], M. Via[2], C. Gignoux[2], B. Henn[1], V. Acuña Alonzo[3], K. Bryc[4], H. Rangel Villalobos[5], S. Cañizales Quinteros[6], A. Ruiz Linares[7], E.G. Burchard[2], C.D. Bustamante[1].* 1) Department of Genetics, Stanford University, Stanford, CA; 2) Institute for Human Genetics, University of California at San Francisco, San Francisco, CA; 3) Molecular Genetics Laboratory, Escuela Nacional de Antropología e Historia (ENAH), Mexico City, Mexico; 4) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 5) Molecular Genetics Research Institute, Universidad de Guadalajara, Ocotlan, Mexico; 6) Molecular Biology and Genomic Medicine, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán (INCMNSZ), Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico; 7) Department of Genetics, Evolution and Environment, University College London, London, UK.

Human populations from the Americas are often of admixed origin, with significant genetic contributions from Native American and European populations (primarily involving local indigenous populations and migrants from the Iberian peninsula and Southern Europe) as well as West Africans brought to the Americas through the trans-Atlantic slave trade. As a result present-day Hispanic/Latino populations exhibit complex population structure, which poses additional challenges for characterizing their genetic makeup. In an effort to contribute to a better understanding of the genetic diversity in the Americas we have generated Affymetrix 6.0 genome-wide genotype data for more than 2,250 individuals from a diverse panel of populations including Mexicans, Colombians, Ecuadorians, Puerto Ricans, and Dominicans, as well as nearly 500 individuals from several Native American populations. We show at a finer scale that the complex historical events have affected patterns of genetic and genomic variation within and among present-day Hispanic/Latino populations in a heterogeneous fashion, resulting in rich and varied ancestry within and among populations as well as marked differences in the contribution of European, Native American, and African ancestry to autosomal, X chromosome, and uniparentally inherited genomes. We also stress the importance of characterizing Native Americans both for identifying potential source populations and for providing comprehensive reference panels when applying methods such as local ancestry estimation or admixture mapping. Also, as part of the 1000 Genomes Project, valuable resequencing data is becoming available for at least 280 genomes from individuals of Mexican, Puerto Rican, Colombian, and Peruvian origin, as well as nearly 60 genomes from HGDP populations, including some Native American genomes, being sequenced at 4x coverage as part of an independent sequencing effort. We are making use of both high-density genotype data as well as resequence data to fine map ancestry break points and understand diversity that is missed by current catalogs of human genomic diversity, which ultimately will help to better design epidemiological studies involving populations from the Americas.

3

Approaches for Discovering Structural Variation in the 1000 Genomes Project. *J. Korbel[1], R. Mills[2], C. Stewart[3], the Structural Variation Analysis Group of the 1000 Genomes Project.* 1) Genome Biology, EMBL, Heidelberg, Germany; 2) Molecular Genetic Research Unit, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 3) Department of Biology, Boston College, Chestnut Hill, MA.

The discovery of genomic structural variants (SVs) from next-generation sequencing data (NGS) is a relatively novel problem with implications for personal genomics. In the course of the 1000 Genomes Project (1000GP) pilot phase, a number of different SV discovery approaches have been developed, which underlie four different rationales: read-depth analysis, paired-end analysis, split-read analysis, and sequence assembly. Groups contributing to the 1000GP have generated approximately 30 SV discovery callsets using variations of these four rationales. Based on these, a set encompassing thousands of SVs is being released, the majority of which have already been successfully validated by PCR, microarray, and sequence assembly based approaches. We have comprehensively analyzed this high confidence SV set with a focus on deletions (i.e., SVs that can be identified by all four rationales) to assess differences in SV ascertainment between approaches along with influences of other factors, such as NGS library choice. Our analysis revealed that SV discovery approaches are partially complementary in terms of the types of SVs they detect (SV-mechanism, repeat/SD-content, SV-size, SV-type, and proximity/overlap with other genetic variants), with regard to how they can be experimentally validated, and with regard to their sensitivity and FDR - with each approach having unique benefits. In our presentation we will cover limitations of current approaches and discuss possible solutions to overcome these. We will further describe a conceptual pipeline for SV discovery in NGS data and comprehensively describe the processes the 1000GP SV analysis group has developed to go from raw individual sequencing data to a highly confident SV discovery set. This pipeline includes a framework for merging and combining information across different SV discovery-sets into a coherent high-confidence SV set. We will show that applying SV genotyping with this SV set enables filling in missing data from genomes sequenced at low coverage and facilitates the removal of sample outliers. Finally, we will discuss how the SV set released by the 1000GP may be used to leverage future NGS based association studies.

4

Characterizing the regulatory landscape of obesity-associated *FTO*. *N.F. Wasserman, J. Westlund, M.A. Nobrega.* Human Gen, Univ Chicago, Chicago, IL.

Genome-wide association studies (GWAS) routinely implicate non-coding DNA variants in the etiology of common complex disease. Such findings imply that many of these associated regions may harbor functional non-coding variants capable of altering the activity of long-range *cis*-regulatory elements - such as enhancers - that control the tissue and temporal expression patterns of nearby genes. Here, we describe a strategy to systematically define and characterize regulatory elements within a 47kb block of linkage disequilibrium (LD) in *FTO* that has shown a robust association with obesity. While no exonic variants have been identified in the associated interval, recent studies have shown that inactivation of *FTO* in mice protects against obesity; this suggests that alleles resulting in an up- or mis-regulation of *FTO* in humans could lead to an increased obesity risk. We utilized an *in vivo* bacterial artificial chromosome (BAC) enhancer trapping strategy in mice to interrogate the *FTO* obesity-associated interval for long-range *cis*-regulatory elements in the context of the endogenous *FTO* promoter. This technique identified enhancer activity in numerous tissues throughout the body, including the hypothalamus, consistent with endogenous *FTO* expression. We have fine-mapped those enhancers of greatest interested using a combination of *in vivo* transgenic reporter assays in zebrafish and mice, chromatin immunoprecipitation (ChIP)-seq in key tissues, and targeted recombineered deletions within our enhancer-trapping BAC. Together, these experiments help to define the regulatory landscape of *FTO* and contribute to our understanding of how tissue-specific mis-regulation may increase risk for obesity.

5

Estimating the Contextual Likelihood of SNP Incidence Genome-Wide and Negative Selection in the Exome via Whole Genome Sequencing of a Human Population. *J.R. Maguire[1], B. Neale[1,2,3], C. Cotsapas[1,2,3], M.J. Daly[1,2,3], B.F. Voight[1].* 1) Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA; 2) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 3) Department of Medicine, Harvard Medical School, Boston, MA.

The mutation rate is known to vary across the mammalian genome, in predictable context-specific ways; for example, C->T SNPs are vastly more likely within CpG islands. New sequencing technologies enable us to observe the genome at unprecedented levels of detail over larger sample sizes than ever before. The SNPs produced by the 1000 Genomes project facilitate not only better understanding of the influx of mutations into the human genome but also how new mutations are maintained or exit the genome due to negative selection or random chance. We present results of an analysis of whole-genome sequence data from the 1000 Genomes Project that shows over time single-nucleotide mutations enter the genome with predictable flanking sequence-specific relative likelihoods regardless of population cohort or locus-specific rate of conservation. We then estimate codon substitution bias in the exome relative to our expectation derived from the above table. We present tables of SNP incidence and Codon- and Amino-acid specific removal. The table of SNP incidence is consistent regardless of where it is computed from (correlation=0.99 across chromosomes or between CEU and YRI derived tables). Similarly the tables of Codon- and Amino-acid removal are consistent across populations and even in deep-coverage exome sequencing. We observe that mutations that have arisen in coding regions have a likelihood of being eliminated by negative selection that correlates strongly with their apparent a-priori “deleteriousness”: i.e. premature STOP codons are seen to be highly suppressed. Finally, we outline a case/control test statistic based on the estimate of the expected level of polymorphism per gene or locus; this framework complements existing statistical approaches for the aggregation of variants which are primarily based on frequency. We show results of applying these analyses in the form of a test-statistic to prioritize rare coding mutations in an Autism exome-sequencing project and Type 2 Diabetes whole-genome sequencing project.

6

Whole genome sequencing and accurate variant calls of a Japanese individual using massively parallel sequencing. *A. Fujimoto[1,2], H. Nakagawa[1], N. Hosono[1], K. Nakano[1], T. Abe[1], M. Nagasaki[3], R. Yamaguchi[3], T. Shibuya[3], M. Kubo[1], S. Miyano[2,3], Y. Nakamura[1,3], T. Tsunoda[1,2].* 1) Ctr Genomic Med, Riken, Yokohama, Japan; 2) Data Analysis Fusion Team, Computational Science Research Program, RIKEN, Japan; 3) Human Genome Center, Institute of Medical Science, University of Tokyo, Japan.

Massively parallel sequencing technology is a powerful tool for whole-genome variation discovery. On current platforms, accurate identification of variants is a critical but difficult task for application to personalized medicine. Here, we report accurate whole-genome variant calls and unbiased detection of rare variants of a Japanese male individual using a high-throughput sequencer to a 38x coverage of a human haploid genome. More than 99 % of sequence reads were mapped to human reference genome (hg18) by bwa and blast. We performed single nucleotide variant (SNV) calling with high accuracy (≥ 99.99 % concordance with SNP array) by read filtering and a Bayesian decision method. We found 3,818,174 SNVs, and the proportions of the novel SNVs were significantly different among SNV categories, suggesting strong biases in the current SNP database. Also, we identified 5,319 deletions smaller than 10kbp with high accuracy, as well as copy number variations and rearrangements. By de novo assembly of unmapped sequences, we indentified contigs that showed high similarity with EB virus and other human reference genomes. Our analysis suggests that considerable undiscovered variations still remain in human genome and that whole genome sequencing is a useful approach to understanding genetic variations.

7

Leveraging the 1000 Genomes Project for Next-Generation Microarrays. *M.A. Eberle[1], K. Kuhn[1], L. Galver[1], J.L. Stone[1], C. Tsan[1], K. Viaud[1], N. Burtt[2], J. Maquire[2], H.M. Kang[3], B. Voight[2], W. Winckler[2], P. Donnelly[4], M. Boehnke[3], M. McCarthy[4], G.R. Abecasis[3], D. Altshuler[2].* 1) Illumina, Inc., 9885 Towne Centre Drive, San Diego, CA 98121, USA; 2) Broad Institute, 7 Cambridge Center, Cambridge, MA 02142, USA; 3) University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109, USA; 4) Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK.

High-throughput sequencing is expanding the catalogue of variation at an unprecedented rate, enabling a more comprehensive understanding of the underlying linkage disequilibrium (LD) patterns within and across populations. This information is required to optimally design the next generation of whole-genome genotyping (WGGT) arrays that will allow the exploration of new hypotheses, including the role of lower frequency variants in disease. Until recently the primary source of information for developing WGGT arrays was the International HapMap Project as it provided detailed information on frequencies and LD between almost four million SNPs in three distinct populations. Selecting an optimal subset of these SNPs allows the creation of arrays that interrogate over 90% of all common HapMap variants either directly or indirectly through LD with 600-800k tagSNPs. These HapMap-based arrays have proven successful in evaluating common variants for association in many diseases and traits, for which, before the genome-wide association era, little was understood. However, the catalog of SNPs available through HapMap represents only a subset of the total variation in the human population. With the advent of high-throughput sequencing, the 1000 Genomes Project has greatly increased the spectrum of known variants and provides an excellent resource for content to develop the next generation of microarrays for “rich” genome-wide association studies (GWAS) - interrogating the entire genome, including rarer content down to <1% MAF. Using the phased genotype data from 180 samples sequenced in phase I of the 1000 Genomes Project, we have evaluated the whole-genome coverage provided by HapMap-based arrays. We calculate that less than 70% of the common (>5%) variants and less than 60% of all variants seen at least twice are tagged by the most comprehensive HapMap-based arrays. To bridge this coverage gap, we created an array (the Omni2.5) with almost 2.5 million tag SNPs selected using the 1000 Genomes data. These ~2.5 million markers can effectively cover 70-88% SNPs from the 1000 Genomes Project in all three HapMap populations, excluding singletons. Genotyping these SNPs across the 1000 Genomes samples also allowed us to evaluate the quality of genotype calls generated by the 1000 Genome Project; assess false positive rates in SNP discovery and categorize potential errors in 1000 Genome Project genotype calls.

8

Analysis of 1000 Genomes Exon Capture Pilot Data. *G. Marth[1], A. Indap[1], W. Leong[1], C. Hartl[2], K. Garimella[2], M. DePristo[2], F. Yu[3], R. Gibbs[3], The 1000 Genomes Project Exon Sequencing Subgroup.* 1) Dept Biol, Boston Col, Chestnut Hill, MA; 2) Genome Sequencing and Analysis, Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Human Genome Sequencing Center, Department of Human and Molecular Genetics, Baylor College of Medicine, Houston, TX.

At present, cost-effective strategies for sequencing hundreds or thousands of individuals entail either low-coverage (2-4X) whole-genome approaches or high-coverage targeted (typically exome) sequencing. Here we examine the high coverage sequence data in the coding regions of ~8,300 exons from 697 individuals sampled from seven different populations, generated by the 1000 Genomes Exon Sequencing Pilot Project. The data was collected using multiple DNA capture technologies combined with two different sequencing platforms (Illumina and 454). Boston College (BC), Baylor College of Medicine-Human Genome Sequencing Center (BCM-HGSC) and the Broad Institute (BI) independently implemented data analysis pipelines that made SNP and INDEL discoveries with high confidence. The results have been released at the DCC website (www.1000genomes.org) for public download. Analysis of all 697 samples by BC and BI yielded ~13,000 total overlapping SNP sites. We carried out extensive experimental validations (>1200 sites) using either Sequenom or PCR-Sanger pipelines, targeting various categories e.g. singletons, low frequency alleles, and functional sites. The validation results indicate that the accuracy of the SNP calls is very high overall with validation rate at 95-98% for rare SNPs including singletons. The deep coverage makes it possible to detect rare variants with very high sensitivity, and therefore, ascertain the low-frequency end of the site frequency spectrum with much better accuracy than achievable in low-coverage data. The 1000 Genomes Project aims to expand the exome sequencing program to deep whole-exome coverage in the same 2,500 individuals targeted for low-coverage sequencing. This will allow us to combine the advantages of the two approaches: detecting millions of variants from the genome-wide low-coverage data, and finding essentially all variants including rare alleles in coding regions. The ability to detect rare alleles in genomic regions of interest, and the modest cost compared to high-depth whole-genome sequencing, make capture-sequencing approaches attractive for medical re-sequencing studies. The extensive catalog of exomic variants we generate here will enable detailed understanding of the full frequency spectrum; linkage patterns of rare alleles; and normal mutational loads across human genes, informing other large-scale exome sequencing studies that will also be able to utilize the comprehensive analysis pipelines we developed.

9

De novo sequence assembly in the HLA, and other applications in medical genetics. *Z. Iqbal[1], G. McVean[2].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Department of Statistics, University of Oxford, South Parks Road, Oxford, United Kingdom.

Modern sequencing technologies are characterised by short read-lengths, high volumes of data, and low cost. We show that improved algorithms and high sequence coverage can compensate for short read-lengths, and allow us to focus on the really tough regions of the genome of serious medical interest. The Human Leukocyte Antigen genes are major determinants of immune response, implicated in many auto-immune and infectious diseases, and are also the most polymorphic genes in the genome - each gene harbours many hundreds of alleles. De novo assembly in this region using whole genome diploid data has hitherto proven intractable. We show that using the Cortex assembler we can de novo assemble a “graph” containing the sequence data of an individual in one “colour”, and all known HLA alleles each in their own separate colour. The genotype of the individual at an HLA gene can then potentially be determined (up to an accuracy that depends on the depth of coverage) by determining which colours (alleles) have the most support from the sequence data.

It is natural to also ask how one can use sequencing data to best understand polymorphism in the human genome, including the missing sequence that we know is not in the reference genome. As part of the 1000 Genomes Project, we have used de novo assembly to build an annotated pan-genome graph from 1.9Tb of data from 180 individuals from European, Yoruba, and Chinese/Japanese populations. This structure encodes polymorphism of all types, and population-specific allele frequencies. In doing so, we recovered 2.7 Mb of novel sequence (including 1.9Mb of human data from the NCBI databases) in a population-genetic context. For example, this novel sequence included a novel olfactory receptor gene transcript, absent from the European and East Asian populations, but present at intermediate frequencies in the Yoruba, and which was also present in chimp, macaque, dog and horse. This approach to population assembly has immediate applications to medical samples, where data from one or multiple samples can be added to the pan-genome graph in a different colour, and one is then able to draw inferences about the variants found in our samples in the context of all known human variation. Polymorphisms that are present in this sample colour at higher frequency than expected by comparing with the general population become candidates for causal variants.

10

TARGETED RESEQUENCING OF SLE SUSCEPTIBLE LOCI. *E. Rai[1], B. Wakeland[1], C. Liang[1], N. Nancy Olsen[1], D. Karp[1], G. Wiley[1], K. Kaufman[2], J. Harley[2], P. Gaffney[2], E.K. Wakeland[1].* 1) Department of Immunology, University of Texas Southwestern Medical Center at Dallas, Dallas, Texas, USA; 2) Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma, USA.

Susceptibility to systemic lupus erythematosus (SLE) is impacted by both genetic and environmental factors. More than 20 genetic loci are associated with susceptibility to SLE, however, the specific genetic lesions and functional variations responsible for the changes in gene function that cause the disease association are unknown. We selected 52 LD segments containing loci associated with susceptibility to SLE for sequencing using a targeted sequencing strategy. A computer program was developed to produce a tiling path of oligonucleotide probes capable of capturing all non-repetitive genomic sequences within the targeted LD blocks and these probe sequences were used to prepare custom slide arrays for sequence capture (Roche NimbleGen). 32 Caucasians females were selected for analysis, based on disease status and content of SNP genotypes at selected SLE susceptibility loci. Genomic DNA was processed for sequencing utilizing a modified protocol that combined the sequence capture protocols from NimbleGen with sequencing library protocols from Illumina. Sequence libraries were sequenced on the Illumina GAIIx genome analyzer using 75 bp Single/Paired End sequencing methodology. Raw sequence reads were assembled and aligned to the human genome using the CLCbiosystems Genomics Workbench. Out of the >14 gigabases of total sequence reads analyzed, 60-80% mapped to the targeted regions, yielding 25-35 fold coverage for all targeted regions. Sequence analysis of 25 LD blocks strongly associated with SLE identified 7129 SNPs and 663 Indels, of which 10.32% were novel variations. Of these, 13% were localized to gene segments that potentially impacted function. To explore the degree of functional variation that was associated with alleles tagged by SNPs strongly associated with disease, phylogenic networks were drawn using haplotypes formed by potentially functional SNPs in complete LD with SLE associated SNPs. This analysis identified a specific CLADE of alleles containing multiple SLE associated SNPs in *ITGAM* and identified extensive variations in functional SNPs among families of alleles that carried disease associated SNPs for many other genes. These results suggest that many disease-associated SNPs actually identify a cluster of functional variants, rather than a single disease allele.

11

Correction of growth defect of *Fgfr3*$^{Y367C/+}$ chondrodysplasic femurs by a novel tyrosine kinase inhibitor: ALG-31. *A. Jonquoy[1], E. Mugniery[1], C. Benoist-Lasselin[1], L. Le Corre[2], P. Busca[2], Y. Le Merrer[2], A. Munnich[1], L. Legeai-Mallet[1].* 1) INSERM U781, Paris Descartes University, Necker-Enfants Malades Hospital, 149 rue de Sèvres, 75015 Paris, France; 2) UMR 8601 CNRS, Paris Descartes University, 45 rue des Saints-Pères, 75006 Paris, France.

Fibroblast Growth Factor Receptor 3 (FGFR3) is a transmembrane tyrosine kinase receptor. Gain of function mutations in this gene lead to constitutive activation of the receptor in the absence of ligand and are responsible for autosomal-dominant skeletal dysplasias, including achondroplasia, the most common form of dwarfism, hypochondroplasia and thanatophoric dysplasia. The phenotype is characterized by rhizomelic disproportionate short stature, narrow trunk and macrocephaly. No treatment is available for these chondrodysplasias at the moment, thus incitating the development of new therapeutic approaches. In collaboration with UMR 8601 CNRS, our project was to synthesize new ATP-mimics as tyrosine kinase inhibitors able to inhibit the constitutive activation of FGFR3. These molecules were designed based on PD173074 pyrido[2,3-*d*]pyrimidine core with modifications of the alkylamino group at position 2 in order to establish better interactions with the FGFR3 active site. We obtained and tested on human chondrocytic cell lines about twenty new molecules. Among them, we selected the most promising one, ALG-31, based on a low IC 50 (<100 μM) and a high percentage of FGFR3 phosphorylation inhibition (>80%). We next determined the effect of ALG-31 on organotypic cultures of femurs isolated from E16.5 *Fgfr3*$^{Y367C/+}$ murine embryos. These mice display a severe dwarfism similar to human chondrodysplasias. The growth plate of long bones is disorganized, with reduced proliferative and hypertrophic zones. Femurs of *Fgfr3*$^{Y367C/+}$ mice treated with ALG-31 regained a normal lenght ($p<0,005$). In addition, the size of proliferative to hypertrophic chondrocytes was increased and the hypertrophic zone extended, as confirmed by collagen type X expression. We observed a surexpression of FGFR3 protein in mutant growth plates, which is decreased after treatment. We pointed out defects in proliferation of pathologic chondrocytes as shown by surexpression of PCNA and cyclin D1 in the prehypertrophic zone and their abnormal expression in the hypertrophic zone. Moreover, we observed in mutants abnormal expression of p57 in proliferative chondrocytes. Femurs treatment by ALG-31 reduced the expression of these three markers in both mutants and controls. To conclude, these results show that this novel tyrosine kinase inhibitor ALG-31 interacts with the activated FGFR3 and rescues bone growth by regulating the balance between proliferation / differentiation.

12

Characterization and Treatment of Beare Stevenson Syndrome Fgfr2 +/Y394C Mouse Model. *Y. Wang[1], X. Zhou[1], J. Friedenthal[1], DL. Huso[2], EW. Jabs[1].* 1) Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York, USA; 2) Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.

Beare Stevenson cutis gyrata syndrome is a rare human autosomal dominant condition characterized by craniosynostosis, cleft palate, hypertelorism, furrowed skin, acanthosis nigricans, digital anomalies, umbilical and anogenital abnormalities and early death. A Tyr375Cys mutation in the transmembrane domain of FGFR2 is the most common mutation associated with this condition. We generated and analyzed an inbred Fgfr2 +/Y394C mouse model on the C57BL/6J genetic background and elucidated the role of this mutation in the development of skeletal, dermatologic and other abnormalities. The mouse Fgfr2 Y394C mutation is analogous to the FGFR2 Y375C mutation in humans. The Fgfr2 +/Y394C mice showed postnatal retarded growth by P7 and 93% of the mutant mice died within 2 weeks of age. Necropsy of P0 Fgfr2 +/Y394C mice revealed malformations of different organs including the skull, sternum, skin, stomach mucosa and umbilicus. The mutant mice presented with synostosis or proximate fronts with disorganized cellularity in the developing coronal suture of the skull. Abnormal osteogenesis indicated by ALP staining was also observed at the coronal suture. Histopathologic examination of Fgfr2 +/Y394C mice revealed furrowed skin with acanthosis and hyperkeratosis. Ki67 staining and the expression of keratin 10, keratin 14 and loricrin at embryonic and newborn stages demonstrated increased or abnormal cell proliferation and differentiation in the mutant epidermis. Signaling pathway studies showed activation of mitogen-activated protein kinases (MAPK) in the Fgfr2 +/Y394C epidermis with an increase in phophorylated p38 as well as ERK1/2. Initial treatment of pregnant female mice with intraperitoneal injections of p38 MAPK inhibitor blocked phosphorylation and activation of p38 and partially rescued the skull and skin phenotypes in mutant offspring. Preliminary treatment with MEK/ERK inhibitor did not generate obvious therapeutic effects on mutant mice. Our studies demonstrate that the Fgfr2 Y394R mutation results in mice with cranial and skin features similar to those seen in the human Beare Stevenson syndrome, and that the p38 signaling pathway mediates these phenotypes. Selectively blocking p38 activation with certain inhibitors could prove beneficial in treating patients with Beare Stevenson syndrome and has implications for other FGFR-related craniosynostosis and skin conditions.

13

Life-Threatening Hypophosphatasia in Infants and Young Children: Results of Long-Term Treatment with ENB-0040, a Bone-Targeted, Enzyme Replacement-Therapy (ERT), and An Algorithm For Patient Management. *C.R. Greenberg[1], S. Craig[2], M. McGinn[2], J. Simmons[3], W. Russell[3], M. Bauer[4], N. Bishop[5], J. Taylor[6], N.J. Salman[7], M.A. Hamdan[7], M. Bober[8], J. Moore[9], R. Lutz[10], D. Wenkert[11], W.H. McAlister[12], A.M. Skrinar[13], H. Landy[13], M.P. Whyte[11].* 1) Dept Pediatrics & Child Hlth, Children's Hosp, Winnipeg, MB, Canada; 2) Royal Belfast Hospital for Sick Children, Belfast, Northern Ireland; 3) Vanderbilt Children's Hospital, Nashville, TN; 4) Arkansas Children's Hospital, Little Rock, AR; 5) Sheffield Children's Hospital, Sheffield, UK; 6) St. Vincent's Hospital, Green Bay, WI; 7) Tawam Hospital, Al Ain, UAE; 8) A.I. duPont Hospital for Children, Wilmington, DE; 9) St. John's Hospital, Springfield, MO; 10) University of Nebraska Medical Center, Omaha, NE; 11) Shiners Hospital for Children, St. Louis, MO; 12) Mallinckrodt Institute of Radiolog, St. Louis, MO; 13) Enobia Pharma, Cambridge, MA.

Background: Hypophosphatasia (HPP) is the heritable rickets/osteomalacia due to inactivating mutation(s) within the gene encoding the tissue nonspecific isoenzyme of alkaline phosphatase (TNSALP). This leads to extracellular accumulation of TNSALP substrates, including inorganic pyrophosphate (PPi), an inhibitor of mineralization, and pyridoxal 5'-phosphate (PLP), a form of vitamin B6. HPP severity ranges from perinatal demise due to respiratory compromise, to only dental problems during adult life. Mortality is high when HPP presents in infancy. Survivors can have rachitic deformity, significant weakness with delayed motor milestones, short stature, respiratory insufficiency, musculoskeletal pain, nephrocalcinosis, craniosynostosis, cranial shaping abnormalities and scoliosis. Results: Eleven infants and young children £3 years of age with life-threatening HPP received ERT using ENB-0040. Ten of 11 pts received from 6 to 18 mos of treatment with ENB-0040, 1-3 mg/kg SC thrice weekly, after a single 2 mg/kg IV dose of ENB-0040. One pt withdrew because of a moderate reaction during the infusion. All 10 pts who continued showed radiographic skeletal improvement. Nine demonstrated substantial to near complete healing of rickets and significant respiratory and/or motor improvement. Growth was largely preserved. Skull mineralization improved, but abnormalities of cranial vault shaping persisted and even progressed. One pt with craniosynostosis developed papilledema and underwent successful craniectomy. One pt, 8 mos into treatment, died of acute respiratory compromise and sepsis thought not to be due to ENB-0040. Baseline hypercalcemia, hypercalciuria, and/or suppressed serum parathyroid hormone levels resolved in parallel with skeletal radiographic improvement and declines in circulating PPi and PLP. Thus, restricted dietary calcium intake was liberalized. Nephrocalcinosis, when present at baseline, did not progress, and there was no evidence of ectopic calcification. An algorithm was developed to guide ENB-0040 dosing and dietary calcium management. Conclusions: HPP presenting in infancy is often life-threatening. Long-term ERT with ENB-0040 seems safe and well-tolerated. Improvements in mineralization are maintained and correlate with continued functional improvement. Cranial vault shaping abnormalities have not responded. As with other forms of rickets, treatment optimization of severe HPP requires attention to both pharmacological and mineral management.

14

Chemical treatment of muscular dystrophy that enhances skipping of the mutated exon in the dystrophin gene. *A. Nishida[1], N. Kataoka[2], Y. Takeshima[1], M. Yagi[1], H. Awano[1], M. Ota[1], K. Itoh[5], M. Hagiwara[3,4], M. Matsuo[1].* 1) Department of pediatrics, Kobe University Graduate School of Medicine, Kobe, Japan; 2) Medical Top Track Program, Medical Research Institute, Tokyo Dental and Medical University, Tokyo, Japan; 3) Laboratory of Gene Expression, School of Biomedical Science, Tokyo Dental and Medical University, Tokyo, Japan; 4) Department of Functional Genomics, Tokyo Dental and Medical University, Tokyo, Japan; 5) Department of Pathology and Applied Neurobiology, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kyoto, Japan.

Duchenne muscular dystrophy (DMD) is the most common and fatal muscle wasting disease caused by a loss of dystrophin protein. Currently, no effective treatment for DMD is available. One of the major therapeutic approaches is to convert from severe to mild phenotype restoring the reading frame by inducing exon skipping with antisense oligonucleotides (AONs). However, to prepare large amount of AONs is still expensive, and it is troublesome for patients to get intramuscular or intravenous injection of AONs regularly. Therefore, small chemicals that can modulate exon skipping are highly awaited for clinical use. Recently we found a dystrophinopathy patient who has a point mutation c.4303G>T in exon31 of the dystrophin gene that makes a stop codon p.Glu1435X. Unexpectedly, this patient expresses faint amount of truncated, but functional dystrophin protein. The analysis of the patient mRNA revealed that the mutation promotes the exon31 skipping and restores the open reading frame of dystrophin, and it was shown that the mutation disrupts exonic splicing enhancer and creates exonic splicing silencer. This means that enough amount of functional dystrophin protein could be obtained if we could induce more exon31 skipping in this patient. Therefore, we searched for small chemicals that enhance the exon31 skipping using the plasmid that contains mutant exon31 and flanking introns of dystorphin gene, and found that a CDC-like kinase (Clk)-specific inhibitor TG003 promoted the exon31 skipping in HeLa cells in a dose-dependent manner. The effect on wild type exon31 plasmid was also investigated and TG003 did not cause exon31 skipping. These results indicate that TG003 specifically promotes exon skipping with mutant exon31, but not with wild type exon31. Furthermore, TG003 promoted the exon31 skipping in the endogenous dystrophin gene in a dose-dependent manner and increased production of dystrophin protein in the patient muscle cells. These results indicate the possibility of a novel therapy of DMD. Our findings may open the way to develop a new type of tailor-made medicine of the fatal genetic diseases with small chemicals.

15

Long-Term Correction of a Mouse Model of Argininosuccinic Aciduria Using a Combination of Pharmacological Intervention and Liver-Directed *Asl* Gene Transfer. *O.A. Shchelochkov[1, 2], P. Campeau[1], S.C.S. Nagamani[1], A.K. Reddy[3], N. Brunetti-Pierri[1], D. Palmer[1], M.H. Premkumar[1], H.K. Garg[4], Y. Tang[4], N.S. Bryan[4], A. Erez[1], B. Lee[1, 5].* 1) Pediatrics/Genetics, University of Iowa Hospitals and Clinics, Iowa City, IA; 2) Current address: Department of Pediatrics, Division of Genetics, University of Iowa Hospitals and Clinics, Iowa City, IA; 3) Department of Medicine, Section of Cardiovascular Sciences, Baylor College of Medicine, Houston, TX; 4) Brown Foundation Institute of Molecular Medicine, University of Texas - Houston Health Science Center, Houston, TX; 5) Howard Hughes Medical Institute, Houston, TX.

Argininosuccinic aciduria (ASA) is the second most common urea cycle disorder caused by a deficiency of argininosuccinate lyase (ASL) which converts argininosuccinic acid to arginine. Biochemically, ASA is characterized by elevated argininosuccinic acid and depletion of arginine, a precursor of nitric oxide (NO). Individuals with ASA are prone to develop unique clinical phenotype, which is independent of hyperammonemia and includes intellectual disability, liver dysfunction, and hypertension. We have generated a hypomorphic ASA mouse model, which can survive beyond the newborn period. Affected mice suffer from multi-organ failure and hypertension. We show that impaired ASL activity results in impaired NO synthesis and propose that dysfunction of NO biology could explain several complications of ASA. To address both cell-autonomous and cell non-autonomous defects caused by ASL deficiency, we have combined the conventional treatment of arginine and sodium benzoate with sodium nitrite (NO donor) and murine *Asl* (m*Asl*) gene transfer to correct both NO deficiency and hyperammonemia. The liver-directed gene transfer was achieved using a helper-dependent adenoviral (HDAd) vector encoding m*Asl* under the control of a liver-specific promoter. Therapy with sodium benzoate, arginine, and sodium nitrite was started at birth and withdrawn at 5 weeks of life. The HDAd-m*Asl* vector was injected at age 4 weeks and resulted in a dramatic improvement of growth within 10 days post-injection. All HDAd-injected mice survived past 10 months while all saline-control mice died by 103 days (P = 0.001). We observed an improvement of plasma amino acids and liver ASL protein content in HDAd-injected animals. To evaluate the effect of the treatment on hypertension, we measured the response to acetylcholine of pre-contracted aortic rings in a relaxation test. HDAd-treated ASA mice failed to respond to acetylcholine suggesting a tissue-autonomous impairment of NO synthesis. These results suggest that gene therapy alone is not sufficient to correct the tissue autonomous NO deficiency. Nevertheless, at 10 months post-injection, the HDAd-injected mice displayed normal motor coordination and balance on the rotarod test, and normal learning and memory on the cued and contextual fear conditioning test. These data demonstrate the great potential of a combination of pharmacological and liver-directed gene therapy for the treatment of most although not all of the complications observed in ASA.

16

N-Acetyl Cysteine (NAC) Reverses Early- Stage Hepatic Phenotype of an Antisense Oligonucleotide Mouse Model of Niemann Pick Disease, Type C. *R. Fu[1,2], A. Incao[3], C.A. Wassif[1], W.J. Pavan[3], F.D. Porter[1].* 1) Program in Developmental Endocrinology and Genetics, NICHD, NIH, DHHS, Bethesda, MD, USA 20892; 2) Health Science Center, Peking University, Beijing, China 100191; 3) Genetic Disease Research Branch, NHGRI, NIH, DHHS, Bethesda, MD, USA 20892.

Niemann-Pick Disease, Type C (NPC) is an autosomal recessive disease characterized by excessive cholesterol and glycosphingolipids storage, and progressive neurological deterioration. Deficiency of either NPC1 or NPC2 leads to failure to efflux unesterified cholesterol and lipids from the late endosome/lysosome compartment. This primary defect in intracellular lipid transport initiates a pathological cascade that includes a deficiency of 27-hydroxycholesterol and neuroactive steroid, perturbed sphingosine metabolism, neuroinflammation, induction of apoptosis, and oxidative stress. Increased oxidative stress in NPC is supported by a number of *in vitro*, *in vivo*, and clinical studies. Based on these data, we hypothesized that N-acetylcysteine (NAC), a prodrug for glutathione supplementation, would reduce oxidative stress and thus potentially provide therapeutic benefit in NPC. To test this hypothesis we used a *NPC1* antisense oligonucleotide (AS oilgo) mouse model. This mouse model replicates many aspects of NPC liver disease and has the potential to be used for rapid *in vivo* testing of candidate drugs. Mice were treated with 1% NAC added to the drinking water. In comparison to untreated mice, we observed a significantly ($p<0.0001$) decreased liver to body weight ratio in NAC treated mice, and serum transaminase levels (ALT and AST) were reduced to control values. Consistent with the bioavailable cholesterol deficiency found in NPC, many SREBP2 target genes (SREBP2, HMGCS1, NSDHL, SQLE) were significantly elevated (more than 1.5 fold > control, $p<0.05$) in *NPC1* AS oligo mice compared to controls. NAC treatment reduced expression of these cholesterol homeostatic genes. Although total cholesterol levels were unchanged, initial results suggested that the fraction of unesterified cholesterol was significantly decreased ($p<0.05$). This series of experiments demonstrates the utility of the *NPC1* AS oligo mouse model for rapid screening of candidate drugs, and suggests that NAC treatment addresses one aspect of the NPC pathological cascade, and thus is a candidate drug for use in combinatorial therapy in the treatment of NPC.

17

Successful use of lipoplex for delivery of the small molecule ManNAc and the *GNE* gene to rescue a mouse model of Hereditary Inclusion Body Myopathy (HIBM). *T. Yardeni[1,4], C. Ciccone[1], S. Hoogstraten-Miller[2], D. Darvish[3], Y. Anikster[4], J. Nemunaitis[5,6], P. Maples[5], C.M. Jay[5], W.A. Gahl[1], M. Huizing[1].* 1) Med Gen Branch, NHGRI/ NIH, Bethesda, MD; 2) OLAM, NHGRI, NIH, Bethesda, MD; 3) HIBM Research group, Encino, CA; 4) Tel - Aviv University, Sackler Faculty of Medicine, ISRAEL; 5) Gradalis Inc., Dallas, TX; 6) Mary Crowley Cancer Research Centers, Dallas, TX.

HIBM is an adult-onset, progressive neuromuscular disorder, caused by *GNE* mutations. *GNE* encodes the ubiquitously expressed, key enzyme in sialic acid (SA) synthesis, UDP-GlcNAc 2-epimerase/ManNAc kinase. We created an HIBM mouse model, mimicking the Persian-Jewish founder mutation M712T. Mutant mice (-/-) unexpectedly died before day 3 of life (P3) from severe glomerulopathy due to hyposialylation, which could be partially rescued by oral supplementation of the SA precursor N-Acetyl-D-mannosamine (ManNAc). We assessed the efficiency and efficacy of ManNAc delivery in liposomes (ManNAc-Lipoplex), and human *GNE* gene delivery in liposomes (*hGNE*-Lipoplex) in our HIBM mouse model. Newborn pups (P1) were retro-orbitally injected with ManNAc-Lipoplex or *hGNE*-Lipoplex. Mice were watched for clinical signs and survival beyond P3, and tissues were tested at P5 for sialylation, histology, glomerular disease and *hGNE* expression. *hGNE*-Lipoplex injections yielded no surviving -/- pups beyond P3, however, wild type survivors showed *hGNE* expression in tested tissues at P5, indicating no toxicity of *hGNE*-Lipoplex and efficient gene delivery to tissues. Interestingly, -/- pups that died before P3 also showed *hGNE* expression in their tissues (as early as P2); treatment at P1 may not allow enough time for sufficient protein translation and SA production. In contrast, ManNAc-Lipoplex injections at P1 yielded survival beyond P3 in >90% of -/- pups. These pups showed improved sialylation of glomerular sialoproteins at P5. ManNAc-Lipoplex treated -/- mice continued to live beyond weaning; the oldest mice are now 4 months. The development of a muscular phenotype in these mice, similar to the symptoms of HIBM, and the effects of ManNAc treatment on these symptoms can now be assessed. Our studies demonstrate: 1) retro-orbital injection in newborn mice is an efficient method for systemic delivery of compounds; 2) small molecules can be efficiently delivered in Lipoplex; 3) systemic delivery of a gene in Lipoplex yields gene expression in tissues after one day; 4) ManNAc-Lipoplex can be applied to increase sialic acid levels and systemic or intramuscular ManNAc-Lipoplex therapy should be considered for the treatment of patients with HIBM, and may also be considered for other disorders of hyposialylation (i.e., certain cancers, certain renal disorders).

18

Over-expression of PGC-1α alleviates huntingtin protein toxicity by restoring mitochondrial activity and inducing reactive oxygen species (ROS) defenses. *T. Tsunemi[1], A.R. La Spada[1,2,3].* 1) Pediatrics, UCSD, La Jolla, CA; 2) Cellular & Molecular Medicine, UCSD, La Jolla, CA; 3) Institute for Genomic Medicine, UCSD, La Jolla, CA.

Huntington's disease (HD) is caused by expansion of a CAG repeat in the huntingtin (htt) gene, yielding a htt protein with an expanded polyglutamine (polyQ) tract that misfolds and is resistant to proteasomal degradation. Important clues to selective neuronal vulnerability in HD have been mitochondrial dysfunction and oxidative stress in striatum. Another feature of HD pathogenesis is production of polyQ-htt peptide fragments that localize to the nucleus and there disrupt transcription. While an interplay may exist among these features, the mechanistic basis of HD striatal degeneration remains to be elucidated. Previous studies have shown that mutant htt interferes with transcriptional programs co-activated by PPARγ co-activator 1α (PGC-1α), a key regulator of mitochondrial biogenesis. To test if restoration of PGC-1α function is sufficient to ameliorate neurological disease, we crossed HD N171-82Q transgenic mice with Rosa26-rtTA mice, and then with TRE-PGC-1α mice to develop "triple" transgenic mice. Induction of PGC-1α expression significantly improved neurological function in HD triple transgenic mice. Analysis of striatum and cortex revealed that mitochondrial activity was significantly increased in triple transgenic mice, and oxidative stress in striatum was significantly reduced due to activation of ROS response genes by PGC-1α. Improvement in the neurological phenotype in HD triple transgenic mice was accompanied by a marked decrease in htt aggregate formation in brain, including oligomeric insoluble species on filter trap assays. When we tested if striatal-like cells (ST-HdH) expressing wild-type htt (Q7/Q7) or mutant htt (Q111/Q111) could tolerate oxidative stress from hydrogen peroxide exposure, we found that Q111/Q111 cells exhibited significantly increased cell death, which could be rescued by PGC-1α expression. We also found that oxidative stress accelerated mutant htt aggregate formation and cell death in Neuro2a cells. Co-transfection of a PGC-1α expression construct similarly reduced mutant htt aggregate formation and caspase-3 dependent cell death in this system. Taken together, our data suggest that PGC-1α alleviates htt toxicity by activating mitochondria to improve bioenergetics status, and by inducing expression of the ROS defense system to reduce htt protein oligomerization. Thus, therapies aimed at restoring PGC-1α function hold great promise for treating HD in human patients.

19

Caspase mediated apoptosis in neural progenitor cells in mouse models of Bardet-Biedl Syndrome reveal novel mechanisms and treatment strategies for neonatal hydrocephalus. *T. Vogel[1], Q. Zhang[2,4], T. Moninger[4], D. Nishimura[2], C. Searby[2,4], G. Beck[2,4], K. Bugge[2,4], D. Thedens[2,4], R. Swiderski[2,4], M. Howard[1], V. Sheffield[2,3,4].* 1) Dept Neurosurgery, Univ Iowa, Iowa City, IA; 2) Dept. Pediatrics, Univ Iowa, Iowa City, IA; 3) Dept. Ophthalmology, Univ Iowa, Iowa City, IA; 4) Howard Hughes Medical Institute, University of Iowa Carver College of Medicine, Iowa, IA.

Hydrocephalus is one of the most common anomalies affecting the developing human nervous system with an estimated incidence of 0.8 to 2.5 per 1,000 live births. Hydrocephalus results from accumulation of cerebrospinal fluid (CSF) in dilated brain ventricles. The majority of neonatal cases are of an unknown etiology and identifying molecular mechanisms for neonatal hydrocephalus and developing medical treatment modalities to target these signaling cascades is a high priority. Here, we employ two mouse models of the human Bardet-Biedl Syndrome (BBS) to better understand the varying severity of neonatal hydrocephalus. BBS is a pleiotropic disorder caused by mutations in a group of genes that contribute to ciliary function. The disorder of the cilium in the CNS is seen in the ependymal cells lining the ventricles. Using a knockin model of Bbs1 (M390R/M390R), we characterize a less severe form of neonatal hydrocephalus when compared to a Bbs3 knockout mouse that models a more severe congenital hydrocephalus. We also identify a novel mechanism for the formation of hydrocephalus through caspase-mediated apoptosis of neural progenitor cells expressing Olig2. Olig2 is a sonic hedgehog induced transcription factor and is essential for the development of neurons, glia, and oligodendrocytes. Apoptosis is not observed in other neural stem cell populations expressing markers such as nestin, doublecortin, or glial fibrillary acidic protein (GFAP). Neonatal hydrocephalus, its associated motor deficit, and apoptosis in Olig2 positive cells are rescued in the Bbs1 knockin model using a polypeptide inhibitor targeting caspase 3 activation. This work provides insight into the neurodevelopmental mechanisms responsible for neonatal hydrocephalus and offers a potential novel therapeutic target for this common neurological disorder.

20

Bortezomib: An answer to the challenge of antibodies in diseases treated with therapeutics proteins? *P.S. Kishnani[1], S.G Banugaria[1], S.N Prater[1], S. Nampoothiri[2], J. Feldman[3], J.A. Kobori[4], J.K McGann[3], D.D. Koeberl[1].* 1) Department of Pediatrics, Division of Medical Genetics, Duke University Medical Center, Durham, NC, USA; 2) Amrita Institute of Medical Sciences & Research Centre, Cochin, Kerala, India; 3) Kaiser Permanente, Santa Clara, CA, USA; 4) Kaiser Permanente, San Jose, CA, USA.

The clinical use of therapeutic enzymes has expanded rapidly in the past decade with enzyme replacement therapies (ERT) now available for previously untreatable diseases. Therapeutic enzymes are potentially immunogenic, evoking antibody responses that may lead to decreased efficacy. Pompe disease is a lysosomal disorder caused by a deficiency of acid alpha glucosidase (GAA). The infantile form is characterized by cardiomyopathy, hypotonia, respiratory insufficiency and death by age 1-2 years. Since the availability of ERT with alglucosidase alfa the history of this disease has changed; several infants live longer, ambulate and remain ventilator free. A challenge is the immune responses to ERT, with loss of previous gains and death. Tolerance inducing therapies have been successful in patients at risk (CRIM negative) when administered in the naïve setting or after early exposure to ERT. Once high-titer anti-GAA antibodies are formed, tolerization therapy has uniformly failed to lower antibody titers or prevent clinical decline. Notably lacking from the therapeutic armamentarium are drugs that target long lived plasma cells whose elimination is vital in reversing entrenched immune responses. Bortezomib is a 26S proteasome inhibitor which targets mature antibody producing plasma cells. It is an FDA approved therapy for multiple myeloma. We report data from Pompe knockout mice demonstrating the unique reduction of anti-GAA antibodies following 4 doses of bortezomib. Dramatically, the first use of bortezomib, (4 doses over 2 weeks) in 2 Pompe infants suppressed the high-titer antibodies developed due to long term ERT exposure. Bortezomib was well tolerated in both cases and a clinical benefit was also noted. Patient 1 regained motor milestones. He also had reversal of cardiomyopathy that had recurred since development of antibody titers. Patient 2 showed acquisition of new motor milestones. This response to plasma cell depletion with bortezomib is unprecedented in the treatment of infants with Pompe disease following the formation of high titer anti-GAA antibodies. We are currently accruing additional data on experience with bortezomib in combination with other agents to suppress and induce tolerance in the setting of an established immune response. The novel use of bortezomib to reverse an entrenched immune response in patients with Pompe disease is a breakthrough in the field. It has implications for other diseases treated with therapeutic proteins.

21

Genome-wide patterns of allele frequency variation estimated from low-coverage Illumina sequencing of 2,000 Danish individuals. *K.E. Lohmueller[1], A. Albrechtsen[2], Y. Li[3], S.Y. Kim[1], T. Corneliusen[2], N. Vinckenbosch[1], G. Tian[4,5], E. Huerta-Sanchez[1], A. Feder[1,6], T. Jiang[3], I. Hellmann[7], O. Pedersen[8,9,10], J. Wang[2,3], R. Nielsen[1,2], the LuCamp Consortium.* 1) Integrative Biology and Statistics, University of California, Berkeley, Berkeley, CA; 2) Department of Biology, University of Copenhagen, Copenhagen, Denmark; 3) Beijing Genomics Institute, Shenzhen, China; 4) Beijing Institute of Genomics, Chinese Academy of Science, Beijing, China; 5) The Graduate University of Chinese Academy of Sciences, Beijing, China; 6) University of Pennsylvania, Philadelphia, PA; 7) Department of Mathematics, University of Vienna, Vienna, Austria; 8) Hagedorn Research Institute, Gentofte, Denmark; 9) Faculty of Health Sciences, University of Aarhus, Aarhus, Denmark; 10) Institute of Biomedical Sciences, University of Copenhagen, Copenhagen, Denmark.

We analyze low-coverage genome-wide next-generation sequencing data from 2,000 Danish individuals. Individual genotypes cannot be directly inferred from such data. However, we have developed a new statistical method to reliably estimate population allele frequencies of single nucleotide polymorphisms (SNPs). Our method eliminates the difficulty of inferring individual genotypes data by using the read counts and quality scores in the sample directly, without first attempting to estimate genotypes. Application of our new methods to the data allows the first estimates of genome-wide patterns of SNP allele frequency variation obtained from re-sequencing data. By comparing allele frequencies in different categories of sites, we are able to quantify the amount of selection acting in various non-coding parts of the genome. Interestingly, we also document a significant negative correlation between minor allele frequency and recombination rate. SNPs in regions of low recombination tend to have lower frequencies than SNPs in regions of the genome with higher recombination rates. Simulations using realistic demographic and selective models show that pattern is expected if positive natural selection is common in the human genome. Our results have important implications for understanding how natural selection has shaped patterns of variation across the human genome.

22

Evidence for archaic admixture in contemporary non-African human populations. *J. Long, R. Ferrucci, S. Joyce, K. Hunley.* Dept of Anthropology, Univeristy of New Mexico, Albuquerque, NM.

Analyses of large-scale genetic data sets show evidence for a series of founder effects that occurred as modern humans left Africa and settled the rest of the world. Nonetheless, research on modern humans has not ruled out the possibility that other processes, such as local gene flow, or mixing between archaic and modern humans, have also contributed to modern human diversity. Recent analyses of the Neanderthal genome make archaic admixture a salient issue because they show evidence for mixing between Neanderthals and out-of-Africa migrants. The present study examines evidence for archaic admixture in genotypes for 619 microsatellite loci collected from over 2,000 individuals from 100 human populations. We obtained these data from the Marshfield Clinic collection. The populations analyzed represent all inhabited continents of the world. In our analysis, we formulate the serial founder effects (SFE) model as a special case of a phylogenetic model promoted by Cavalli-Sforza and his associates. In this light, the SFE process makes four predictions: 1) A tree of descent according to the pattern of fissions. 2) The root of the tree lies in Africa. 3) The length of each branch is proportional to ratio of evolutionary time to effective population size. 4) The gene identity between all pairs of populations that share the same most recent common ancestor is equal in expectation. Using hypothesis tests based on generalized hierarchical statistical models, we find good agreement between the SFE predictions and diversity within and between African populations, and we find good agreement between the SFE predictions and diversity between non-African populations. However, there is more diversity within the non-African populations than the SRE model can account for. This makes for greater genetic distance between Africans and non-Africans than otherwise expected. How and where did the non-Africans obtain this diversity? A simple explanation for the finding is that the earliest migrants out-of-Africa mixed with an archaic population such as Neanderthals prior to their expansion throughout Europe and Asia. Coalescent based computer simulations of the SFE model with mixing support our interpretation. The time and place that we detect mixing coincides perfectly with that detected in a recent examination of Neanderthal genome sequences. Our study shows that genomic diversity in modern humans still reflects ancient events and processes.

23

Demographic inference using the pilot experiments of the 1000 genomes project. *S. Gravel, C.D. Bustamante, Thousand Genomes Project.* Genetics, Stanford University, Stanford, CA.

We discuss human demographic inference and allele frequency distribution using data from the pilot 1 and pilot 3 phases of the 1000 genome project. The pilot 1 phase consists of whole-genome, low coverage next generation sequencing (2x to 6x) of 180 samples from 4 HapMap populations (CEU,YRI, JPT, CHB), whereas the pilot 3 phase consists of deep exon capture sequencing (50x over 1000 genes) of 800 samples from 7 populations. Whereas the whole genome data provides millions of variable sites, the low coverage results in a large fraction of rare variants being missed (approximately 60% for singletons). Conversely, the targeted sequencing experiment allows for more precise determination of the rare variant frequency, but at thousands rather than millions of variant sites. We combine these two data sets to obtain accurate estimates of the joint site frequency spectra for individuals of European, Asian, and African origins. Using a diffusion approximation as implemented in the Dadi package (Gutenkunst & al, Plos Genetics, 2009), we obtain likelihood estimates for various demographic scenarios and infer parameters such as out-of-Africa divergence times, population growth rates, and migration rates. We discuss possible implications for the distribution of fitness effects, the missing heritability problem, and the design of large scale sequencing experiments.

24

Distorted population genetic principles in a super-exponentially growing population. *A.G. Clark[1], A. Keinan[2].* 1) Dept Molec Biol & Gen, Cornell Univ, Ithaca, NY; 2) Biol Statistics and Computational Biology, Cornell Unv, Ithaca, NY.

The global human population has been growing super-exponentially over the last 20-30 centuries, so that the current census size is approaching one million times the historical effective population size. This extraordinary situation implies a massive departure from equilibrium, such that human populations will continue to accumulate genetic variability for many millennia to come until a new equilibrium is reached. This presents an interesting theoretical problem regarding the structure of that variation, and encouraged us to revisit all the principles of population genetics in the context of this extraordinary demography. While many principles still hold, recent super-exponential growth produces an unusual site frequency spectrum, with nearly normal counts of SNPs in different frequency classes other than the very rarest classes, which are highly inflated. But this effect is only seen in sample sizes in the thousands. Our logistic branching process modeling also predicts a distortion of linkage disequilibrium in rapidly growing populations. The amplitude of random genetic drift is greatly attenuated as the population grows, and the probabilities of fixation and of loss are likewise reduced. Last, natural selection operates more efficiently in a large population and as a result, super-exponential growth has the effect of activating natural selection on for many standing variants. This phenomenon, combined with the skewed allele frequency spectrum of new mutations, produces key predictions for the frequency and characteristics of variants that are implicated in complex disease. All of these features show how poorly the historical effective size represents many population genetic attributes in a rapidly expanding population.

25

Coalescence-time distributions in a serial founder model of human evolutionary history. *M. DeGiorgio[1], J.H. Degnan[2], N.A. Rosenberg[1,3].* 1) Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Mathematics and Statistics, University of Canterbury, Christchurch, NZ; 3) Department of Human Genetics and the Life Sciences Institute, University of Michigan, Ann Arbor, MI.

Simulation studies have demonstrated that a serial founder model—in which populations expand outward from an initial source via a process where new colonies contain only subsets of the genetic diversity present in their parental colonies—can explain several patterns in worldwide genetic variation. Here, we provide analytical results for key quantities under the serial founder model, deriving distributions of coalescence times for pairs of lineages sampled from either the same colony or from different colonies. We use these distributions to obtain expected coalescence times and expected homozygosity and heterozygosity values. A predicted linear decline in expected heterozygosity with increasing distance from the source reproduces a pattern that has been observed both in human genetic data and in simulations. Our formulas predict that populations close to the source location have lower between-population gene identity than populations far from the source, also mirroring results obtained from human genetic data and simulations. Thus, our analytical framework is consistent with previous studies, and it lays a foundation for further investigation of trends in worldwide human genetic variation.

26

Characterization of rare exonic haplotypes in 697 individuals across 7 populations from 1,000 Genomes Pilot 3. *K.V. Garimella, C. Hartl, S. Schaffner, C. Sougnez, M. DePristo, M. Daly, D. Altshuler, S. Gabriel, The 1000 Genomes Project.* Broad Institute, Cambridge, MA.

The genetic architecture of rare, segregating variation in human populations at a genome-wide scale provides both a window into the fine details of human demography and evolution as well as an important background for interpreting rare variation in medical resequencing projects like the NHLBI Exome Sequencing Project. Here we discuss rare haplotypes discovered in the 1,000 Genomes project from a combination of deep exonic target resequencing (Pilot 3) as well as low-pass whole genome resequencing. Specifically, we map highly reliable (95-99% validation rate) SNP calls in the 8.5K sequenced exons in 697 samples in 7 populations (~100 samples/population) onto the large-scale haplotypes from the multi-population 1,000 Genomes WGS 4x coverage callsets. We characterize the novel rare haplotypes emerging from this data set, highlighting singleton and doubletons haplotypes unique to a single population as well as those shared across populations. Furthermore we derive coalescence trees for each segregating set of haplotypes to characterize the ancestral haplotype on which these rare haplotypes arose as well as their likely emergence time. We end with a discussion of the architecture of rare but segregating variation in normal human population and how this informs our approaches to analyzing medical resequencing projects.

27

The relationship between identity by state and identity by descent in population sequence data and its implications for genotype imputation. *K. Harris[1], H. Li[2], R. Durbin[1].* 1) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 2) Broad Institute, 5 Cambridge Center, Cambridge, MA.

When two DNA sequences are identical by state (IBS), meaning that they match along a stretch of genotyped markers, it is often assumed that the material comes from a recent common ancestor, inherited in one piece with no internal recombinations. Without this ability to infer identity by descent (IBD) from IBS, neither linkage analysis nor genome-wide association would be possible. We use coalescent theory to calculate how strongly IBS implies IBD as a function of sequence length, incorporating assumptions about marker density and population history. In addition to calculating p(IBD|IBS) between matching haploid sequences, we consider the problem of matching an unphased diploid sequence to a reference haplotype panel. The results have bearing on inference practices like haplotype phasing and the imputation of missing markers. Both practices become more reliable if the ends of an IBS alignment are not assumed to be IBD. To make the diploid computation possible, it was necessary to develop a new approximation to the full coalescent, one that may prove useful elsewhere. The key to computing p(IBD|IBS) is predicting the length distribution of IBS stretches in the genome. We correctly predict this distribution as measured in eleven complete human genomes once we account for sequencing errors. We also correctly predict the probability of true IBS given IBS at a thinned marker subset, the natural measure of imputation accuracy. The probability of IBS given thinned IBS varies with sequence length in a way that splits clearly along ethnic lines, as judged by data from five Africans, four Europeans, and two Asians. The predictions are consistent with different population histories that are compatible with what is known about the migration out of Africa and the later Eurasian split. Using the population histories that we fit to the genome data, we obtain ethnically specific plots of p(IBD|IBS) versus sequence length. We find IBS between Africans to be a better IBD predictor than IBS between non-Africans, suggesting that extra care should be taken when inferring IBD in bottlenecked populations. That being said, IBD can be inferred as accurately as desired by collecting dense enough marker data and working from the centers of sufficently long IBS stretches.

28

The simulation, detection and effect of inversion polymorphisms in the human genome. *P.F O'Reilly, K.S Basatena, C.J Hoggart, L.M Coin.* Epidemiology & Biostatistics, Imperial College London, London, United Kingdom.

The discovery of inversion polymorphisms in the human genome has been relatively minimal, largely because of the particular challenge that their detection presents. Here we present a statistical method designed to scan the genome for inversions from SNP data. The method highlights loci where there exists a bipartition of individuals showing no, or little, between-group recombination. This captures the main discernible feature of inversions from genotype data: that recombinations between the inverted and non-inverted types are suppressed. The method can also exploit knowledge about the location of inversions gained from sequencing project data to genotype individuals for inversion status in genotyped data sets, in cases underpowered for detection from genotype data alone. We also present new software, invertFREGENE, for simulating inversions in population genetic data. First we use this software to test our detection method, finding that our method has good power to detect even small inversions, and outperforms current alternatives. We apply the method to several large human data sets, building a genome-wide map of candidate inversions. Finally, we use invertFREGENE, as well as real data, to investigate the potentially disruptive effects of large inversions on population genetics methods of inference, in particular those for inferring recombination rates, for detecting selection, and for controlling for population structure.

29

Inferring recombination rates in recently admixed human populations. *D. Wegmann[1], K. Veeramah[1], D. Kessner[2], N. Freimer[3], J. Novembre[1,2].* 1) Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, CA; 2) Interdepartmental Program in Bioinformatics, University of California Los Angeles, Los Angeles, CA; 3) Center for Neurobehavioral Genetics, University of California Los Angeles, Los Angeles, CA.

To fully understand the evolution of recombination rates requires understanding how they vary among individuals and across populations. Here we focus on estimating recombination rates in recently admixed human populations. We take a novel approach based on identifying the ancestry of chromosomal segments. Wherever ancestry changes along a chromosome, a recombination event must have occurred in the history of the chromosome since admixture. Genome wide variation data, along with novel statistical methods, make it possible to infer the ancestry of local segments and identify ancestry switch points in large samples. Assuming a hidden Markov model, we compute the probability of ancestry switches between neighboring SNPs in the genomes of 3000 African Americans and compile a de-novo human recombination map. We account for the possibility that some observed recombination events may be identical-by-descent by a calibration based on simulations of African-American demography. Our recombination map allows us to characterize genomic regions with an excess or deficit in recombination when compared to existing maps. Finally, we aim to characterize the modulation of recombination rates by the genetic background of divergent populations, both at a genome-wide and local scale.

30

Age-dependent recombination events in human pedigrees. *J. Hussin, R. Gendron, M.-H. Roy-Gagnon, G. Andelfinger, P. Awadalla.* Faculty of Medicine, University of Montreal, Montreal, Quebec, Canada.

In humans, chromosome-number abnormalities in offspring have been associated with altered recombination and increased maternal age. The underlying causes of the latter association remain unknown, therefore age-related effects on recombination are of major importance, especially in relation to the mechanisms involved in human trisomies. In rodent oocytes, the frequency of recombination events has been found to decrease with age. We focus here on determining whether recombination rate is related to the age of the mother in humans. We localized crossovers at high resolution using a dense genome-wide SNP survey (6.0 Affymetrix platform) genotyped among French-Canadian multi-generation pedigrees, providing information about 195 maternal meiosis. Overall, we observed similar variation in fine-scale recombination rates and patterns as previously observed in Hutterites families (Coop et al. Science. 2008). However, contrary to what has been previously reported for humans (Kong et al. Nature Genetics. 2004), we observed that viable offspring of older mothers tend to have reduced recombination rates, in agreement with the findings in mouse and hamster. The most pronounced effect is seen for mothers over the age of 30. The observation is a genome-wide effect but, among submetacentric chromosomes, the effect was significantly more pronounced, suggesting a subtelomeric effect. Furthermore, we observed that in females, recombination frequencies drop dramatically in subtelomeric regions. Finally, we propose a model that reconciles our findings with earlier reports that found associations between maternal age and recombination in trisomy cases.

31

Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. *L. Conde[1], E. Halperin[2,3], N.K. Akers[1], K.M. Brown[4], M.T. Smith[1], P.M. Bracci[5], C.F. Skibola[1].* 1) Division of Environmental Health Sciences, School of Public Health, UC Berkeley, CA; 2) The Blavatnik School of Computer Science, Tel-Aviv University, Israel; 3) International Computer Science Institute, Berkeley, CA; 4) Integrative Cancer Genomics Division, Translational Genomics Research Institute (TGen), Phoenix, AZ; 5) Department of Epidemiology and Biostatistics, University of California, San Francisco.

Non-Hodgkin lymphoma (NHL) is a complex group of B- and T-cell neoplasms originating primarily from the lymph nodes. Family and epidemiological studies suggest an important role for genetics in the etiology of lymphoma, though the inherited genetic basis of the disease is largely unknown. Recently, we conducted a genome-wide association study (GWAS) of three common histological subtypes of NHL, follicular lymphoma (FL), chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL), and diffuse large B-cell lymphoma (DLBCL), using a pooled DNA genotyping strategy [1]. Due to the experimental and technical noise associated with pooled DNA GWAS, we have now conducted a new individual genotyping-based study on a larger subset of NHL samples from case-control study participants using a three-stage GWAS study design, with the aim of discovering new genetic variants associated with susceptibility to the major lymphoma subtypes. In the first stage, we conducted a GWAS using 213 FL, 211 CLL/SLL and 257 DLBCL cases and 750 controls from a case-control study of NHL based in the San Francisco Bay Area. The 40 SNPs with the lowest trend p-values for each NHL subtype were genotyped in Stage 2 in three independent population-based case-control studies. In Stage 3, validated SNPs were genotyped in additional independent NHL case-control populations. We identified a FL susceptibility locus in the MHC Class II region, which is independent of the rs6457327 locus in the MHC Class I PSORS1 region that we previously identified [1]. We also confirmed the positive associations previously reported [2] for CLL. We did not identify associated alleles in DLBCL, which may be, in part, due to the heterogeneity of this NHL subtype. Of note, we did not observe significant associations of the MHC region with risk of CLL/SLL or DLBCL, which suggest that the influence of MHC genes differs by NHL subtype. Further studies also will be necessary to elucidate the mechanisms by which MHC alleles mediate susceptibility to FL. REFERENCES 1. Skibola, C.F. et al. Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. Nat Genet 41, 873-5 (2009). 2. Di Bernardo, M.C. et al. A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. Nat Genet 40, 1204-10 (2008).

32

Genome-wide association study of pancreatic cancer in Japanese population. *S-K. Low[1,2], A. Kuchiba[3], H. Zembutsu[1], A. Saito[3,5], A. Takahashi[4], M. Kubo[4], Y. Daigo[1], N. Kamatani[4], S. Chiku[3,6], H. Totsuka[3,7], S. Ohnami[3], H. Hirose[8], K. Shimada[9], T. Okusaka[10], T. Yoshida[3], Y. Nakamura[1], H. Sakamoto[3].* 1) Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, the University of Tokyo, Japan; 2) Department of Medical Genome Sciences, Graduate School of Frontier Sciences, the University of Tokyo, Japan; 3) Genetics Division, National Cancer Center Research Institute, Japan; 4) Center of Genomic Medicine, RIKEN, Japan; 5) Statistical Genetics Analysis Division, StaGen Co., Ltd., Japan; 6) Science Solutions Division, Mizuho Information and Research Institute, Inc., Japan; 7) Bioinfomatics Group, Research and Development Center, Solution Division 4, Hitachi Government and Public Corporation System Engineering Ltd., Japan; 8) Department of Internal Medicine, Keio University School of Medicine, Japan; 9) Hepatobiliary and Pancreatic Surgery Division, National Cancer Center Hospital, Japan; 10) Hepatobiliary and Pancreatic Oncology Division, National Cancer Center Hospital, Japan.

Pancreatic cancer shows very poor prognosis and is the fifth leading cause of cancer death in Japan. Previous studies indicated some genetic factors contributing to the development and progression of pancreatic cancer; however, there are limited reports for common genetic variants to be associated with this disease, especially in the Asian population. We have conducted a genome-wide association study (GWAS) using 991 invasive pancreatic ductal adenocarcinoma cases and 5,209 controls, and identified three loci showing significant association (P-value< $5x10^{-7}$) with susceptibility to pancreatic cancer. The SNPs that showed significant association carried estimated odds ratios of 1.29, 1.32, and 3.73 with 95% confidence intervals of 1.17-1.43, 1.19-1.47, and 2.24-6.21; P-value of 3.30×10^{-7}, 3.30×10^{-7}, and 4.41×10^{-7}; located on or near to genes *FOXQ1* (6p25.3), *BICD1* (12p11.21) and *DPP6* (7q36.2), respectively. These genes have been implicated to play an important role in pathogenesis, telomere dysfunction and invasion of pancreatic cancer cells. Our findings may contribute to a better understanding of the pancreatic carcinogenesis.

33

Identification of a SNP in the *KLK3* gene which is associated with prostate cancer - the first coding SNP to be discovered in this disease. *Z. Kote-Jarai[1], A. Amin Al Olama[2], D. Leongamornlert[1], M. Tymrakiewicz[1], E. Saunders[1], N. Brown[3], M. Guy[1], G. Giles[4], G. Severi[4], J. Hopper[4], M. Southey[4], F. Hamdy[5], D. Neal[2], J. Donovan[6], K. Muir[7], J. Morrison[2], J. Harris[8], J. Clements[8], D. Easton[2], R. Eeles[1], The UK Genetics Prostate Cancer Study Collaborators.* 1) Oncogenetics, The Institute of Cancer Research, Sutton, Surrey, United Kingdom; 2) University of Cambridge, Cambridge, UK; 3) Centre for Cancer Therapeutics, The Institute of Cancer Reserach; 4) University of Melbourne, Melbourne, Australia; 5) University of Oxford, Oxford, UK; 6) University of Bristol, Bristol, UK; 7) Univeristy of Warrick, Warrick, UK; 8) Queensland University of Technology, Brisbane, Australia.

To identify common prostate cancer (PrCa) susceptibilility alleles, we conducted a multistaged genome-wide association study (GWAS). In the first stage (stage 1) 541,129 SNPs were genotyped in 1,854 PrCa cases with clinically detected PrCa diagnosed at <60 years or having family history of disease, and 1,894 population screened controls with low PSA level from the UK. We have identified 7 novel PrCa susceptibility loci one of which contains a strong candidate susceptibility gene *KLK3* which codes for PSA. PSA is widely used as a biomarker for PrCa detection and disease monitoring and our tag-SNP in this region lies between *KLK2* and *KLK3*. To refine the association of PrCa with variants in this region we imputed SNPs from HapMap2 and the 1000 Genomes that were not in our GWAS. We used MACH 1.0 software and the HapMap CEU population as a reference panel. At the location of *KLK3* on 19q13.3 we found a previously unidentified SNP, rs17632542, associated with PrCa at genome-wide significance, (P =$5.9x10^{-26}$). Using a Taqman assay we genotyped this novel SNP in our stage 1 and an additional stage 2 sample set of 3650 PrCa cases and 3940 controls from the UK and Australia. The combined analysis showed that the association of rs17632542 with PrCa is much more significant when compared with the original tag-SNP (P value $1.6x\ 10^{-24}$ compared with P value $2.3x10^{-17}$). The newly identified variant is a non-synonymous coding SNP in the *KLK3* gene causing an Ile to Thre substitution at aa179. Molecular dynamic modelling predicts that this substitution might have an effect on the stability of the protein. In addition this SNP is in an exonic splice enhancer (ESE) site. We conclude that this novel association signal is a strong candidate for a functionally important SNP within the *KLK3* gene and that this alteration at 19q13.3 has potential to be the causative variant influencing PrCa risk. Acknowledgements: This study was funded by Cancer Research UK, The Prostate Cancer Research Foundation, and NIHR support to the Biomedical Research Centre at The Institute of Cancer Research and Royal Marsden NHS Foundation Trust.

34

Prediction of prostate cancer - Improving the PSA test with information on 33 genetic risk variants. *M. Johansson[1], B. Holmström[2], S. Hinchliffe[3], A. Bergh[2], UH. Stenman[4], G. Hallmans[2], F. Wiklund[5], P. Stattin[2].* 1) International Agency for Research on Cancer (IARC), Lyon, France; 2) Umeå University, Umeå, Sweden; 3) University of Leicaster, Leicaster, UK 4) Helsinki University Hospital, Helsinki, Finland; 5) Karolinska Institutet, Stockholm, Sweden.

Background Genome-Wide Association Studies (GWAS) have identified multiple genetic variants associated with prostate cancer risk. How much these variants add to prostate-specific antigen (PSA) in prediction of prostate cancer is unknown. Method We created a genetic risk score based on 33 single nucleotide polymorphisms (SNPs) that had been associated with prostate cancer in previous GWAS. In a case-control study nested within the longitudinal Northern Sweden Health and Disease Cohort (NSHDC), 520 incident prostate cancer cases were identified by linkage to the regional cancer register along with 988 individually matched controls. Genotypes of the 33 SNPs, and concentrations of total PSA (tPSA) and free PSA were measured in DNA and plasma, respectively. Receiver Operating Characteristics (ROC) curves with Area Under Curve (AUC) estimates were used as validity measures for prediction of prostate cancer. Results AUC for the genetic risk score was 64.3% (95% confidence interval [CI]: 61.4%-67.2%) and AUC for tPSA and free to total PSA (%fPSA) was 86.2% (CI: 84.4%-88.1%). A risk model including the genetic score, tPSA and %fPSA increased AUC to 87.2% (CI: 85.4%-89.0%, Pdifference=0.002). At a specificity of 95%, the sensitivity for tPSA and %fPSA combined was 37.7% (positive/negative likelihood ratios [+/- LR]: 7.5/0.66), and adding the genetic risk score increased the sensitivity to 43.7% (+/- LR: 8.7/0.59). The AUC of the genetic risk score for low risk cancer was 65.5% (CI: 62.1%-68.8%) and 63.4% for high risk cancer (CI: 59.1%-67.6%). Conclusion The addition of a genetic risk score to PSA resulted in a modest, but notable improvement in prediction of prostate cancer.

35

Risks of colorectal and extracolonic cancers for carriers of monoallelic *MUTYH* mutation. *A.K. Win[1], S.P. Cleary[2,3], J.G. Dowty[1], N.M. Lindor[4], P.A. Newcomb[5], J.P. Young[6], D.D. Buchanan[6], R.W. Haile[7], L. Le Marchand[8], R. Green[9], J.L. Hopper[1], S. Gallinger[2,3], M.A. Jenkins[1], the Colon Cancer Family Regisry.* 1) Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The University of Melbourne, Parkville, Victoria, Australia; 2) Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada; 3) Cancer Care Ontario, Toronto, Ontario, Canada; 4) Department of Medical Genetics, Mayo Clinic, Rochester, Minnesota, USA; 5) Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA; 6) Familial Cancer Laboratory, Queensland Institute of Medical Research, Brisbane, Australia; 7) Department of Preventive Medicine, University of Southern California, Los Angeles, California, USA; 8) Cancer Research Center, University of Hawaii, Honolulu, Hawaii, USA; 9) Memorial University of Newfoundland, St. John's, Newfoundland, Canada.

Background: Several studies have shown an increased risk of colorectal and extracolonic cancers for carriers of germline *MUTYH* mutations inherited from both parents (biallelic mutations). Extracolonic cancer risks for carriers of a *MUTYH* mutation inherited from only one parent (monoallelic mutation) are yet to be established. **Methods:** We identified 144 families of *MUTYH* mutation carriers from three countries that we ascertained through population-based sources of the multi-site, international Colon Cancer Family Registry. Mutation status, sex, age, and histories of cancer, polypectomy, and hysterectomy were sought from 2,241 of their relatives. Using Cox regression weighted to correct for the method of ascertainment, we estimated the age-specific cumulative risks and country-, age- and sex-specific standardized incidence ratios (SIRs) of colorectal and extracolonic cancers for monoallelic mutation carriers, compared with the general population. **Results:** Monoallelic mutation carriers had a significantly increased incidence of CRC (SIR = 2.28; 95% confidence interval, CI = 1.78 - 2.94; P <0.001), gastric cancer (SIR = 3.16; 95% CI = 2.13 - 4.86; P <0.001), and endometrial cancer (SIR = 2.17; 95% CI = 1.10 - 4.70; P = 0.03) compared to the general population. The estimated cumulative risks to age 70 years based on the population cancer incidence of the United States were as follows: for CRC, 7% (95% CI = 6 - 9%) for men, 5% (95% CI = 4 - 6%) for women; for gastric cancer, 2% (95% CI = 1 - 3%) for men, 0.7% (95% CI = 0.5 - 1%) for women; and for endometrial cancer, 4% (95% CI = 2 - 8%). There was no evidence of increased risks for cancer of the brain, bladder, lung, ovary, breast and prostate. **Conclusion:** Monoallelic *MUTYH* mutation carriers with a family history of CRC are at increased risk of colorectal, gastric and endometrial cancers.

36

High resolution genetic analysis of twenty four loci implicated in colorectal cancer. *V.D. Peltekova[1], M. Lemire[1], Q. Trinh[1], T. Chong[1], D. D'Souza[1], L. Timms[1], L. Hodgson[1], R. De Borja[1], M. Chan-Seng-Yue[1], M. Volar[1], T. Beck[1], J. McPherson[1], T.J. Hudson[1], S. Gallinger[2], B. Zanke[3].* 1) Cancer Genomics, Ontario Institute for Cancer Research, Toronto, Ontario, Canada; 2) Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada; 3) The Ottawa Health Research Institute, Ottawa, Canada.

Despite of the recent contributions of genome-wide association (GWAS) and linkage studies to map the genomic regions associated with colorectal cancer (CRC), identification of the actual disease-contributory genetic variants has been a challenge. To localize causative variants detected by GWAS, and to uncover new variants at loci previously implicated in familial forms of CRC, the Cancer Genome Platform at the Ontario Institute for Cancer Research has employed microarray-based target sequence-capture method, coupled to Illumina Genome Analyzer, to interrogate 3.14 Mb of the 11 GWAS loci that were published for CRC, and additional 13 genes implicated in hereditary forms of CRC. We generated 218 Giga-bases of sequencing data, and identified 13,967 variants in a panel of 40 sporadic CRC cases, 40 controls, and 40 probands and siblings from families with autosomal dominant CRC transmission (AMS I and II) for which no mutations have been previously identified. Out of the 13,967 variants identified, 6,496 (47%) were previously described and 7,471 (53%) were novel. We observed 183 SNPs in coding exons, of which 108 could affect protein structure. Six rare nonsense variants and a splice-site variant were also observed, with three at GWAS loci and four in genes already implicated in hereditary forms of CRC. We next utilized custom high density arrays to genotype 7871 variants from the 11 GWAS loci and 60 coding non-synonymous polymorphisms from the 13 GWAS loci in a panel of 2,400 CRC cases and controls. Fine mapping of the two loci on chromosomes 8q23 and 11q23 has revealed sets of SNPs located at previously uncharacterized genes that may independently contribute to the CRC risk. By rapid amplification of cDNA ends, RT-PCR and whole transcriptome-sequencing of normal and cancerous tissues we were able to identify multiple alternatively spliced transcripts which are specific to the CRC. Taken together, these data enhance our understanding of the CRC-etiology and may identify new therapeutic targets and diagnostic markers.

37

A novel colorectal cancer susceptibility SNP in the EIF3H promoter influences patient survival and response to treatment. *J.P. Cheadle[1], C.G. Smith[1], D. Fisher[2], R. Harris[1], T.S. Maughan[3], S. Idziaszczyk[1], V. Moskvina[4], J. Colley[1], A. Meade[2], C. Bonnet[1], H. West[1], R. Kaplan[2], The COIN Collaborative Group[5], The COIN-B Collaborative Group[6].* 1) Inst Medical Genetics, School of Medicine, Cardiff Univ, Cardiff, S Wales, United Kingdom; 2) MRC Clinical Trials Unit, 222 Euston Road, London, United Kingdom; 3) Department of Oncology and Palliative Care, School of Medicine, Cardiff Univ, Cardiff, S Wales, United Kingdom; 4) Biostatistics and Bioinfomatics Unit, School of Medicine, Cardiff University, Cardiff, S Wales, United Kingdom; 5) The COIN Trial steering group and collaborators; 6) The COIN-B Trial steering group and collaborators.

To-date, genome wide association studies (GWAS) have identified ten loci associated with colorectal cancer (CRC) risk but their role in advanced/metastatic CRC (aCRC) is unknown. We analysed 24 SNPs identified from GWAS or that were highly significant in a subsequent meta-analysis, in 2,186 patients with aCRC and 2,176 healthy controls. We found that rs4939827 at 18q21, rs16892766 at 8q23, rs4779584 at 15q13 and rs10808555 at 8q24 were significantly over-represented in aCRC cases or controls. We tested whether these SNPs influenced patient survival by utilising data from the randomised controlled trial COIN. We found that patients carrying one or more of the minor alleles for rs16892766 (AC or CC genotypes) showed a significant decline in overall survival (OS) as compared to patients with the AA genotype (HR 1.279, 95% CI 1.129-1.449, P<0.001). This effect was strongly significant independent of known prognostic factors. To determine the underlying mechanism at the 8q23 locus, we sequenced EIF3H and identified a SNP, rs28649280, in a potential Sp1 promoter element. rs28649280 was over-represented in aCRC cases as compared to controls (P=0.022) and reduced OS (HR 1.263, 95% CI 1.116-1.430, P<0.001) and response to treatment (OR 0.646, 95% CI 0.481-0.868, P=0.004). Our data represent the first report of a common SNP identified from GWAS for CRC-susceptibility alleles, to have an effect on patient outcome together with insights into the underlying cellular mechanism.

38

A genome-wide association study of survival in ovarian cancer identifies a locus at chromosome 19 that is associated with susceptibility to ovarian cancer. *K.L. Bolton[1,2], J. Tyrer[1], H. Song[1], S.J. Ramus[3], M. Natoridou[3], C. Jones[3], T. Sher[3], A. Gentry-Maharaj[3], E. Wozniak[3], Y. Tsai[4], S. Johnatty[5], P.M. Webb[5], J. Beesley[5], S. Chanock[2], M. Garcia-Closas[2], T. Sellars[4], D.F Easton[1], A. Berchuck[6], G. Chenevix-Trench[5], P.D.P. Pharoah[1], S.A. Gayther[3], The Ovarian Cancer Association Consortium.* 1) Department of Oncology, University of Cambridge, Cambridge, UK; 2) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, USA; 3) 3Department of Gynaecological Oncology, University College London, EGA Institute for Women's Health, London, UK; 4) Moffitt Cancer Center, Tampa, FL, USA; 5) 57The Queensland Institute of Medical Research, Post Office Royal Brisbane Hospital, Australia; 6) 58Division of Gynecologic Oncology, Duke University Medical Center, Durham, North Carolina, USA.

Epithelial ovarian cancer (EOC) is the leading cause of death from gynecological malignancy in the developed world. There is evidence that inherited genetic variation influences EOC survival but no common variants have been definitively identified to date. The aim of our study was to detect polymorphisms that influence EOC prognosis using a genome wide approach. Subjects were from a recently completed, multi-stage, genome-wide association study (GWAS) for genetic susceptibility to EOC. We tested 507,094 SNPs genotyped in 1,768 invasive EOC cases from the UK for association with survival using Cox proportional hazards regression (Stage 1). The 4,649 SNPs showing the strongest association with survival were genotyped in an additional 4,238 EOC cases and 4,809 controls (Stage 2). Based on the combined analysis of Stage 1 and 2 data, no SNP showed association with survival at genome-wide significance (P<10-8). However, rs8170 on 19p13, showed a strong association with risk of serous EOC (P-trend =1x10-7) and a moderate association with survival (P-trend =2x10-5). This SNP and an additional disease-associated variant in the same region, rs2363956, were genotyped in an additional 4,043 EOC cases and 5,951 controls (Stage 3). The combined analysis confirmed the association with risk of serous EOC (combined data odds ratio = 1.16 95% CI 1.11 - 1.21, P-trend = 4 x 10-11). In the phase 3 data there was no evidence for the association of rs8170 or rs2363956 with survival time (combined data hazard ratio =1.09 95% CI 1.04 - 1.14, P-trend = 6 x 10-4). These SNPs are in two genes on 19p13 - MERIT40 and ANKLE1. MERIT40 is a functionally important component of the BRCA1 complex. CGH microarray data on 102 primary EOC tumors showed the 19p13 region to be amplified in 40% of tumors. We also found MERIT40 to show higher expression levels in EOC compared to normal epithelial ovarian cells (p=5.6x10-9), but neither SNP was significantly correlated with MERIT40 expression. We have identified a novel locus on 19p13 that confers risk for serous ovarian cancer. The clear evidence of association with risk suggests that the survival association could still be of interest, but further study will be required to clarify the magnitude of the association. The functional data implicating the BRCA1 interacting gene MERIT40 as the gene underlying the genetic association add weight to the significance of the 19p13 locus for susceptibility in EOC.

39

Cancer Risk in Children with Birth Defects: a Longitudinal, Population-Based Assessment Among 2.7 Million Births. *P. Romitti[1], T. Flood[2], M.L. Feldkamp[3], S. Krikov[3], S. Puzhankara[1], R. Goedken[1], M. Fluchel[3], J. Little[4], L.D. Botto[3].* 1) Dept Epidemiology, Univ Iowa, Iowa City, IA; 2) Arizona Department of Health Services, Phoenix, AZ; 3) Div Medical Genetics, University of Utah, Salt Lake City, UT; 4) Dept Epidemiology, University of Ottawa, Ottawa, Canada.

Assessing cancer risk among children with birth defects has important implications for etiologic research and clinical management. With a few exceptions, the literature is so far inconclusive. To develop this information, we used a population-based approach that combined longitudinal statewide birth cohorts from Arizona, Iowa, and Utah, among over 2.7 million births born from 1983 through 2006. Birth defects were identified through statewide birth defect registries, cancer cases through linked SEER cancer registries. A population-based cohort of over 43,000 children with major birth defects (including trisomies 13, 18, and 21) was compared to an three times larger cohort of nearly 148,000 births without birth defects, randomly sampled from the same underlying birth population and frequency-matched to the birth defects cohort by birth year. We used a Kaplan-Meier time-to-event approach, accounting for censoring by death, to estimate cancer risk through up to age 15 years. Compared to the reference cohort, children with birth defects appeared to have a statistically significant, nearly three-fold increase in risk for cancer (Relative Risk [RR], 2.73). Risk was highest among children with chromosomal conditions, specifically Down syndrome (RR, 13.2), and was driven largely by leukemias. However, the risk for cancer remained moderately increased also among those with a birth defect but without these chromosomal anomalies (RR, 1.82). In this group, cancer risk was driven mainly by brain tumors and some embryonal tumors (neuroblastoma, hepatoblastoma), and occurred mainly in children with brain malformations, cleft palate, rectal malformations, and possibly some heart defects. These population-based findings support and extend previous findings that suggest an increased risk for cancer in children with birth defects, including non-chromosomal birth defects, and suggest selected case groups in which further research could help identify a common genetic susceptibility to cancer and birth defects.

40

Ts65Dn "Down syndrome" mice extend survival of NPcis cancer model. *A. Yang, R.H. Reeves.* Dept of Physiology and Institute for Genetic Medicine, Johns Hopkins Univ., School of Medicine, Baltimore, MD, 21205.

Epidemiological studies report conflicting results as to whether there is a lower incidence of solid tumors in people with trisomy 21 (Down syndrome, DS). We showed that the frequency of adenomas in Apc^{Min} mice is significantly repressed on the trisomic Ts65Dn background, and that tumor number is directly correlated with the dosage of the *Ets2* tumor repressor gene (Sussan et al., Nat 451, 73-75, 2008). To determine whether Ts65Dn protects against multiple kinds of cancer, as suggested by data from DS, we crossed these mice to the NPcis model which has null alleles of the adjacent genes, *NF1* and *Trp53*, and develops sarcoma, lymphoma, adrenocortical carcinoma (ACC) and astrocytoma as a consequence of LOH.

Ts65Dn mice developed the same four tumor types seen in euploid NPcis, however, survival was extended significantly on the trisomic background. This protective effect was complex, but a substantial part of the effect was due to a change in the types of tumors observed. Rapidly growing sarcomas were the most common type in euploid mice, while trisomic mice had a preponderance of ACC which survives longer than sarcoma. The incidence of either lymphoma or astrocytoma was not changed. Trisomic ACC showed elevated apoptosis relative to ACC in euploid mice, which may also contribute to extended survival.

In contrast to intestinal tumors in Apc^{Min}, *Ets2* dosage had no effect on tumor incidence in NPcis in euploid or trisomic mice. It has been reported recently that the microenvironment in Ts65Dn slows growth of xenografts made from long-established, highly aggressive cancer cell lines as a consequence of reduced angiogenesis. Using tumor cell lines newly derived from euploid sarcomas in this study, we also saw reduced growth of xenografts in Ts65Dn, however, no reduction in angiogenesis was detected. Rather, the smaller size was correlated with decreased proliferation. Our biological evidence shows that, as predicted from statistical analysis, trisomy is protective against multiple forms of cancer. The reduced cancer burden in Ts65Dn mice and by extension in people with DS, results from the action of multiple genes affecting multiple growth mechanisms.

41

Towards evidence-based criteria for clinical interpretation of CNVs. *E.B. Kaminsky[1], J. Mulle[1], V. Kaul[1], D. Saul[2], D.L. Pickering[3], D.M. Golden[3], E. Aston[4], T.J. Gliem[5], T. Ackley[6], S. Huang[7], J. Paschall[8], D.M. Church[8], J.C. Barber[7], J.A. Crolla[7], R. Iyer[6], E.C. Thorland[5], S. Shetty[4], S. South[4], A.R. Brothman[4], W. Sanger[3], S. Aradhya[2], S.T. Warren[1], M.K. Rossi[1], M.K. Rudd[1], D.H. Ledbetter[1], C.L. Martin[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; 4) ARUP Laboratories, Salt Lake City, UT; 5) Mayo Clinic, Rochester, MN; 6) Michigan Medical Genetics Laboratories, Ann Arbor, MI; 7) Wessex Regional Genetics Laboratory, Salisbury, UK; 8) National Center for Biotechnology Information, Bethesda, MD.

Copy number variation in the human genome has a profound impact on human health, particularly related to neurodevelopmental disorders. However, large case-control studies of copy number variants (CNVs) are lacking to differentiate between those CNVs that are disease-causing versus those that are found in the normal population. The International Standard Cytogenomic Array (ISCA) consortium was established to leverage the vast amounts of copy number array data generated from clinical genetic testing. Here, we describe CNV analysis of the first ISCA dataset which includes data from 15,749 whole genome arrays. We previously reported that pathogenic CNVs were identified in ~17% of this cohort based on known clinically relevant regions, size/gene content and inheritance pattern. Recurrent CNVs, mediated by segmental duplications, represented ~24% of the imbalances. For these recurrent CNVs, we selected 13 of the most common regions (1q21 TAR, 1q21.1, 3q29, 7q11.23, 8p23.1, 15q11.2-q13, 15q13, 16p11.2, 16p13.11, 17p11.2, 17q12, 17q21.31 and 22q11.2) for formal case-control analyses as an evidence-based strategy to determine clinical significance. We compared the ISCA cases to data from 10,118 normal controls. Each of the 13 microdeletion regions analyzed were significantly overrepresented in cases compared to normal controls (minimum OR>4.7 and p<0.008 for all), confirming each as a pathogenic CNV. The case-control analysis aided in the interpretation of several microduplication regions. For 6 microduplications initially classified as pathogenic, case-control analysis confirmed this classification (minimum OR>3.4; p<0.02). For 16p11.2 duplications, laboratories varied in their initial clinical interpretation, but our analysis demonstrates that this region is most likely pathogenic (OR=6.28; p=2.5E-05). For the 1q21 TAR region, laboratories also had variable interpretations for the duplications; case-control analysis demonstrated that the cases were not significantly different than controls, thus suggesting that duplications of this region may be benign (OR=0.27; p=0.01). The ISCA dataset will be publicly available at NCBI and will serve as a valuable resource for large case-control studies of copy number variation. By using this neurodevelopmental disorders patient dataset together with data from normal control populations, it will be possible to establish a gene dosage map for human development and translate this knowledge for the improvement of patient care.

42

Prevalence of rearrangement mechanisms detected by microarray as multiple genomic imbalances: Results from The International Standard Cytogenomic Array (ISCA) Consortium public database. *N. Uddin[1], R. Toydemir[1], E.B. Kaminsky[2], D. Saul[3], D.L. Pickering[4], D.M. Golden[4], S. Shetty[1], T.J. Gliem[5], T. Ackley[6], S. Huang[7], J.C. Barber[7], J.A. Crolla[7], R. Iyer[7], E.C. Thorland[5], A.R. Brothman[1], W. Sanger[4], S. Aradhya[3], D.H. Ledbetter[2], C.L. Martin[2], S.T. South[1].* 1) Department of Pathology & ARUP Laboratories, University of Utah, Salt Lake City, UT; 2) Emory University, Atlanta, GA; 3) GeneDx, Gaithersburg, MD; 4) University of Nebraska, Omaha, NE; 5) Mayo Clinic, Rochester, MN; 6) University of Michigan, Ann Arbor, MI; 7) Wessex Regional Genetics Laboratory, Salisbury, United Kingdom.

Chromosomal rearrangements, resulting from diverse mechanisms, are important mediators of constitutional genomic imbalances. Frequencies of common unbalanced rearrangements in patients with developmental delay/intellectual disability have been described in the literature based on conventional cytogenetics. Recently, the utility of copy number microarrays in identifying mechanisms underlying genomic imbalances, though limited, has been demonstrated. We analyzed whole-genome microarray data in 15,749 cases, with clinical indications of intellectual disability, autism, and developmental delay; available through the ISCA Consortium, to identify chromosomal rearrangement mechanisms mediating the genomic imbalances. A total of 2688 out of the 15,749 cases were interpreted to have pathogenic imbalances by participating institutes. Out of those, 583 cases had two or more significant genomic imbalances. Based on the position and combination of the imbalances, we were able to postulate the rearrangement mechanism in 208 (35%) of these cases. The overall prevalence of each postulated mechanism from the 2688 abnormal cases was as follows: unbalanced translocations 147 (5.47%), recombination from pericentric inversions 20 (0.74%), ring chromosomes 4 (0.15%) and terminal deletion with adjacent duplication from U-type exchange 26 (0.97%), recombination from paracentric inversions 2 (0.07%), or from non-allelic homologous recombination (NAHR) 9 (0.33%). These frequencies are comparable with the published data, except for terminal deletion with adjacent duplication due to either U-type exchange or NAHR; both of which are more easily distinguished by microarray. U-type exchange was the most common mechanism (70%) for terminal deletion with adjacent duplication, and overall it was the second most common cytogenetic mechanism (after unbalanced translocations) resulting in pathogenic unbalanced genomic changes. In conclusion, in a subset of cases genomic imbalances can be attributed to a defined cytogenetic mechanism based on microarray data. Widespread use of microarray platforms provides an opportunity to examine the overall burden of each of those mechanisms and gives insight into many cases which could not be ascertained by classical cytogenetic techniques alone.

43

Increased detection rate and reliability in pre- and postnatal genetic diagnosis using genome wide 250k SNP array analysis in 4,776 individuals. *N. de Leeuw, R. Pfundt, J. Hehir-Kwa, B. Faas, B. van Bon, N. Leijsten, T. Machielsen, S. van Gessel, M. Wunderink, M. Banning, R. van Beek, H. Yntema, B. de Vries, J. Veltman, D. Smeets.* Department of Human Genetics, Radboud Univ Nijmegen MC, Nijmegen, Netherlands.

Genome wide array-based Comparative Genome Hybridisation (array CGH) was implemented in 2003 in our diagnostic cytogenetic laboratory for the analysis of patients with mental retardation (MR) and / or multiple congenital anomalies (MCA). Since January 2009, genome wide SNP array analysis has replaced routine chromosome studies as the first line diagnostic test for MR / MCA patients in Nijmegen. We also expanded the indications for array analysis and we now include fetal samples, patients with leukemia (not presented here) as well as patients with disorders for which locus heterogeneity is known (homozygosity prescreening). So far, 4,776 samples have been analyzed using the 250k Affymetrix SNP array platform (Affymetrix, Inc., Santa Clara, CA, USA). In about 50% of the patients, array analysis was preceded by routine cytogenetic analysis. The copy number variation (CNV) detection rate in patients was on average 35%: in 20-25% of them, a (potentially) clinically significant CNV was detected by array analysis, including CNVs as small as 8 kb. About a quarter of these imbalances were cytogenetically visible. In 7% of patients, one or more “unique” CNVs were detected that were also observed in one of the parental samples. The clinical significance of these rare inherited CNVs is as yet unclear. Moreover, in an additional 5% of the patients, we detected a significantly increased percentage of homozygosity that in a number of cases led to the identification of recessive disease genes, a mosaic aneuploidy or uniparental disomy (UPD). In 12% of the patients, parental testing was performed by 250k SNP array analysis which enabled us to perform detailed genotype analysis of the patient-parent (trio) information. This analysis is used to screen for the presence of any form of UPD in the patient and/or to determine the parental origin of *de novo* CNVs. This trio analysis is routinely performed when both patient and parents are analyzed by SNP array. Moreover, the outcome of a genotype analysis (which can be performed with any combination of two or three array data sets) is also used as a final quality control by ruling out potential sample mismatches due to non-paternity or sample mix-up. We demonstrate the power of a SNP-based platform for molecular karyotyping and conclude that genome wide SNP array analysis not only significantly improves the diagnostic yield in pre- and postnatal genome diagnostics, but also the quality of the diagnostic laboratory workflow.

44

Affymetrix 2.7Mb high-resolution genome-wide SNP array in clinical practice: preliminary experience. *Y. Qiao[1,2], C. Tyson[3], M. Hrynchak[3], C. Fawcett[3], S. Martell[1], C. Harvard[1], J. Holden[4], S. Lewis[2], E. Rajcan-Separovic[1].* 1) Dept Pathology (Cytogenetics), BC Child and Family Res Inst, Vancouver, BC, Canada; 2) Dept. Medical Genetics , UBC, Vancouver, Canada; 3) Cytogenetics Lab, Royal Columbian Hospital, New Westminster, Canada; 4) Depts Psychiatry & Physiology, Queen’s University, Kingston, Ontario, Canada.

Pathogenic copy number variants (CNVs) are found in 5-15% of individuals with intellectual disability (ID) using different array CGH platforms. It is expected that arrays with higher resolution genomic coverage detect CNVs more accurately and allow identification of smaller CNVs. We applied Affymetrix Cytogenetics Whole-Genome 2.7M Array to assess CNV detection in 22 ID cases that contain 29 pathogenic CNVs (ranging from 45Kb-5.4Mb) previously detected and confirmed by other array platforms and/or FISH or Q-PCR. All of the 29 positive CNVs were detected with 85% confidence level when CNV size cut-off was set to 100Kb. However, the detection rate dropped to 76% of positive CNVs (22/29) when the confidence level was increased to 90%. When no CNV size or confidence level restrictions were applied (confidence 0, segment size 0 Kb), 428 CNVs were detected (18.6 in average/subject, range 6-50). The CNVs on chromosome X/Y were not counted due to the noisy data on these chromosomes. 31% of CNVs were >100Kb, 55% were 10~100Kb while 14% CNVs <10Kb. More than half of the CNVs within each size group were benign (i.e. common variants reported in at least two studies in the Database of Genomic Variants -DGV, or had no gene content). The proportion of benign CNVs was highest for the <10Kb CNV size group (73%). In addition to the known positive CNVs we identified 20 unique CNVs that were not detected previously when a size cut-off of 100kb and confidence of 85% were applied. They all contain gene(s) and do not overlap with common CNVs from DGV, therefore they could have clinical relevance. These CNVs are being followed up by FISH. In light of the recent observation that 9-24% of cases with specific recurrent microdeletions/duplications have a second putatively pathogenic CNV (“two-hit” model Girirajan et al., Nature Genetics 2010), and the validation of the 6.0 Affy array which showed additional pathogenic findings in 3/19 cases with one known pathogenic CNVs detected by lower resolution array (Bernandini et al., EJHG, 2010), we anticipate that high resolution Affy 2.7M Array could reveal further clinically relevant CNVs in patients with previously identified aberrations.

45

Whole exome sequencing as a tool to identify both sequence variants and CNVs. *A. Bale, P. Gordon, M. Choi, S. Mane, R. Lifton.* Dept Gen, Yale Univ Sch Med, New Haven, CT.

Whole-exome sequencing (WES), based on array capture followed by high-throughput analysis, is a method of global screening for single base changes and small indels in expressed regions of the genome. Although not designed for detection of large deletions, theoretically exome sequences contain within them data about gene dosage based on representation; e.g., compared with the diploid state, a heterozygous deletion results in one half the normal sequencing “depth” (analogous to RNA seq. for gene expression). In practice, the representation of different exons varies by an order of magnitude or more because of differences in array capture efficiency, but representation of any particular exon is highly reproducible from one assay to the next. We used SeqCap arrays and two lanes on an Illumina GAII for an average exome coverage of 70X. By comparison to SNP arrays, the sensitivity and specificity for detecting single base changes were both 99% (Choi et al., 2009). A python script was developed to evaluate gene dosage systematically at the level of individual exons or larger windows ranging up to 1Mb. To assess sensitivity and false discovery rate, we analyzed X chromosome representation in 3 females and 5 males. At the level of individual exons, 2N dosage (females) could be distinguished consistently from 1N dosage (males) for any exon with a coverage of 50X or greater but was not reliable for exons with lower representation. On the other hand, sampling of all regions of the X chromosome with window sizes ranging from 100kb to 1 Mb gave almost exactly 2N dosage in females compared with males and yielded no false discoveries when exons with homologs elsewhere in the genome were masked. Blind analysis of the female subjects correctly discovered 1N gene dosage for all heterozygous X chromosome CNVs larger than 100 kb. In addition to its use for identifying point mutations, WES presently has approximately the same resolution for CNVs as 44k array CGH platforms. In its current implementation it cannot serve as a comprehensive tool for discovering single-exon deletions, but minor improvements in array capture would increase the resolution to the single-exon level. While this method under-represents gene-poor regions, it is likely that most clinically relevant structural variants exert their effects through altered dosage of expressed sequences. As it transitions to clinical laboratories, WES will provide a comprehensive platform for nearly all genetic testing.

46

Long-range expression consequences of copy number variation: insights from Smith-Magenis and Potocki-Lupski syndrome mouse models. *G. Ricard[1], J. Molina[2], J. Chrast[1], W. Gu[3], N. Gheldof[1], S. Pradervand[1,4], F. Schütz[1,4], J.I. Young[2,5], J.R. Lupski[3], A. Reymond[1], K. Walz[2,5].* 1) Genopode, Ctr Integrative Genomic, Lausanne, Switzerland; 2) Centro de Estudios Científicos (CECS), Valdivia, Chile; 3) Molecular & Human Genetics, Baylor College of Medicine, Houston, Texas, USA; 4) Swiss Institute of Bioinformatics (SIB), 1015 Lausanne, Switzerland; 5) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Florida.

A large fraction of genome variation between individuals is comprised of submicroscopic copy number variation of genomic DNA segments. We assessed the relative contribution of structural changes and gene dosage alterations on phenotypic outcomes with mouse models of Smith-Magenis and Potocki-Lupski syndromes. We phenotyped mice with 1n (Deletion/+), 2n (+/+), 3n (Duplication/+) and balanced 2n compound heterozygous (Deletion/Duplication) copies of the same region. Parallel to the observations made in humans, such variation in gene copy number was sufficient to generate phenotypic consequences: in a number of cases diametrically opposing phenotypes were associated with gain versus loss of gene content. Surprisingly, some neurobehavioral traits were not rescued by restoration of the normal gene copy number. Transcriptome profiling showed that a highly significant propensity of transcriptional changes map to the engineered interval in the five assessed tissues. A statistically significant overrepresentation of the genes mapping to the entire length of the engineered chromosome was also found in the top-ranked differentially expressed genes in the mice containing rearranged chromosomes, regardless of the nature of the rearrangement; an observation robust across different cell lineages of the central nervous system. Our data indicate that a structural change at a given position of the human genome may affect not only locus and adjacent gene expression, but also “genome regulation”. Furthermore, structural change can cause the same perturbation in particular pathways regardless of gene dosage. Thus, the presence of a genomic structural change, as well as gene dosage imbalance, contributes to the ultimate phenotype.

47

Phenotypic, Genomic and Functional Characteristics of the 1q21.1 Copy Number Variant (CNV) in Multigenerational families. *C. Harvard[1], E. Strong[1], P. Pavlidis[2], M. O'Driscoll[3], E. Chow[4,5], S. Martell[1], C. Tyson[6], B. McGillivary[7], S. Hamilton[7], S. Marles[8], A. Mhanni[8], Y. Qiao[1], J.J.A. Holden[9], M.E.S. Lewis[7], E. Rajcan-Separovic[1], *first and second author contributed equally to this work.* 1) Department of Pathology, Child & Family Research Institute, Vancouver, BC, Canada; 2) Centre for High-throughput Biology Michael Smith Laboratories, UBC, Vancouver, BC, Canada; 3) Human DNA Damage Response Disorders Group, University of Sussex, Brighton, UK; 4) Clinical Genetics Service, Centre for Addiction and Mental Health, Toronto , ON, Canada; 5) Department of Psychiatry, University of Toronto, Toronto, ON, Canada; 6) Cytogenetics Laboratory, Royal Columbian Hospital, New Westminster, BC, Canada; 7) Department of Medical Genetics, Vancouver, BC, Canada; 8) Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, MN, Canada; 9) Departments of Physiology and Psychiatry, Queen's University, Kingston, ON, Canada.*

Copy number variations (CNVs) at chromosomal position 1q21.1 are associated with variable congenital anomalies and neurodevelopmental phenotype (ND) including intellectual disability (ID). This CNV is also found in unaffected individuals, albeit at lower frequencies. We have characterized 4 families carrying the 1q21.1 CNV in both affected and unaffected members using a variety of approaches, including; a) high resolution (2.7 Mb Affymetrix) array; b) comparison of phenotypes of probands and unaffected carriers; c) comparison of whole genome (WG) RNA expression profiles for 1q21.1 CNV carriers, and; d) studies of function of genes integral to the 1q21.1 CNV which may contribute to phenotypic variability.Our study includes 9 subjects with 1q21.1 deletion (7) or duplication (2). Five subjects present with ID and congenital anomalies; two subjects had learning difficulties or ADHD associated with milder congenital anomalies and 2 subjects are phenotypically normal relatives of subjects with ID and congenital anomalies. Genomic breakpoints mapped to the previously reported 1.35 Mb critical region (144.5-146.3 Mb). The presence of an additional CNV hit was excluded in all subjects. Phenotypic analysis revealed ND phenotype and short stature as the most common finding in all affected subjects. As previously reported, microcephaly was associated with deletion in 3/5 affected subjects and macrocephaly with duplication (1/1 affected subject). No other consistent phenotypes were found. One family showed a progressively more severe neurodevelopmental phenotype through 3 generations. We prioritized genes based on expression changes associated with copy number. The top hit was a gene in 1q21.1 (CHD1L, p = 2.4x10e-5, though not significant after multiple test correction), and overall enrichment for genes in 1.q21.1 was very significant in the top 100 genes (p<10e-14). The changes in RNA expression of CHD1L were also confirmed by increased and decreased CHD1L protein levels in patient cell lines with a 1q21.1 dup and del respectively. This gene is particularly interesting since both up- and down-regulated expression has been reported to increase sensitivity to DNA damaging agents. CHD1L copy number and expression changes may represent a susceptibility factor in the presence or absence of DNA damaging agents, rendering variable phenotypic consequences in 1q21.1 CNV carriers. DNA damage response in cells from carriers of 1q21.1 CNV is currently investigated.

48

Recurrence and submicroscopic complexity in 71 probands with unbalanced insertions characterized by array CGH and FISH. *B.C. Ballif[1], N.J. Neill[1], A. Lamb[1], S. Parikh[2], J.B. Ravnin[1], R. Schultz[1], B. Torchia[1], J. Rosenfeld[1], L.G. Shaffer[1].* 1) Signature Genomic Laboratories, Spokane, WA; 2) Center for Pediatric Neurology, Cleveland Clinic, Cleveland, OH.

Insertions occur when a segment of one chromosome is translocated and inserted into a new region of the same chromosome, the other homolog, or a non-homologous chromosome. The incidence of unbalanced insertions is not well established, with estimates ranging from 1:500 to 1:80,000 live births. We report 71 insertions, representing 0.83% of 8,530 abnormalities detected in samples referred to our laboratory for array-based comparative genomic hybridization (array CGH). All rearrangements were visualized by FISH after microarray analysis detected a gain or loss of a chromosomal segment. Among our 71 probands, three recurrent, paternally inherited der(Y)ins(Y;18)(q?11.2;p11.32p11.32) were observed. In addition, an intrachromosomal, maternally inherited der(X)ins(X)(q28p22.33p22.33) was identified in two unrelated male probands. The clinical significance of these recurrent rearrangements is unclear, although the inheritance pattern of each suggests that the phenotypic consequences of these insertions may be benign, or in the case of the der(X)ins(X)(q28p22.33p22.33), may only be significant when present in male progeny of a carrier female. In addition, cryptic, submicroscopic duplications were observed near the insertion sites in a patient with a *de novo* der(1)ins(1;X)(p36.32;q22.2q22.2) and a patient with a *de novo* der(21)ins(21;X)(q?22;q28q28), highlighting the power of high-resolution array CGH analysis compared to traditional karyotyping techniques in the characterization of complex rearrangements and emphasizing the mechanistic complexity underlying the formation of insertions. Finally, we used linear amplification with primers specific to the inserted genomic segment coupled with array CGH and DNA sequencing to identify the precise insertion site in a der(X)ins(X;19)(q28;p13.3p13.3). This approach identified an insertion of a 126 kb segment of chromosome 19 into *MECP2* in a patient with symptoms of Rett syndrome. Our results confirm that copy number gains identified by arrays may represent unbalanced insertions that can be pathogenic due to the disruption of a critical gene or through increased dosage due to duplication of a critical gene. Furthermore, our data demonstrate the necessity of FISH visualization of abnormal microarray results to determine the origin and nature of DNA copy-number gain/loss. This information is essential for estimation of recurrence risks and effective genetic counseling.

49

High incidence of genomic disorder-associated microdeletions in isolated and syndromic Mullerian aplasia. *S. Nik-Zainal[1], R. Strick[2], M. Storer[1,3], N. Huang[1], R. Rad[1], L. Willatt[4], T. Fitzgerald[1], V. Martin[1,3], R. Sandford[4], N.P. Carter[1], A.R. Janecke[5], S.P. Renner[2], P.G. Oppelt[2], P. Oppelt[2], C. Schulze[2], M. Hurles[1], M.W. Beckmann[2], P.L. Strissel[2], C. Shaw-Smith[1,3].* 1) Wellcome Trust Sanger Institute, Cambridge, UK; 2) University-Clinic Erlangen, Erlangen, Germany; 3) Institute of Child Health, London, WC1N 1EH UK; 4) Addenbrooke's Hospital, Cambridge CB2 2QQ UK; 5) Innsbruck Medical University, A-6020 Innsbruck, Austria.

Complete aplasia of the uterus, cervix and upper vagina, also termed Mullerian aplasia or Mayer-Rokitansky-Kuster-Hauser (MRKH) syndrome occurs with an incidence of around 1 in 4,500 female births and results in failure of menstruation and infertility. It may occur in isolated or syndromic form with associated fallopian, ovarian, urinary tract or spine abnormalities. Previous reports in the literature suggest that a proportion of such cases, in particular syndromic cases, are caused by variation in copy number at a variety of genomic loci. In order to obtain an overview of the contribution of copy number variation to both isolated and syndromic Mullerian aplasia, we studied a series of 63 cases, 25 with syndromic and 38 with apparently isolated forms of the condition. We report a strikingly high incidence of microdeletions associated with previously characterized genomic disorders in this cohort: of the 63 samples studied, 9 (14 %) had a copy number variant of this type. These comprised four cases of microdeletion at 16p11.2, an autism susceptibility locus not previously associated with Mullerian aplasia; four cases of microdeletion at 17q12; and one case of a distal 22q11.2 microdeletion. Microdeletions at 16p11.2 and 17q12 were found in four of 38 (10.5%) of cases with isolated Mullerian aplasia, and at 16p11.2, 17q12 and 22q11.2 (distal) in five of 25 cases (20%) with syndromic Mullerian aplasia. Our finding of microdeletion at 16p11.2 in 2/38 (5%) of isolated and 2/25 (8%) of syndromic cases suggests that this genomic disorder alone makes a significant contribution to the pathogenesis of Mullerian aplasia. Overall, the high incidence of genomic disorder-associated copy number variants in all forms of Mullerian aplasia has implications for our understanding of the aetiopathogenesis of the condition, and for genetic counseling in families affected by it.

50

Large inverted repeats within Xp11.2 are present at the breakpoints of isodicentric X chromosomes in Turner syndrome. *S.A. Scott, N. Cohen, T. Brandt, P.E. Warburton, L. Edelmann.* Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY.

Turner syndrome (TS) results from whole or partial monosomy X and is mediated by haploinsufficiency of genes that normally escape X-inactivation. Although a 45,X karyotype is observed in half of all TS cases, the most frequent variant TS karyotype includes the isodicentric X chromosome alone [46,X,idic(X)(p11)] or as a mosaic [46,X,idic(X)(p11)/45,X]. Given that the mechanism of idic(X)(p11) rearrangements is poorly understood and breakpoint sequence information is unknown, this study sought to investigate the molecular mechanism of idic(X)(p11) formation by determining their precise breakpoint intervals. Karyotype analysis and fluorescence in situ hybridization mapping of eight idic(X)(p11) cell lines and three unbalanced Xp11.2 translocation lines identified the majority of breakpoints within a 5 Mb region from ~53 to 58 Mb in Xp11.1-11.22, clustering into four regions. To further refine the breakpoint intervals, a high resolution (average of ~350 bp) oligonucleotide microarray was designed and array-based comparative genomic hybridization (aCGH) was performed on all eleven idic(X)(p11) and Xp11.2 translocation lines. aCGH analyses identified all breakpoint regions, including an idic(X)(p11) line with two potential breakpoints, one breakpoint shared between two idic(X)(p11) lines, and two Xp translocations that shared breakpoints with idic(X)(p11) lines. Four of the ten breakpoint regions included large inverted repeats composed of repetitive gene clusters and segmental duplications, which corresponded to regions of copy-number variation. In addition, four breakpoints harbored direct and/or inverted repetitive LINE elements and one breakpoint occurred in the pericentromeric region. These data indicate that the rearrangement sites on Xp11.2 which lead to isodicentric chromosome formation and translocations are likely not random and suggest that the complex repetitive architecture of this region predisposes it to rearrangements, some of which are recurrent.

51

Autosomal Recessive Fibrochondrogenesis Caused by *COL11A1* Mutations. *S.W. Tompson[1], C.A. Bacino[2], N.P. Safina[3], M.B. Bober[4], T. Funari[1], L. Nevarez[1], D. Krakow[1,5,6], D.H. Cohn[1,5,7].* 1) Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, CA; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Children's Hospital of the King's Daughters, Norfolk, VA; 4) Nemours/Alfred I duPont Hospital for Children, Wilmington, DE; 5) Department of Human Genetics, University of California, Los Angeles, CA; 6) Department of Orthopedic Surgery, University of California, Los Angeles, CA; 7) Department of Pediatrics, University of California, Los Angeles, CA.

Fibrochondrogenesis is a distinct, autosomal-recessive, neonatally lethal, short-limbed skeletal dysplasia first described in 1978. The disease is clinically characterized by protuberant eyes, a flat midface with a small nose and anteverted nares, shortening of all limb segments with relatively normal hands and feet, and a small bell-shaped thorax. Radiographically, individuals display short long-bones with broad metaphyses, the vertebral bodies are flat and, on lateral view, have a distinctive appearance with a hypoplastic posterior end and a rounded anterior end. The ribs are typically short, wide and cupped at both ends. In a single case of fibrochondrogenesis, whole-genome SNP genotyping identified unknown ancestral consanguinity by detecting three autozygous regions. Because of the cartilage selective nature of the phenotype, the 389 genes in the autozygous intervals were prioritized by correlating their expression with known cartilage-selective genes using the UCLA Gene Expression Tool (UGET). The gene encoding the α1 chain of type XI collagen (*COL11A1*) was the most significantly correlated gene and the only cartilage-selective gene in the three candidate intervals. Sequence analysis of *COL11A1* in two genetically independent fibrochondrogenesis cases demonstrated that each was a compound heterozygote for a loss-of-function mutation on one allele and a substitution for a triple helical glycine residue on the other. Electron microscopy on human cartilage tissue identified thickened collagen fibers in the extracellular matrix and protein retention in the rough endoplasmic reticulum of chondrocytes. The human fibrochondrogenesis phenotype is similar to that of the chondrodysplasia (*cho*/*cho*) mouse, a recessive phenotype resulting from homozygosity for a *Col11a1* loss of function allele.

52

An inherited hand arthropathy is caused by mutations in *TRPV4*. *R. Savarirayan[1], S. Lamande[1], I. Gresshoff[1], Y. Yuan[2], P. McIntrye[2], L. Rowley[1], D. Amor[1], K. Kaluarachchi[1], W. Cole[3], C. Little[4], E. Botzenhart[5], J. Bateman[1].* 1) Murdoch Children's Res Inst, Melbourne, Australia; 2) Dept. Pharmacology, University of Melbourne; 3) University of Alberta; 4) Institute of Bone and Joint Research, Sydney; 5) University of Freiburg.

Familial Digital Arthropathy-Brachydactyly (FDAB;MIM 606835) is a dominantly inherited condition involving aggressive osteoarthritis of the fingers and toes and consequent shortening of the middle and distal phalanges, with no other evidence of skeletal abnormality. Here, we show in two unrelated families and a sporadic case that the disorder is caused by missense substitutions (G270V, R271P and F273L) in the intracellular ankyrin repeat domain of the transient receptor potential cation channel TRPV4. Functional testing of mutant TRPV4 in stably transfected HEK293 cells revealed the mutant proteins showed poor cell surface localization, despite being expressed at similar levels to wild type protein, and calcium influx in response to the TRPV4 agonist GSK1016790A was significantly reduced. Others have shown that gain of function TRPV4 mutations cause skeletal dysplasias and peripheral neuropathies. Our data showing tightly clustered TRPV4 mutations that reduce channel activity in a third phenotype, inherited osteoarthritis, demonstrate the importance of TRPV4 activity in articular cartilage homeostasis and raises the possibility that this cation channel may play a role in age-related osteoarthritis.

53

Loss-of-function mutations of *CHST14* cause a new type of autosomal recessive Ehlers-Danlos syndrome. *N. Miyake[1], T. Kosho[2], S. Mizumoto[3], T. Furuichi[4], A. Hatamochi[5], S. Ikegawa[4], S. Yamada[3], K. Sugawara[3], N. matsumoto[1].* 1) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 2) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Japan; 3) Laboratory of Proteoglycan Signaling and Therapeutics, Hokkaido University Graduate School of Life Science, Sapporo, Japan; 4) Laboratory for Bone and Joint Disease, Center for Genomic Medicine, RIKEN, Tokyo, Japan; 5) Department of Dermatology, Dokkyo Medical University, School of Medicine, Tochigi, Japan.

Ehlers-Danlos syndrome (EDS) is a heterogeneous connective tissue disorder involving skin and joint laxity and tissue fragility affecting as many as 1 in 5,000 individuals. In 2005, we reported two patients of new type of EDS, similar to kyphoscoliosis type but without lysyl hydroxylase deficiency. The patients also present with the distinct craniofacial characteristics, multiple congenital contractures, progressive joint and skin laxity, and multisystem fragility-related manifestations. Through long-term clinical evaluation of them as well as additional four unrelated patients, we confirmed that these patients represent a new type of EDS. As two of six patients are consanguineous families, we hypothesized this is autosomal recessive disorder. We mapped its locus at 15q15.1 by homozygosity mapping using two consanguineous families and identified a homozygous *CHST14* mutation in the two familial cases and compound heterozygous mutations in four sporadic cases. *CHST14* encodes dermatan 4-*O*-sulfotransferase 1 (D4ST1), which transfers active sulfate from 3'-phosphoadenosine 5'-phosphosulfate to position 4 of the N-acetyl-D-galactosamine (GalNAc) residues of dermatan sulfate (DS). Transfection experiments of mutants and enzyme assays using fibroblast lysates of patients showed the loss of D4ST1 activity. *CHST14* mutations altered the glycosaminoglycan (GAG) components in patients' fibroblasts. Interestingly, DS of decorin proteoglycan, a key regulator of collagen fibril assembly, was completely lost and replaced by chondroitin sulfate (CS) in the patients' fibroblasts. The structure of DS chains is flexible while that of CS chains is rigid. Thus, the transition from the CS/DS hybrid chain of decorin to a CS chain probably decreases the flexibility of the GAG chain and breaks the GAG antiparallel complex after compression stresses. The histopathological examination revealed that the collagen bundle formation was impaired while the sizes and shapes of collagen fibrils are unchanged in the patients' dermal tissues. This might indicate the loss of the decorin DS proteoglycan due to *CHST14* mutations may preclude proper collagen bundle formation or maintenance of collagen bundles. In conclusion, we could confirm that *CHST14* mutations cause a new type of EDS and the impaired bundle formation is suggested as a main pathomechanism in this disease.

54

Genomic analysis in a single affected individual reveals *BMPER* mutation in diaphanospondylodysostosis. *V.A. Funari[1,2], D. Krakow[1,3,4,5], L. Nevarez[1], Z. Chen[4], T. Funari[1], W.R. Wilcox[1,2], D.L. Rimoin[1,2,4,6], S.F. Nelson[2,4], D.H. Cohn[1,2,4].* 1) Medical Genetics Inst, Cedars-Sinai Med Ctr, Los Angeles, CA; 2) Pediatrics, David Geffen School of Medicine at UCLA, Los Angeles, CA; 3) Orthopedic Surgery, David Geffen School of Medicine at UCLA, Los Angeles, CA; 4) Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA; 5) Obstetrics and Gynecology, David Geffen School of Medicine at UCLA, Los Angeles, CA; 6) Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA.

Diaphanospondylodysostosis (DSD) is a rare, recessively-inherited, perinatal lethal skeletal disorder. The low frequency and perinatal lethality of DSD makes assembling a large set of families for traditional linkage based genetic approaches challenging. New genomic technologies make alternative approaches feasible and efficient for single gene disorders. Using 250K Affymetrix SNP microarrays, we empirically searched for unknown ancestral consanguinity in a cohort of four unrelated patients and identified autozygosity in a single patient of Mexican descent. We identified two autozygous intervals comprising 34 Mbps, unique to the single case of DSD. A custom CGH array and protocol were designed to capture the exons in these intervals and the captured fragments were subjected to Illumina 75 bp paired-end sequencing. Exon sequence analysis in these intervals revealed that the affected individual was homozygous for a null mutation in the BMPER gene, which encodes the bone morphogenetic protein-binding endothelial cell precursor-derived regulator. Mutations in BMPER were subsequently found in three additional DSD cases, confirming that defects in BMPER cause DSD. Bmper knockout mice accurately phenocopy most of the skeletal and kidney defects observed in humans with DSD. This study and several other cases of unknown consanguinity in recessive disorders in individuals of Mexican origin, as well as a recent study finding high rates of homozygosity in a random set of individuals from Guadalajara, Mexico, suggest that the Mexican population represents a significant resource for genetic mapping of loci for recessive disorders. Empirical identification of autozygosity due to unknown ancestral consanguinity in autosomal recessive diseases, particularly in certain populations such as Mexicans, can significantly refine candidate disease gene intervals and thus the genomic regions for targeted resequencing using a next generation sequencing platform. Finally, these data indicate that BMPER-related BMP signaling plays an essential role in vertebral segmentation early in human development.

55

A novel skeletal disorder, Bent Bone Dysplasia-FGFR2 Type, has deficient canonical FGFR2 signaling and implicates a regulatory role for FGFR2 in Runx2 activity. *A. Merrill[1], A. Sarukhanov[1], P. Krejci[5], N. Camacho[1], K. Estrada[3], K. Lyons[1,3], W. Wilcox[4,5], D. Krakow[1,2].* 1) Department of Orthopaedic Surgery, David Geffen School of Medicine at UCLA, Los Angeles, CA; 2) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA; 3) Department of Molecular Cellular and Developmental Biology, UCLA, Los Angeles, CA; 4) Department of Pediatrics, David Geffen School of Medicine at UCLA, Los Angeles, CA; 5) Medical Genetics Research Institue, Cedars Sinai Medical Center, Los Angeles, CA.

We have characterized a new autosomal dominant lethal skeletal dysplasia. The findings include bent long bones, hypoplastic scapula and clavicles, deficient skull ossification, coronal craniosynostosis, prenatal teeth, and thickened periosteum of the long bones. In three unrelated affected individuals, heterozygosity for an identical de novo missense mutation was identified that introduces a polar amino acid in the transmembrane (TM) domain of FGFR2. Thus, we have termed this novel disorder Bent Bone Dysplasia FGFR2 type (BBD). Heterozygosity for mutations in FGFR2 cause a heterogenous group of skeletal disorders including Aperts, Crouzon, and Pfeiffer syndromes, all of which are characterized by craniosynostosis and result from receptor activating mutations. The BBD phenotype is extremely different from the aforementioned disorders. The mutation is predicted to disrupt the membrane anchor motif (TM) by the TMHMM algorithm. Others have shown that the FGFR2 TM domain has an inherently weak association to cellular membranes and introduction of a polar amino acid in a homologous region of FGFR1 decreases localization of the receptor to the plasma membrane. In support of these findings, BBD cells have little identifiable FGFR2 localized to the plasma membrane compared to control cells. Further, affected cells have reduced levels of the maturely glycosylated form of FGFR2 normally targeted to the plasma membrane. This suggests that this disorder results from loss of function in FGFR2 signaling at the plasma membrane. A cell based assay in BaF3 cells to test the direct signaling effect of the mutation in the absence of endogenous FGF signaling showed that the mutant receptor is unable to activate the MAPK/ERK pathway upon stimulation with FGF2, FGF9, and FGF18. Reduced skull ossification and hypoplastic clavicles in BBD have phenotypic overlap with Cleidocranial Dysplasia, suggesting a defect in osteoblast differentiation and a role for Runx2 in the disease process. Further, FGF-activated ERK has been shown to affect Runx2 activity and defective intranuclear targeting of Runx2 results in decreased activity and osteogenic differentiation. BBD cells show defective subnuclear targeting of Runx2 compared to control implicating a loss of Runx2 activity. The delineation of this new disorder suggests canonical FGF signaling regulates Runx2 activity and intranuclear localization through FGFR2 activation of the MAPK/ERK pathway.

56

Mutations in *PTPN11* cause Ollier disease and Maffucci syndrome. *N.L.M. Sobreira[1], E.T. Cirulli[2], A.B.A. Perez[3], R. Jesus-Garcia[4], E. Wohler[5], F. Lissa[6], J. Costa[7], S. Nikkel[8], D. Avramopoulos[1,9], G. Thomas[1,5], D.B. Goldstein[2], J.E. Hoover-Fong[1,10,11], D. Valle[1,10].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Center for Human Genome Variation, Duke Institute for Genome Sciences and Policy, Duke University, Durham, NC; 3) Universidade Federal de Sao Paulo, Centro de Genetica Medica, SP, Brazil; 4) Escola Paulista de Medicina, Universidade Federal de Sao Paulo, SP, Brazil; 5) Kennedy Krieger Institute of Cytogenetics, Baltimore, MD; 6) Department of Surgical Oncology, Hospital São José, Criciúma, Brazil; 7) Serviço de Reumatologia do CHAM, EPE - Ponte de Lima; 8) Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario; 9) Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, MD; 10) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD; 11) Greenberg Center for Skeletal Dysplasias, Baltimore, MD.

Metachondromatosis (MC, MIM156250) is an AD disorder characterized by multiple exostoses and enchondromas without a risk for malignant transformation. We utilized whole genome sequencing coupled with weak linkage information to identify loss of function mutations in *PTPN11* as the cause of MC (Sobreira et al. in press). *PTPN11* encodes the protein tyrosine phosphatase, SHP-2, which plays a central role in RAS/MAPK signaling downstream of several receptor tyrosine kinases. Germline gain of function missense mutations in *PTPN11* cause Noonan syndrome (MIM163950), Noonan-like disorders with multiple giant cell lesion syndrome (MIM163950), and LEOPARD syndrome (MIM151100). Our MC results suggested that *PTPN11* mutations might also be involved in two additional unexplained phenotypes with enchondromas: Ollier disease (OD, MIM166000), characterized by multiple enchondromas, typically with unilateral predominance and onset in early childhood, resulting in skeletal deformities; and, Maffucci syndrome (MS, MIM166000) characterized by multiple enchondromas, hemangiomas and phlebectasia typically with onset in puberty. There are a few instances of familial OD, while MS is sporadic. Schwartz et al. (1987), using a life-table method, calculated that 25% of OD and 100% of MS patients eventually develop a related malignancy. We sequenced *PTPN11* exons and flanking splice sites in constitutional DNA of one patient with OD with multiple deforming enchondromas of the upper and lower extremities (Costa et al., 2008); and, in tumor DNA extracted from a paraffin block of a retiform hemangioendothelioma of a patient with MS who also had enchondromas of her ribs and upper extremities (Lissa et al, 2009). In the OD patient we identified a heterozygous missense mutation in exon 14 of *PTPN11* (L560F); and in the MS sample we identified a heterozygous nonsense mutation in exon 4 of *PTPN11* (R138X). L560F has not been observed in 200 controls chromosomes in which *PTPN11* was sequenced (Sarkozy et al., 2003) and L560 is conserved to zebrafish. Our results implicate *PTPN11* mutations in the pathogenesis of OD and MS. Studies in progress will determine the interaction of inherited versus acquired *PTPN11* mutations as well as the effect of these mutations on SHP-2 function and on the involved RAS/MAPK pathways. Our results, together with those of others, indicate that *PTPN11* mutations contribute to at least 6 discrete phenotypes.

57

The skeletal disorder metachondromatosis is caused by loss-of-function mutations in *PTPN11*. *E.D. Boyden[1], M. Bowen[1], I.A. Holm[1], B. Campos-Xavier[2], L. Bonafé[2], S. Ikegawa[3], V. Cormier-Daire[4], R. Savarirayan[5], S. de Sousa[6], J.R. Kasser[1], M.L. Warman[1], K.C. Kurek[1].* 1) Children's Hospital Boston/Harvard Medical School, Boston, MA; 2) Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland; 3) Laboratory for Bone & Joint Diseases, RIKEN, Tokyo, Japan; 4) Groupe Hospitalier Necker-Enfants Malades, Paris, France; 5) Royal Children's Hospital, Melbourne, Australia; 6) Hospital Pediátrico de Coimbra, Coimbra, Portugal.

Metachondromatosis is an autosomal dominant, incompletely penetrant disorder for which affected individuals develop skeletal growth plate enchondromas and osteochondromas. The disease is unlinked to EXT1 and EXT2, the genes responsible for the phenotypically similar multiple hereditary exostoses (MHE). In contrast to MHE, the osteochondromas in metachondromatosis tend to point toward the joint, affect digits more frequently than long bones, and resolve spontaneously. To identify the gene responsible for metachondromatosis, we genotyped affected individuals from a single family segregating the disorder on Affymetrix 6.0 SNP arrays. The maximum attainable LOD score of 2.7 was observed for an 8.6 Mb interval on chromosome 12, and the high marker density excluded the remainder of the genome. Genotyping in two smaller families was also compatible with linkage to chromosome 12. We then performed multiplexed targeted capture, using an Agilent 1M oligonucleotide array, of all exons and promoters within the interval on DNAs from 16 affected individuals among 11 unrelated families. We sequenced the libraries on the Illumina Genome Analyzer II and identified variants using Novoalign and SAMtools. We observed heterozygous mutations in *PTPN11* in 4 of the 11 families, including 1 nonsense and 3 frameshift mutations. Sanger sequencing of *PTPN11* in the remaining 7 families and 2 additional participants identified 1 additional nonsense and 1 additional frameshift mutation. In sum, putative loss-of-function mutations in *PTPN11* were found in 6 of 13 unrelated families with metachondromatosis. *PTPN11* encodes SHP2, a member of a family of protein tyrosine phosphatases known to be regulatory signaling molecules. Two other autosomal dominant disorders, Noonan and LEOPARD syndromes, are caused by heterozygous missense mutations in *PTPN11* postulated to be neomorphic, antimorphic, or partial loss-of-function. We are currently testing our hypotheses that the *PTPN11* mutations found among individuals with metachondromatosis are functionally null, and that second somatic null mutations causing complete loss of SHP2 expression are required for the development of the skeletal lesions. If correct, then by applying Knudson's hypothesis relating differences in the occurrence of tumors in isolated and familial cancer syndromes to isolated and familial enchondroma syndromes, we speculate that two somatic *PTPN11* mutations may be a frequent cause of isolated enchondromas.

58

Spondylo-megaepiphyseal-metaphyseal dysplasia: a rare autosomal recessive chondrodysplasia due to inactivating mutations in the NKX3-2 gene. *G. Mortier[1], M. Simon[2], A. Dheedene[3], Y. Alanay[4], E. Mihci[5], L. Rifai[6], A. Sefiani[6], Y. van Bever[2], M. Meradji[7], A. Superti-Furga[8], J. Hellemans[3].* 1) Dept Medical Genetics, Antwerp Univ Hosp, Edegem, Belgium; 2) Dept Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands; 3) Center for Medical Genetics, Ghent Univ Hosp, Ghent, Belgium; 4) Dept Pediatrics, Ped Genet Unit, Hacettepe Univ School of Medicine, Ankara, Turkey; 5) Dept of Pediatrics, Clin Genet Unit, Akdeniz Univ School of Medicine, Antalya, Turkey; 6) Dept Medl Genet, National Institute of Health, Rabat, Morocco; 7) Dept of Radiology, Erasmus MC, Rotterdam, The Netherlands; 8) Centre for Pediatrics and Adolescent Medicine, Univ Freiburg, Freiburg, Germany.

Spondylo-megaepiphyseal-metaphyseal dysplasia (SMMD) is a rare autosomal recessive skeletal dysplasia with only a few cases reported in the literature. Affected individuals have a disproportionate short stature with a short neck and trunk and relatively long limbs that may show flexion contractures of the distal joints. The most remarkable radiographic features are the delayed and impaired ossification of the vertebral bodies as well as the presence of large epiphyseal ossification centers and wide growth plates in the long tubular bones. Genome wide homozygosity mapping followed by a candidate gene approach resulted in the elucidation of the genetic cause in three new consanguineous families with SMMD. Each proband was homozygous for a different inactivating mutation (c.336_337delGGinsT; c.337dupG; c.104_110delCGCCCG) in NKX3-2, a homeobox-containing gene located on chromosome 4p15.33. Expression studies in skin fibroblasts of one patient showed partial nonsense mediated decay of the mutant transcripts and upregulation of mutant NKX3-2 mRNA compared to control, the latter suggesting a disturbed autoregulatory loop. We analyzed the expression of NKX3-2 across different adult cell types. The highest expression was observed in chondrocytes and gut. In contrast to previous reports, we were able to demonstrate NKX3-2 expression in the nervous system (brain) with the more sensitive and quantitative qPCR assay. Striking similarities were found when comparing the vertebral ossification defects in SMMD patients with those observed in the Nkx3-2 null mice. Distinguishing features were the asplenia found in the mutant mice and the radiographic abnormalities in the limbs only observed in SMMD patients. This discrepancy may be explained by the perinatal death of Nkx3-2 null mice, before defects in the appendicular skeleton become apparent. In summary, our findings define SMMD as a distinct entity within the group of rare, autosomal recessive skeletal dysplasias. The identification of inactivating mutations in NKX3-2 underscores the crucial role of this homeobox-containing protein in ossification of the human vertebral column. The presence of mega- and pseudo-epiphyses with wide growth plates in tubular bones of SMMD patients confirms the more generalized role of NKX3-2 in endochondral ossification of both the axial and appendicular human skeleton.

59

Gene-gene interaction between SQSTM1 and TNFRSF11A increases the severity of the Paget's disease of bone. *F. Gianfrancesco[1], D. Rendina[2], A. Mingione[1], M. Di Stefano[3], T. Esposito[1], D. Merlotti[4], D. Formicola[1], S. Magliocca[1], S. Gallone[5], G. De Filippo[2], G. Morello[1], R. Nuti[4], P. Strazzullo[2], G.C. Isaia[3], G. Mossetti[2], L. Gennari[4].* 1) Institute of Genetics and Biophysics, Italian National Research Council, Naples, Italy; 2) Department of Clinical and Experimental Medicine, Federico II University Medical School, Naples, Italy; 3) Department of Internal Medicine, University of Turin, Italy; 4) Department of Internal Medicine, Endocrine-Metabolic Sciences and Biochemistry, University of Siena, Italy; 5) Department of Neuroscience, University of Turin, Italy.

Mutations in the p62/SQSTM1 gene were identified as a common cause of Paget's disease of bone (PDB) but experimental evidence demonstrates that mutation is not sufficient to induce PDB and additional factors acting together with p62 mutations are necessary for the development of PDB in vivo. We identified 2 non-synonymous SNPs in TNFRSF11A gene encoding Rank protein, associated with PDB and with the severity of phenotype in a large population of 642 unrelated patients that were previously screened for SQSTM1 gene mutations (the largest SQSTM1 mutation screening performed to date). Both SNPs (rs35211496, H141Y and rs1805034, V192A), that are conserved along evolutive scale and are localised in the two of four tandem cysteine-rich pseudo-repeat domain that are characteristic of the TNFRSF superfamily, were associated with PDB in the overall cohort as well as after the exclusion of patients with SQSTM1 gene mutations (n= 102). The largest effect was found for rs1805034, yielding an odds ratio of 1.3 (p=0.003), with the C allele as the risk allele. Moreover, an even more significant p-value (p=0.0003) was observed in the subgroup of patients with SQSTM1 mutation, with an odds ratio of 1.8. Interestingly patients with the C allele also showed an increased prevalence of polyostotic disease (68%, 53%, and 51% in patients with CC, CT, and TT genotypes, respectively, p=0.01) as well as an increased number of affected skeletal sites (2.9, 2.5, and 2.0 in patients with CC, CT, and TT genotypes, respectively, p=0.008). These differences increased in magnitude when analyses were restricted to cases with SQSTM1 mutation, both for the prevalence of polyostotic disease (91%, 77%, and 73% in patients with CC, CT, and TT genotypes, respectively, p=0.03) and the number of affected sites (4.4, 3.3, and 2.8 in patients with CC, CT, and TT genotypes, respectively, p=0.03). NFkB signaling was assessed in HEK293 cells co-transfected with expression plasmids for p62 wildtype (P392) or mutated (L392) and RANK (A192 or V192) and an NFkB luciferase reporter gene. Co-trasfection with L392 (p62) and A192 (Rank) produced a level of activation of NFkB signaling greater than co-trasfection with P392 and V192 confirming genetics and clinics studies. In conclusion, these results provide the first evidence that genetic variation within OPG/RANK/RANKL system influences the severity of PBD in synergistic action with SQSTM1 gene mutations.

60

Expression of the Hutchinson-Gilford progeria mutation during osteoblast development leads to irregular bone structure and impaired skeletal integrity. *E. Schmidt[1], O. Nilsson[2], J. Tuukkanen[3], C. Ohlsson[4], B. Rozell[5], M. Eriksson[1].* 1) Biosciences and Nutrition, Karolinska Institutet, Huddinge, Stockholm, Sweden; 2) Department of Women's and Children's Health, Karolinska Institutet and University Hospital, Stockholm, Sweden; 3) Department of Anatomy and Cell Biology, University of Oulu, Finland; 4) Department of Internal Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; 5) Clinical Research Center, Department of Laboratory Medicine, Karolinska Institutet, Karolinska University Hospital, Huddinge, Stockholm, Sweden.

Hutchinson-Gilford Progeria Syndrome (HGPS) is a very rare genetic disorder characterized by features of premature aging. Most commonly the disease is caused by a de novo point mutation in exon 11 of the LMNA gene (1824C>T, G608G). The resulting activation of a cryptic splice site leads to an aberrant splice product of lamin A and subsequently to an abnormally processed prelamin A protein named progerin. Lamins are intermediate filament proteins that form the network underlying the inner nuclear membrane. In this study we aim to explore the mechanisms of HGPS bone disease. Affected children develop growth retardation, joint stiffness and skeletal abnormalities including decreased longitudinal bone growth, skeletal dysplasias and atypical distal long bone demineralization diagnosed as focal osteoporosis.In our laboratory we have established a tissue-specific mouse model system utilizing the tet-ON/OFF for temporal and tissue specific expression of a minigene of human lamin A with the most common HGPS mutation (Sagelius et al. JCS 121, 2008). Expression of the HGPS mutation was directed to osteoblasts using the Osx1-GFP::Cre transactivator (Rodda et al. Development 133, 2006). Already at 5 weeks of age HGPS mutant mice showed postnatal growth retardation, imbalanced gait and spontaneous bone fractures. Histopathological examination show an irregular bone structure displayed by loss of osteocytes, mineralization defects and adipocyte invasion in the bone marrow space. pQCT analysis revealed reduced trabecular bone mineral density and increased cortical bone thickness. To explore the bone defect we focus on studying bone turnover by quantitative expression analysis of bone biomarkers including Runx2, Alp, Col1a1, OC, Dmp1, TRAP, RANKL, and OPG. Bone quality was assessed by biomechanics analysis and bone mineralization assays. Our results indicate underlying molecular mechanisms of bone disease in HGPS. The observed phenotype also shows similarities to reported bone abnormalities in aging mouse and may thus help to uncover general principles of the aging process.

61

Comprehensive evaluation of *ITGAM* and SLE susceptibility using HapMap and 1000Genomes data. *X. Kim-Howard[1], C. Sun[1], K. Kaufman[1,3], J.A. Kelly[1], J. Merrill[1], P.M. Gaffney[1], K.L. Moser[1], C.D. Langfeld[4], R.P. Kimberly[2], G. Gilkerson[5], T.J. Vyse[6], J.A. James[1], J.B. Harley[3], S.K. Nath[1], SLEGEN, PROFILE.* 1) Oklahoma Med Res Foundation, Oklahoma City, OK; 2) Musculoskeletal, and Autoimmunity Center, Department of Medicine, University of Alabama at Birmingham, USA; 3) Oklahoma City VA Medical Center Home, Oklahoma City, OK; 4) Department of Biostatistical Sciences and Center for Public Health Genomics, Wake Forest University Health Sciences, Winston-Salem, NC; 5) Ralph H. Johnson VAMC and the Department of Medicine, Division of Rheumatology, Medical University of South Carolina, 96 Jonathan Lucas Street, Suite 912, Charleston, SC; 6) Imperial College London, South Kensington Campus, London, England.

Introduction: We recently identified a novel non-synonymous variant, rs1143679, within *ITGAM* associated with systemic lupus erythematosus (SLE) susceptibility. Using a genome-wide association approach, 3 studies independently confirmed association between SLE susceptibility and the *ITGAM-ITGAX* region. We hypothesize that multiple independent, possibly ethnic-specific predisposing, variants within the *ITGAM-ITGAX* region are associated with SLE. **Objective**: Our objective is to perform a comprehensive analysis to assess whether single or multiple causal variants from the *ITGAM-ITGAM* region are involved in SLE susceptibility across 4 independent data sets (N=12,098). Data sets include 2 European-derived populations, European-American (EA) (3980 cases/3546 controls) and UK: (445/588), as well as 2 African-derived populations, African-American (AA) (1511/1759) and Gullah (GH) (135/134). **Methods**: We used 1000Genomes Project data as a reference panel to impute SNPs in this region, which allowed analysis of virtually all common SNPs (at least five-fold increase in marker density over HapMap) within the *ITGAM-ITGAX* region. Imputation-based association analyses were performed for single- and multiple-SNP association. To correct for population stratification effect in association we used 343 AIMs and applied both logistic regression and PCA. LD patterns and conditional logistic regression analysis were performed to assess the independence between SNPs. A meta-analysis was performed to obtain an overall odds ratio and p-value. **Results**: We genotyped 32 SNPs from the *ITGAM-ITGAX* region. We identified 195 SNPs in CEU and 281 SNPs in YRI from 1000Genomes database. Once rare (minor allele frequency<3%) SNPs were removed 167 (CEU) and 226 (YRI) SNPs were available for imputation based analysis. Although our analysis identified multiple associated variants, based on LD structure and subsequent conditional analysis the best significant SNP was identified as rs1143679 in all 4 populations (EA p=1.3E-30, UK p=5.7E-9, AA p=2.5E-9, GH p=0.004). Meta-analysis results are significant, the overall p-value, using all datasets are: p=3.6E-46 (OR=1.69, 95% CI (1.58-1.82)). However, we could not identify other SLE predisposing variants besides rs1143679. **Conclusion**: Our comprehensive imputation-based association analysis re-confirmed the exon-3coding SNP, rs1143679 (R77H), explained the entire *ITGAM*-SLE association.

62

A Common APOBEC3H Haplotype Associates with Decreased HIV-1 Sequence Editing and Lower HIV-1 RNA Set-Point In Early Untreated HIV-1 Infection. *J. Barbour[2], P.A. Gourraud[1], J. Woo[2], T. Schmidt[2], G.S. Spotts[2], F.M. Hecht[2], J.R. Oksenberg[1], T.J. Liegler[1].* 1) Neurology, Univ California, San Francisco, San Francisco, CA; 2) Medicine, Univ California, San Francisco, San Francisco, CA.

Introduction: We mapped genetic variation of the human APOBEC3 (A3) locus in a cohort of treatment naïve, recently HIV-1 infected adults to measures of sequence editing activity in the integrated HIV-1 DNA genome, and HIV-1 clinical markers. We hypothesized genetic variation within APOBEC3G and APOBEC3F would associate with variation in HIV-1 sequence in vivo. Methods: We examined single nucleotide polymorphisms (SNP) in the APOBEC3 locus on chromosome 22, paired to population sequences of pro-viral HIV-1 vif of peripheral blood mononuclear cells (PBMC), from 96 recently HIV-1 infected treatment naïve adults. Results: We found evidence for the existence of an APOBEC3H linkage disequilibrium (LD) block associated with variation in GA->AA, or APOBEC3F signature, sequence changes in pro-viral HIV-1 vif sequence. We identified a common 5 position risk haplotype telomeric to APOBEC3H (A3Hrh). These positions were in high LD (D' = 1; r^2=0.98) to a previously described A3H 'RED' haplotype containing a variant (E121) with enhanced susceptibility to HIV-1 Vif (Zhen et al 2009 [1]). Homozygote carriers of the A3Hrh had lower HIV-1 RNA levels over time during early, untreated HIV-1 infection, (p = 0.018 and 0.015 mixed effects model) and lower GA->AA (A3F) sequence editing on pro-viral HIV-1 vif sequence (p = 0.01). Discussion: We found an association of A3H genetic variants with an A3F (GA->AA) editing pattern upon HIV-1 vif DNA sequence. Our results suggest genetic variants of A3H that do not bear the E121 mutation controlling Vif susceptibility may exert a steady GA->AA sequence editing pattern upon HIV-1, contribute to viral diversification in vivo, and associate with elevated HIV-1 RNA levels.

63

Evidence for HPA axis gene variation in MS using supervised machine learning and meta-analysis in 12,566 individuals. *F.B.S. Briggs[1], S.E. Bartlett[2], B.A. Goldstein[1,3], J. Wang[4,5], J.L. McCauley[6], R.L. Zuvich[7], P.L. De Jager[8,9], J.D. Rioux[10], A.J. Ivinson[11], A. Compston[12], D.A. Hafler[9,13], S.L. Hauser[4,5], J.R. Oksenberg[4,5], S.J. Sawcer[12], M.A. Periack-Vance[6], J.L. Haines[7], B. Acuna[14], L. Shen[14], A. Bernstein[14], C.A. Schaefer[14], L.F. Barcellos[1], International Multiple Sclerosis Genetics Consortium.* 1) Genetic Epidemiology and Genomics Laboratory, Division of Epidemiology, School of Public Health, University of California, Berkeley, CA; 2) Ernest Gallo Clinic and Research Center, University of California, San Francisco, CA; 3) Division of Biostatistics, School of Public Health, University of California, Berkeley, CA; 4) Department of Neurology, School of Medicine, University of California, San Francisco, CA; 5) Institute for Human Genetics, School of Medicine, University of California, San Francisco, CA; 6) John P. Hussman Institute for Human Genomics, University of Miami, Miller School of Medicine, Miami, FL; 7) Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN; 8) Program in NeuroPsychiatric Genomics, Center for Neurologic Diseases, Department of Neurology, Brigham & Women's Hospital, Boston, MA; 9) Program in Medical & Population Genetics, Broad Institute of Harvard University and Massachusetts Institute of Technology, Cambridge, MA; 10) Laboratory in Genetics and Genomic Medicine of Inflammation, Montreal Heart Institute, Université de Montréal, Montreal, Quebec, H1T1C8, Canada; 11) Harvard NeuroDiscovery Center, Harvard Medical School, Boston, MA; 12) University of Cambridge, Department of Clinical Neurosciences, Addenbrooke's Hospital, BOX 165, Hills Road, Cambridge, CB2 2QQ, UK; 13) Department of Neurology, Yale School of Medicine, New Haven, CT; 14) Division of Research, Kaiser Permanente, Oakland, CA.

The primary genetic risk factor in multiple sclerosis (MS) is the *HLA-DRB1*1501* allele; however, much of the remaining genetic contribution to MS has yet to be elucidated. Several lines of evidence support a role for the involvement of hypothalamic-pituitary-adrenal (HPA) axis in autoimmunity which may, in part, be genetically determined. The HPA axis is a principal component of the neuroendocrine system that regulates individual response to physical and emotional stress, and maintains homeostasis with strong neuroimmune modulating properties. Impaired HPA axis activity has been observed in MS cases in several studies. The HPA axis is further implicated in pathology of MS by evidence from several studies supporting psychological stress as a risk factor for MS onset and exacerbations. Here, we comprehensively investigated variation within 8 candidate HPA axis genes and susceptibility to MS. A total of 326 SNPs were investigated in 1,343 MS cases and 1,379 healthy controls of European ancestry using a multi-analytical strategy. Random Forests, a supervised machine learning algorithm, identified 8 intronic SNPs within the corticotropin releasing hormone receptor 1 or *CRHR1* locus on 17q21.31 as important predictors of MS. Based on univariate analyses: six *CRHR1* variants were associated with decreased risk for disease following a conservative correction for multiple tests. Independent replication was observed for *CRHR1* in a large meta-analysis comprised of 2,624 MS cases and 7,220 healthy controls of European ancestry. Results from a combined meta-analysis of all 3,967 MS cases and 8,599 controls provide strong evidence for the involvement of *CRHR1* in MS (rs242936: p=9.7 x 10^{-5}). Further investigation of mechanisms involved in HPA axis regulation and response to stress in MS pathogenesis is warranted.

64

Genome-wide association study for psoriasis. *E. Ellinghaus[1], D. Ellinghaus[1], P.E. Stuart[2], R.P. Nair[2], S. Debrus[3], J.V. Raelson[3], M. Belouchi[3], H. Fournier[3], C. Reinhard[3], J. Ding[4], Y. Li[4], T. Tejasvi[2], J. Gudjonsson[2], S.W. Stoll[2], S. Lambert[2], S. Weidinger[5,6], B. Eberlein[5], M. Kunz[7], P. Rahman[8], D. Gladman[9], C. Gieger[10], H.E. Wichmann[10,11,12], T.H. Karlsen[13], D. Kabelitz[14], U. Mrowietz[6], G.R. Abecasis[4], J.T. Elder[2,15], S. Schreiber[1], M. Weichenthal[6], A. Franke[1].* 1) Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany; 2) Department of Dermatology, University of Michigan, Ann Arbor, USA; 3) Genizon BioSciences, Inc., St. Laurent, QC, Canada; 4) Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, USA; 5) Department of Dermatology and Allergy, Technische Universität München, Munich, Germany; 6) Department of Dermatology, University Hospital Schleswig-Holstein, Christian-Albrechts-University, Kiel, Germany; 7) Comprehensive Center for Inflammation Medicine, University of Lübeck, Lübeck, Germany; 8) Department of Medicine, Memorial University, St. John's, Newfoundland, Canada; 9) Division of Rheumatology University Toronto, Psoriatic Arthritis Program, University Health Network, Ontario, Canada; 10) Institute of Epdiemiology, Helmholtz Centre Munich, German Research Center for Environmental Health, Neuherberg, Germany; 11) Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-University, Munich, Germany; 12) Klinikum Grosshadern, Munich, Germany; 13) Medical Department, Rikshospitalet University Hospital, Oslo, Norway; 14) Institute for Immunology, University Hospital Schleswig-Holstein, Christian-Albrechts-University, Kiel, Germany; 15) Ann Arbor Veterans Affairs Hospital, Ann Arbor, USA.

Psoriasis is a chronic immune-mediated and hyperproliferative disorder of the skin that affects up to 3% of the Caucasian population. The most common form is psoriasis vulgaris (PsV) which is characterized by red, raised, scaly plaques that commonly occur on the elbows, knees, scalp and lower back. The first identified and now well-confirmed susceptibility locus is PSORS1 on chromosome 6 in the MHC-class1-region. In addition, part of the genetic susceptibility can be explained by polymorphisms at the established susceptibility loci *IL12B*, *IL23R*, *IL23A*, *IL4*/*IL13*, *TNIP1* and *TNFAIP3*. To identify additional psoriasis susceptibility loci, we successfully genotyped 561,466 SNPs in 472 psoriasis cases and 1146 controls of German ancestry. We followed up the 145 most significant SNPs in 2746 psoriasis cases and 4140 controls, consisting of three independent replication panels from Germany, the United States and Canada, respectively. In order to combine the results of all 3 cohorts and to include effect size parameters, we performed a Meta-analysis using METAL. Our results provide support for the association of psoriasis and a gene that encodes a protein which is involved in IL-17 signaling and regulating responses to cytokines by members of the Rel/NF-κB transcription factor family (intronic SNP: P=7.3×10^{-9} and missense SNP: P=2.7×10^{-10}). Logistic regression analysis revealed that the associations at this gene could be fully accounted for by the missense SNP (D19N) in exon 2. Both associations were further corroborated by findings in two additional replication panels (full study sample: 6487 cases vs. 8037 controls; P_{comb}=2.36×10^{-10} for the intronic SNP and P_{comb}=1.24×10^{-16} for the missense SNP). In a subset of German cases and controls (993 cases and 2277 controls) for which HLA-Cw6 status was available, we tested for the presence of a statistical interaction between HLA-Cw6 carriership and the missense SNP but found no evidence for it (P=0.77). Since about 15% of psoriasis patients develop psoriatic arthritis (PsA), we performed a PsA stratified analysis. We compared all PsA patients within our combined sample (1919 cases) versus 1919 randomly selected PsV cases and obtained no significant difference (P=0.13) which suggests that our gene locus represents a shared susceptibility for PsA and PsV.

65

Testing the role of common genetic variation in lean cases of type 2 diabetes reveals potential novel loci. *J.R.B Perry[1], V. Steinthorsdottir[2], Y. Labrune[3], P. Navarro[4], A. Petersen[5], B. Voight[6], M. Ganser[7], N. Amin[8], J. Dupuis[9], L. Qi[10], M. Li[11], K. Stefansson[2], R. Sladek[13], J.F. Wilson[14], T. Illig[5], D. Altshuler[6], M. Boehnke[7], C.M van Duijn[8], J.B. Meigs[15], F.B. Hu[10], J.S Pankow[16], P. Froguel[3], M. McCarthy[12], T.M. Frayling[1], S. Cauchi[3], DIAGRAM Consortium.* 1) Gen Complex Traits, Peninsula Med Sch, Exeter, United Kingdom; 2) deCODE Genetics, 101 Reykjavik, Iceland; 3) CNRS-UMR-8199, Institute of Biology and Lille 2 University, Pasteur Institute, F-59019 Lille, France; 4) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh, EH4 2XU, UK; 5) Institute of Epidemiology, Helmholtz Zentrum Muenchen, 85764 Neuherberg, Germany; 6) Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts 02142, USA; 7) Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, USA; 8) Department of Epidemiology, Erasmus University Medical Center, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands; 9) Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts 02118, USA; 10) Department of Nutrition, Harvard School of Public Health, 665 Huntington Ave, Boston, MA 02115, USA; 11) Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland 21287, USA; 12) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, OX3 7LJ, UK; 13) Department of Human Genetics, McGill University, Montreal H3H 1P3, Canada; 14) Centre for Population Health Sciences, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, UK; 15) Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA; 16) Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, Minnesota 55454, USA.

Type 2 diabetes is a highly heterogeneous common condition. Patients vary appreciably in the two major risk factors, age and extent of obesity. Few genome wide association (GWA) studies have tested the role of common genetic variants in subsets of patients. We hypothesised that lean type 2 diabetic patients would have stronger genetic risk because of their reduced phenotypic risk and that testing only lean type 2 diabetes patients (BMI <25Kg/m²) would yield novel loci not previously discovered in larger, more heterogeneous groups of patients.

We performed a case-control GWAS of lean type 2 diabetic patients versus healthy controls not selected for BMI, as part of the DIAGRAM consortium. Approximately 1500 cases and 50,000 controls were used, representing ~18% of diabetes cases from the total DIAGRAM discovery sample set. Analysis was performed using an additive model, meta-analysing 8 studies.

A total of 8 signals reached a significance threshold of P=1x10^{-6}, compared to an expected one signal by chance. These included 4 known Type 2 diabetes loci; *TCF7L2* (P = 3x10^{-29}, per-allele odds ratio (OR) = 1.61), *CDKAL1* (P = 4x10^{-10}, per-allele OR = 1.3), *HHEX* (P = 3x10^{-7}, per-allele OR = 1.22) and *WFS1* (P = 4x10^{-7}, per-allele OR = 1.24). For all four variants the point estimate effect size is larger than that observed in the main DIAGRAM GWAS. Novel associations at this threshold were observed at four additional loci: in or near *FOXJ2* (P = 1x10^{-7}, per-allele OR = 1.25), *TGFA* (P = 1x10^{-7}, per-allele OR = 1.3), *CD34* (P = 5x10^{-7}, per-allele OR = 2.1) and *LAMA1* (P = 9x10^{-7}, per-allele OR = 1.23). Of 41 variants known to associate with autoimmune forms of diabetes, the lowest P-value = 0.0004, with all remaining signals P > 0.05. This is consistent with minimal influence of a genetically-mediated autoimmune role in this form of diabetes, or minimal misclassification of patients between the two major forms of diabetes - both of which are more likely in lean compared to all diabetic patients.

Further validation of our novel signals is needed, but our study shows that focusing on sufficiently powered subsets of more homogeneous groups of patients may identify novel loci and help dissect the heterogeneity of common conditions.

66

Replication of the effect of *SLC2A9* genetic variation on serum uric acid levels in American Indians. *V.S. Voruganti[1], N. Franceschini[2], K. Haack[1], S. Laston[1], J.W. MacCluer[1], J.G. Umans[3], A.G. Comuzzie[1], P.G. Zager[4], K.E. North[2,5], S.A. Cole[1].* 1) Genetics, Southwest FNDN Biomedical Res, San Antonio, TX; 2) Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Medstar Research Institute, Washington DC; 4) Department of Medicine, Division of Nephrology, University of New Mexico Health Sciences Center, Albuquerque, NM; 5) Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Increased serum uric acid level or hyperuricemia is a risk factor for gout, renal disease and cardiovascular disease and is known to aggregate in families. The solute carrier protein 2 family, member 9 gene (*SLC2A9*) encodes a urate transporter that mediates urate flux across the renal proximal convoluted tubule. Genome-wide association studies have shown consistently that variants in the *SLC2A9* gene are associated with serum uric acid levels in Caucasian, African American, Mexican American and Asian populations. Our aim was to test whether this association generalized to American Indian participants of the Strong Heart Family Study (SHFS). We genotyped rs16890979 of *SLC2A9* in SHFS as part of the CALICO-PAGE consortium. Measured genotype analyses based on variance components decomposition method was conducted to assess the association between rs16890979 and serum uric acid. A strong association was found between this SNP (Minor allele frequency = 45 %) and serum uric acid in all centers combined ($p = 7.3 \times 10^{-31}$). When evaluated center-wise, rs16890979 was significantly associated with serum uric acid; ($p = 3.4 \times 10^{-11}$), ($p = 1.8 \times 10^{-08}$) and ($p = 2.6 \times 10^{-12}$), in Arizona, Dakotas and Oklahoma, respectively. This SNP is a missense variant leading to a valine-to-isoleucine amino acid substitution (V253I) and was associated with lower levels of serum uric acid in American Indians. The proportion of serum uric acid variance explained by this SNP was 4.4%, 3.8 %, 3.0 % , 4.8 % in all centers combined, Arizona, Dakotas and Oklahoma, respectively. We tested whether this association replicated in Zuni Indians of the Genetics of Kidney Disease in Zuni Indians Project and found a strong association between the SNP and serum uric acid levels ($p = 4.5 \times 10^{-08}$) and the SNP was associated with lower levels of serum uric acid. In summary, the association of *SLC2A9* SNP with serum uric acid levels generalizes to American Indians.

67

Phenotype-Wide Association Study (PheWAS) for Detection of Pleiotropy within the Multi-Ethnic Studies of the Population Architecture Using Genomics and Epidemiology (PAGE) Network. *S. Pendergrass[1], K. Brown-Gentry[1], S. Dudek[1], J.L. Ambite[2], C.L. Avery[3], S. Buyske[4], C. Cai[2], G. Heiss[3], L. Hindorff[5], C. Kooperberg[6], Y. Lin[6], T.A. Manolio[5], T. Matise[4], L. Wilkens[7], M.D. Fesinmeyer[6], C.-N. Hsu[2], D.C. Crawford[1], M.D. Ritchie[1].* 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Information Sciences Institute, University of Southern California, Marina del Rey, CA; 3) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 4) Department of Genetics and Department of Statistics, Rutgers University, Piscataway, NJ; 5) National Human Genome Research Institute, Bethesda, MD; 6) Public Health Sciences, Fred Hutchinson Cancer Research Institute, Seattle, WA; 7) Cancer Research Center, University of Hawaii, Honolulu, HI.

Over the past five years, genome-wide association studies (GWAS) have identified associations between SNPs and complex traits. The Population Architecture using Genomics and Epidemiology (PAGE) network was established in 2008 to characterize GWAS-identified variants in diverse population-based studies and identify pleiotropy for replicated GWAS variants by taking advantage of the rich phenotypic data gathered by each participating study. We describe here the results of the PAGE Phenotype-Wide Association Study (PheWAS) for three studies: the National Health and Nutrition Examination Surveys (NHANES; N), the Women's Health Initiative (WHI; W), and Atherosclerosis Risk in Communities (ARIC; A). We used phenotypic and genotypic information from these PAGE studies for the initial analysis: NHANES (93 SNPs, 152 phenotypes, ~7,000 participants), WHI (95 SNPs, 1,457 phenotypes, ~2,000-13,000 participants), and ARIC (69 SNPs, 611 phenotypes, ~4,000-10,000 participants). Overall, 37 SNPs were genotyped across the three studies and 52 SNPs were genotyped across at least two studies. After tests of association were performed between SNPs and phenotypes stratified by race-ethnicity within each study, six significant and novel SNP-Phenotype associations at $p<0.01$ were identified in more than one study. One SNP-phenotype association was found consistently in African-Americans across the three studies: *NOTCH2* rs10923931, originally associated with type 2 diabetes, was associated with white blood cell count (p=0.0014 (A); 0.0012 (N); 1.41E-10 (W)). Within the subset of analyses performed by only two studies, five novel SNP-phenotype associations were consistent in both studies for a single race-ethnicity. These include: *HHEX* rs1111875 in African-Americans for blood pressure and heart rate (p=0.0028 (A); 0.0088 (W)); *GALNT2* rs2144300 in European-Americans for serum total calcium (p=0.0017 (A); 0.0059 (N)); *GALNT2* rs2144300 in African-Americans with forced expiratory volume in three seconds (p=0.0028 (A); 0.0088 (N)); *IL6*R rs2228145 in African-Americans for white blood cell count, (p=7.97E-10 (A); 0.000392 (N)); and *LIPG* rs2156552 in European-Americans for goiter (p=0.0076 (A); 0.0086 (N)). This growing PAGE PheWAS resource will allow researchers to identify novel SNP-trait associations in multiple and diverse study populations, providing new directions for examination of clinical and functional implications of human genomic variation.

68

European Genetic Ancestry Protects Against the Development of Renal Disease in Systemic Lupus Erythematosus. *I.B. Richman[1], K.E. Taylor[1], S.A. Chung[1], L. Trupin[1], M. Petri[2], E. Yelin[1], R.R. Graham[3], A. Lee[4], T.W. Behrens[3], P.K. Gregersen[4], M.F. Seldin[5], L.A. Criswell[1].* 1) Rosalind Russell Medical Research Center for Arthritis, Division of Rheumatology, University of California, San Francisco, San Francisco, CA; 2) Division of Rheumatology, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Immunology Biomarkers Group, Genentech, South San Francisco, California, USA; 4) The Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, NY; 5) Rowe Program in Genetics, University of California at Davis, Davis, CA.

Background: African Americans, Asians, and Hispanics with systemic lupus erythematosus (SLE) are at greater risk of developing renal disease than Caucasians. The extent to which genetic or environmental factors contribute to these disparities remains uncertain. We hypothesized that in a multiethnic SLE case series, European genetic ancestry would be protective against the development of renal disease, and that the effect of ancestry would be partially attributable to variation in genes associated with renal disease in SLE. Methods: This was a cross-sectional study of 1910 adults with SLE enrolled in two case collections. Clinical data were obtained by chart review. Self-reported socioeconomic data were available for a subset of participants (n=1129). All participants were genotyped for 116 single nucleotide polymorphisms (SNPs) informative for continental ancestry. A subset of participants (n=753) was also genotyped for 80 SNPs in 14 candidate genes for renal disease in SLE and had *HLA-DRB1* typing available. Using the program STRUCTURE, we estimated genetic ancestry for each participant assuming 5 populations. We used logistic regression to test the association between European ancestry and renal disease, adjusting for disease duration and sex. Subsequent analyses assessed whether also adjusting for other continental ancestries, educational attainment, or variation in genes associated with renal disease attenuated the relationship between European ancestry and renal disease. Results: Genetic ancestry estimation for 1910 SLE cases demonstrated that participants had on average 62.4% European, 15.8% African, 11.5% East Asian, 6.5% Amerindian, and 3.8% South Asian ancestry. Among participants, 34.4% (n=656) had a history of renal disease. After adjustment for ancestry, SNPs in *IRF5* (rs4728142), *BLK* (rs2736340), *STAT4* (rs3024912), *ITGAM* (rs9937837) and the HLA alleles *DRB1*0301* and *DRB1*1501* were associated with renal disease in SLE ($p<0.05$). In multivariable logistic regression, a 10% increase in European ancestry was associated with a 15% reduction in the odds of having renal disease (OR 0.85, 95% CI 0.82-0.87, p=1.9e-30). Adjusting for other genetic ancestries, education, or genes associated with renal disease did not substantively alter this relationship. Conclusions: European ancestry is associated with lower risk of renal disease in SLE, an effect independent of other genetic ancestries, common risk alleles, and socioeconomic status.

69

Copy Number and SNP Variation in Oxidative Stress Pathway Genes Modify Health Effects of Air Pollution Reduction during the 2008 Beijing Olympics. *S. Diehl[1,2], F. Kuo[1,2], C. Huang[1], W. Huang[3], C. Chou[1,2], H. Kipen[4], D. Rich[5], S. Lu[5], P. Ohman-Strickland[5], Y. Wang[6], J. Zhao[7], M. Hu[3], G. Wang[7], P. Zhu[7], D. Thomas[8], T. Zhu[3], J. Zhang[5].* 1) Center for Pharmacongenomics and Complex Disease Research, UMDNJ- NJ Dental School, Newark, NJ; 2) Biomedical Informatics, UMDNJ-School of Health Related Professions, Newark, NJ; 3) Peking University College of Environmental Sciences and Engineering, Beijing; 4) UMDNJ-Robert Wood Johnson Medical School and EOHSI, Piscataway, NJ; 5) UMDNJ-School of Public Health, Piscataway, NJ; 6) Peking University Health Sciences Center, Beijing; 7) Peking University First Hospital, Beijing; 8) University of Southern California, Los Angeles, CA.

Epidemiological and experimental studies show that short-term increases in air pollution are associated with cardiopulmonary morbidity and mortality. Understanding of the mechanistic pathways underlying these associations and especially the role of genetic variation is extremely limited. In this study, biomarkers of pulmonary and systemic inflammation and oxidative stress were obtained in 128 healthy Chinese medical residents who experienced major reduction in air pollution during the 2008 Beijing Olympics. Blood samples, exhaled breath condensate, blood pressure and heart rate variability were obtained before the Olympics when air pollution levels were high, during the Olympic period when pollution was substantially reduced, and after the Olympics when pollution increased. Pollutants measured included PM2.5, elemental carbon (EC), organic carbon (OC), CO, NO2, SO2 and O3. QPCR methods were used for copy number polymorphisms (GSTM1 and GSTT1) and TaqMan methods for SNPs (GSTP1 and HMOX1). These genes are potential mediators of oxidative stress related to air pollution exposure. Analyses of the dose-response relationship between each biomarker and pollutant were conducted using random effects regression, controlling for temperature and humidity and evaluated across lags of 0-6 days for pollution exposures. Genotype influences on biomarker levels (main effects) and in altering the impact of changes in air pollution (GXE interactions) were determined, including gender-specific effects. Preliminary analyses show that exhaled nitric oxide, a measure of airway inflammation, exhibits highly significant changes associated with reduction in levels of several pollutants. The magnitude of change was significantly influenced by GSTP1 genotype for CO exposure (P=0.0007) and HMOX1 genotype for O3 exposure (P=0.00008). Systolic blood pressure showed a highly significant reduction associated with decreases in several pollutants especially CO, NO2, EC and OC. No gene interactions were found for CO effects. However, the three other pollutants showed evidence of interactions with HMOX1 in male subjects only (NO2 P=0.0002; EC P=0.00007; OC P=0.0003). No strong associations were found with copy number polymorphisms in GSTM1 or GSTT1 in analyses completed to date. These findings demonstrate the importance of genetic variation in modifying the health effects of air pollution exposure. Supported by HEI 4760-RPFA05-3, NIEHS ES015864, NIEHS ES05022 and the Foundation of UMDNJ.

70

Looking for Genes in All the Wrong Places? *N. Cox[1], E. Gamazon[1], S. Das[2], N. Rasouli[3,4], P. Kern[5], S. Elbein[2].* 1) Section of Genetic Medicine, University of Chicago, Chicago, IL; 2) Section on Endocrinology and Metabolism, Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, NC 27152; 3) Division of Endocrinology, Department of Internal Medicine, University of Colorado School of Medicine, Aurora, CO; 4) Research Service, Eastern Colorado Veterans Healthcare Network, Denver, CO; 5) Division of Endocrinology, Department of Internal Medicine, University of Kentucky School of Medicine, Lexington, KY.

SNPs associated through GWAS with complex human traits are more likely than minor-allele-frequency matched SNPs on high throughput GWAS platforms to be eQTLs in lymphoblastoid cell lines (LCLs) (Nicolae et al, PLoS Genetics 2010), and we have now extended these studies to adipose and muscle tissue collected from 62 individuals selected from 182 total study participants for extremes of insulin sensitivity. The 41 SNPs reproducibly associated with T2D were significantly enriched for eQTLs identified through either adipose ($p<3.6 \times 10^{-8}$) or muscle ($p<3.5 \times 10^{-7}$), but not LCLs. Moreover, the enrichment of eQTLs in both muscle and adipose was driven entirely by SNPs regulating distant (> 4 Mb) rather than local transcripts. To determine whether this observation would hold more generally for T2D-associated SNPs, we examined the top 1000 SNPs associated with T2D in the WTCCC (WTCCC, Nature 2007) and found marginal support for enrichment of LCL eQTLs ($p< 0.08$), whereas the enrichment of muscle and adipose eQTLs among the top 1000 T2D-associated SNPs was highly significant ($p< 2.4 \times 10^{-7}$, and $p< 4.6 \times 10^{-8}$, respectively) and driven entirely by SNPs predicting expression of distant transcripts. Results were qualitatively similar for the top 10,000 T2D-associated SNPs in the WTCCC studies. To determine if the predominance of SNPs regulating distant transcripts among T2D-associated SNPs is the exclusive pattern observed for complex traits, we examined the enrichment of eQTLs from LCLs, adipose, and muscle among SNPs reproducibly associated with type 1 diabetes (T1D). We found that SNPs regulating local and distant transcripts were both significantly enriched among SNPs reproducibly associated with T1D, and this pattern was consistently observed for all three tissues examined. Our results confirm that SNPs reproducibly associated with complex disorders may often affect susceptibility to disease through the regulation of transcript levels. For at least some disorders, including T2D, there is substantial evidence that trait-associated SNPs are enriched primarily for SNPs regulating distant rather than local transcripts. Efforts to understand the biology underlying T2D-associated SNPs may benefit from consideration of distant transcripts regulated by those SNPs, and sequencing strategies designed to identify genes more likely to harbor rare variants increasing risk of T2D may need to consider genes other than those physically near T2D-associated SNPs.

71

Evidence for a role of GABA and CRH system genes in depressive symptoms: The Collaborative Study on the Genetics of Alcoholism. *D.A. Kertes[1], J. Kramer[2], H.J. Edenberg[3], J.I. Nurnberger Jr.[4], V. Hesselbrock[5], M.A. Schuckit[6], L.J. Beirut[7], B. Porjesz[8], J.A. Tischfield[9], B.P. Riley[10], D.M. Dick[11].* 1) Department of Psychology and UF Genetics Institute, University of Florida, Gainesville, FL; 2) Department of Psychiatry, University of Iowa College of Medicine, Iowa City, IA; 3) Department of Biochemistry and Molecular Biology, and Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN; 4) Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN; 5) Department of Psychiatry, University of Connecticut, Farmington, CT; 6) Department of Psychiatry, University of California-San Diego, La Jolla, CA; 7) Department of Psychiatry, Washington University School of Medicine, St. Louis, MO; 8) Department of Psychiatry, State University of New York, Brooklyn, NY; 9) Department of Genetics, Rutgers University, Piscataway, NJ; 10) Departments of Psychiatry and Human Genetics, Virginia Commonwealth University, Richmond VA; 11) Departments of Psychiatry, Psychology, and Human Genetics, Virginia Commonwealth University, RIchmond, VA.

Background: Depressive symptoms are common among individuals with alcohol use disorders and impact treatment outcomes. Substantial overlap exists among the neurobiological systems proposed in the pathophysiology of depressive and alcohol use disorders. Specific genetic effects contributing to risk for alcohol use disorders with comorbid depression remain poorly understood. The purpose of this study was to conduct a genetic association analysis of depressive symptoms among alcohol dependent (AD) individuals. Methods: Participants were 847 AD individuals enrolled in the Collaborative Study on the Genetics of Alcoholism (COGA) genotyped using the Illumina 1M BeadChip. Nineteen candidate genes were selected from 120 genes implicated in AD or major depression, based on an initial screen of an independent AD sample (the Irish Affected Sib-Pair Study of Alcohol Dependence). The initial screen included genes involved in opioid, dopamine, serotonin, glutamate, and GABA neurotransmission, cell signaling, pharmacokinetics, stress biology and behavioral control. Genes with markers showing nominally significant association in the screening sample were tested for association in the COGA participants. Mantel Haenszel χ^2 test for linear association was followed by matrix spectral decomposition and multiple test correction. Results: A genetic variant in GABRA2 was significantly associated with depressive symptoms among AD individuals in the COGA sample. The associated allele was the same as that observed in the screening sample, indicating a marker-level replication across two independent studies. Gene-level replication with depressive symptoms was observed for CRHBP and GABRB2. An examination of LD indicated these signals were at least partially independent from those observed in the screening sample but remained significant following correction for multiple testing. Conclusions: These findings highlight potential risk genes for depressive symptoms in alcohol dependent individuals and underscore the utility of examining comorbid conditions in genetic association analyses.

72

Galanin receptor 1 (GalR1) gene associates with tobacco craving in smokers seeking treatment cessation therapy. *A. Lori[1], T. Yilang[1], S. O'Malley[2], M. Picciotto[2], R. Wu[2], J. Cubells[1,3].* 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322; 2) Department of Psychiatry, Yale University School of Medicine, New Haven, CT 06510; 3) Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA 30322.

Background: Cigarette smoking is the leading preventable cause of death worldwide and is perpetuated by nicotine dependence (ND). Craving for tobacco is a major challenge for people with ND who try to quit smoking. However few genetic studies of tobacco craving have been conducted. Galanin and its receptors have been associated with addiction related behaviors, such as withdrawal response, probably because they interact with mesolimbic dopamine transmission. Methods: We performed a genetic association study in 590 subjects (432 European American) with ND (DSM-IV), recruited for smoking cessation trials. Twenty-six candidate genes for ND-related phenotypes were selected based on the literature and typed at tag SNPs. Subjects were assessed using the Minnesota Withdrawal Scale (MNWS), which included an assessment of craving, and several other ND-related instruments. General linear models examined associations between genotypes and ND-related scores in the Caucasian, non-Caucasian or pooled samples, incorporating appropriate covariates and including estimates of proportions of chromosomal ancestry based on a panel of ancestry-informative markers (AIMs). Results: Our analyses found one SNP in GalR1, rs2717162, significantly associated with severity of craving in Caucasian samples (p=3.68x10-6) and in the combined sample (p=1.02x10-8). Individuals with TT and TC genotypes had significantly higher MNWS craving scores than CC subjects. Proportions of chromosomal ancestry did not associate with the craving score. Alcohol-related behaviors and depression scores likewise did not account for the finding. 40 SNPs were found nominally associated with ND, withdrawal symptoms or craving. Among those were rs16969968 and rs684513 in CHRNA5, both of which have been associated with ND-related phenotypes in prior studies. Conclusion: The current results replicate previously reported associations between CHRNA5 and ND. Our findings also suggest that variation at GalR1 associates with differences in the severity of craving for tobacco in smokers motivated to quit. The results appear unlikely to reflect population stratification. Together with preclinical evidence, these results, if replicated, suggest that galanin and its receptors may be useful therapeutic targets for the pharmacological treatment of ND.

73

Genome-wide association of bipolar disorder suggests an enrichment of true associations in regions near exons. *E.N. Smith[1], D.W. Craig[2], N.J. Schork[1,3], J.R. Kelsoe[4,5], The Bipolar Genome Study (BiGS) Consortium.* 1) Scripps Genomic Med, Scripps Translational Sci, La Jolla, CA; 2) Neurogenomics Division, The Translational Genomics Research Institute, Phoenix, AZ, 85004; 3) Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA 92037; 4) Department of Psychiatry, University of California, San Diego, La Jolla, CA 92093; 5) Department of Psychiatry, VA San Diego Healthcare System, La Jolla, CA, 92151.

Bipolar disorder is a highly heritable disorder that has not easily yielded genetic loci from family and population based mapping strategies. We report results from a genome-wide association study with 2,191 bipolar disorder (BD) cases and 1,434 controls genotyped on the Affymetrix 6.0 platform including the GAIN Bipolar samples and 1,591 newly genotyped samples. We do not find any associations individually reaching genome-wide significance in the study or in a meta-analysis with the Wellcome Trust Case Control Consortium (WTCCC) BD study. However, when we highlight SNPs with stronger associations from the WTCCC BD study, we see an enrichment of associations at $P < 0.05$ in SNPs for which we have the highest power, indicating the presence of underlying true associations and consistent with a polygenic model of inheritance for BD. The association between power and replication is significant ($P=7x10^{-8}$) and is not observed when SNPs are prioritized based on results from phenotypes other than BD. Removal of previously implicated regions does not reduce the association, supporting the involvement of novel loci. We partitioned SNPs by their distance relative to exon coding regions and show that SNPs close to exons have a stronger enrichment for replication than SNPs away from exons, supporting an enrichment of true associations for bipolar disorder in regions near exons. These results indicate that despite a lack of genome-wide significance for any individual variant in a large study, analysis of groups of SNPs that show consistent effects across studies can provide insight into the underlying genetic mechanism of bipolar disorder.

74

Machine Learning Methods Predict the Serotonin Transporter Promoter Polymorphism Genotypes (5-HTTLPR) with SNPs. *A.T. Lu[1], S. Bakker[3], R.A. Ophoff[1,2,3], R.M. Cantor[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, CA; 3) Rudolf Magnus Institute of Neuroscience, Department of Psychaitry, University Medical Center Utrecht, Utrecht, The Netherlands.

The serotonin transporter gene (SLC6A4) and its promoter (5-HTTLPR) polymorphism have been the focus of a large number of association studies of neurodegenerative and neuropsychiatric disorders, neural activity in the limbic and prefrontal brain regions in response to emotional stimuli, and treatment responses of those with behavior disorders. The 5-HTTLPR polymorphism is difficult to genotype, probably due to high GC content of target sequence of the promoter region. We postulated that a good model for predicting 5-HTTLPR from nearby SNPs would benefit association studies between this hypothesized functional variant and many neurobehavioral syndromes, now that large study samples with SNP genotypes are being established. To enable such association studies, we fit a prediction model with eight SNPs from the 200kb region surrounding 5-HTTLPR on 17q11.2. Two machine learning methods, multicategory vertex discriminant analysis (VDA) and support vector machine (SVM), were used. Both approaches allow a nonlinear relationship between the genotypes and the predictors. Ten-fold cross validation was used to search the associated parameters under the possible models in order to optimize misclassification rates, and a step-wise approach was applied to select the minimal set of predictors. Two study samples genotyped with 77regional tagging SNPs in common were used to develop, train and test the learning models. A sample from The Northern Finland Birth Cohort from 1966 consisting of 2,147 individuals, genotyped with the Illumina 370K SNP Chip and manually genotyped for the 5HTTLPR polymorphism was used to train and test the learning models, and a second sample of 150 Dutch schizophrenia patients and 126 controls genotyped on the Illumina HumanHap550 BeadChip platform and manually genotyped for the 5-HTTLPR polymorphism was used to test the model in an independent sample. The prediction model of eight SNPs achieved a 92 % accuracy rate and 0.98 ± 0.01 AROC (area under the receiving operating characteristic). These evaluation statistics support the ability this model to predict the 5-HTTLPR genotypes with substantial. Using this model can reduce the financial and time costs of genotyping the 5-HTTLPR polymorphism for large association studies, and applying this approach can predict other important polymorphisms that also prove difficult to genotype.

75

Genome-wide association study identifies variants at four loci as genetic risk factors for Parkinson's disease. *W. Satake[1,2], Y. Nakabayashi[1], I. Mizuta[1], M. Kubo[3], T. Kawaguchi[3], T. Tsunoda[3], M. Watanabe[4], A. Takeda[5], H. Tomiyama[6], K. Nakajima[7], K. Hasegawa[8], F. Obata[9], T. Yoshikawa[10], H. Kawakami[11], S. Sakoda[2], M. Yamamoto[12], N. Hattori[6], M. Murata[13], Y. Nakamura[3,14], T. Toda[1].* 1) Div Neurol/Mol Brain Sci, Kobe Univ Grad Sch Med, Japan; 2) Dept Neurol, Osaka Univ Grad Sch Med, Japan; 3) Center for Genomic Med, RIKEN, Japan; 4) Dept Neurol, Univ Tsukuba, Japan; 5) Div Neurol, Tohoku Univ Grad Sch Med, Japan; 6) Dept Neurol, Juntendo Univ Sch Med, Japan; 7) Dept Neurol, Tottori Univ Faculty Med, Japan; 8) Dept Neurol, Sagamihara Natl Hsp, Japan; 9) Div Clinical Immunol, Kitasato Univ, Japan; 10) RIKEN Brain Sci Institute, Japan; 11) Dept Epidemiol, Hiroshima Univ, Japan; 12) Dept Neurol, Kagawa Pref Cen Hsp, Japan; 13) Dept Neurol, Natl Center Hsp Neurol and Psy, Japan; 14) Human Genome Center, Univ Tokyo, Japan.

Parkinson's disease (PD) is one of the most common neurodegenerative diseases worldwide. To identify susceptibility variants for PD, we performed a genome-wide association study (GWAS) using >550K SNPs in a sample set of 1,078 cases and 2,628 controls from Japan. After strict sample and SNP quality controls (MAF >0.05 etc.), we tested for association between each SNP and PD. We selected the 337 most associated SNPs (P <0.00054) from the GWAS, and conducted two subsequent replication studies in a total of 933 cases and 15,753 controls from Japan. We identified a new susceptibility locus on 1q32 and designated this as *PARK16* (rs947211, $P = 1.52 \times 10^{-12}$). Three genes (*NUCKS1*, *RAB7L1* and *SLC41A1*) were contained in the same LD block as rs947211. Interestingly, e-QTL analyses revealed that rs947211 was the principal genetic determinant of variation in expression levels of *NUCKS1* ($P = 6.0 \times 10^{-15}$). This highlights *NUCKS1* as a promising susceptibility gene for PD. We also identified *BST1* on 4p15 as a second new risk locus (rs4538475 , $P = 3.94 \times 10^{-9}$). This supports that Ca^{2+} dishomeostasis is a possible cause of selective vulnerability of dopaminergic neurons in PD. We also detected strong associations at α-*synuclein* (rs11931074, $P = 7.35 \times 10^{-17}$) and *LRRK2* (rs1994090, $P = 2.72 \times 10^{-8}$), both of which are implicated in autosomal dominant forms of parkinsonism. These data show that genes causative for autosomal dominant parkinsonism through their mutations also confer risk of typical PD through their common variants. Population attributable risks for rs947211 (*PARK16*), rs4538475 (*BST1*), rs11931074 (α-*synuclein*) and rs1994090 (*LRRK2*) were estimated to be 13%, 8%, 18% and 3%, respectively. To further identify relatively rare genetic risks, we also analyzed our GWAS data in detail by diminishing a SNP quality control (MAF from 0.05 to 0.01). We detected other association signals at 4 SNPs showing $P < 10^{-6}$; 9 SNPs, $P = 10^{-6}$-10^{-5}; and 12 SNPs, $P = 10^{-5}$-10^{-4}, which are candidates of PD genetic risks to be confirmed. To evaluate population genetic heterogeneity of PD risks, we compared results of a Caucasian GWAS. We identified *PARK16*, α-*synuclein* and *LRRK2* as shared risk loci for PD and *BST1* and *MAPT* as loci showing population differences. Our results identify two new PD susceptibility loci, show involvement of autosomal dominant parkinsonism loci in typical PD and suggest that population differences contribute to genetic heterogeneity in PD.

76

Genome-Wide Association Study Identifies Five Novel Susceptibility Loci. *P. Gejman[1], S. Ripke[2], M. Daly[2], A. Sanders[1], K. Kendler[3], P. Sullivan[4], P. Sklar[5], T. Werge[6], A. Corvin[7], M. O'Donovan[8], D. Rujescu[9], S. Cichon[10], B. Mowry[11], D. St. Clair[12], R. Ophoff[13], M. Owen[8], H. Gurling[14], D. Blackwood[15], D. Levinson[16], O. Andreassen[17], C. Pato[18], A. Malhotra[19], T. Lehner[20], A. Fanous[21], P. Visscher[22], P. Holmans[8], s. GENE[23], Schizophrenia Group, Psychiatric Genetics Consortium.* 1) NUH RI & University of Chicago, Evanston, IL; 2) Massachusetts General Hospital, Boston, MA; 3) Virginia Commonwealth University School of Medicine, Richmond, VA; 4) University of North Carolina at Chapel Hill, Chapel Hill, NC; 5) Broad Institute, Massachusetts General Hospital, Boston, MA; 6) Institute of Biological Psychiatry, Sct. Hans, University of Copenhagen, Copenhagen, Denmark; 7) Institute of Molecular Medicine, Trinity College Dublin, Dublin, Ireland; 8) School of Medicine, Cardiff University, Cardiff, Wales, United Kingdom; 9) Ludwig-Maximilians-University, Munich, Germany; 10) Institute of Human Genetics, University of Bonn, Bonn, Germany; 11) Queensland Brain Institute; University of Queensland; Brisbane, Queensland, Australia; 12) Institute of Medical Sciences, University of Aberdeen, Aberdeen, Scotland, United Kingdom; 13) University of California at Los Angeles, Los Angeles, CA; 14) University College London Medical School, Windeyer Institute of Medical Sciences, London, England, United Kingdom; 15) University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, Scotland, United Kingdom; 16) Stanford University; Stanford, CA; 17) Institute of Clinical Medicine, University of Oslo, Oslo, Norway; 18) Keck School of Medicine, University of Southern California, Los Angeles, CA; 19) The Zucker Hillside Hospital, North Shore-Long Island Jewish Health System, Glen Oaks, NY; 20) National Institute of Mental Health, Bethesda, MD; 21) Washington Veterans Administration Medical Center, Washington, DC; 22) Queensland Institute of Medical Research, Brisbane, Queensland, Australia; 23) deCODE Genetics, Reykjavik, Iceland.

Schizophrenia is a common, severe, and highly heritable psychotic disorder for which biological insight and effective therapeutics remain largely elusive. Previous studies have identified several common and rare genetic loci associated with schizophrenia; however, together these explain only a small fraction of the heritability. We performed a combined genome-wide association study (GWAS) of schizophrenia in samples of European ancestry comprised of a total of 21,856 individuals (9,394 cases and 12,462 controls), and followed up the strongest signals in 30,273 independent samples - by far the largest schizophrenia sample investigated to date. Strong experiment-wide evidence for replication of our initial findings supported a multigenic model for schizophrenia inheritance. By combining the initial meta-analysis and follow up results, we found a total of seven genome-wide significant loci, of which five are novel (1p21.3, 2q32.3, 8p23.2, 8q21.3, and 10q24.32) and two (6p21.3-22.1, and 18q21.2) are replicated from previous studies.

77

Local brain structure alteration is a mediator of genetic risk for neuropsychiatric disorders and facilitates gene finding in these disorders. *B. Franke[1,2,3], A. Arias Vasquez[1,2,3], J.A. Veltman[1], P. Hagoort[3,4], H.G. Brunner[1], M. Rijpkema[3,4], G. Fernandez[3,4,5].* 1) Dept. of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 2) Dept. of Psychiatry, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 3) Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen, Nijmegen, The Netherlands; 4) Donders Centre for Cognitive Neuroimaging, Nijmegen, The Netherlands; 5) Dept. of Neurology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

Alteration of regional brain structure is an inherent feature of many neuropsychiatric disorders, but it remains unclear if this is causally related to pathogenesis. If this were the case, alterations of brain structure should also be present in healthy individuals carrying genetic risk factors for such disorders. This, in turn, would imply that brain structure in the healthy population can be used as an intermediate phenotype for gene finding in neuropsychiatric disorders. We enrolled over 1500 healthy individuals in the Brain Imaging Genetics (BIG) study (mean age 28 years). For all BIG participants, structural MRI brain images are available from scans at 1.5 or 3 Tesla. FSL-FIRST is used to assess volumes of specific brain structures. To date, genotyping of several candidate genes was performed, as was genome-wide genotyping on Affymetrix 6.0 arrays in 1000 participants. Hypothesis-driven testing of candidate genes for neuropsychiatric disorders lead (among others) to the following findings: Gene-wide analysis of CACNA1C, a gene showing genome-wide association with bipolar disorder, revealed a strong effect of a cluster of SNPs in intron 3 on brainstem volume (n=600; p=3.62E-05). This effect was seen in two independent cohorts and was confirmed by voxel-based morphometry (VBM). Brainstem, with its central control over motor, cognitive, affective and arousal functions, constitutes an interesting novel mode of action of risk factors for bipolar disease. Testing of the GWAS-derived risk SNP for Alzheimer's disease (AD) in CR1, we found a significant reduction of gray matter volume in entorhinal cortex, the first structure to be affected by AD. This finding was achieved using VBM analysis in a discovery sample of 430 BIG-participants (p(SVC)=0.032) and replicated in a second sample of 492 subjects (p(SVC)=0.039). Preliminary GWAS of amygdala volume in 600 samples resulted in association findings for several suspected psychiatric risk genes such as CDH13, featuring among the top-findings of GWAS in disorders including ADHD, addiction and schizophrenia, but also the personality traits extraversion and neuroticism. In conclusion, risk factors for neuropsychiatric disorders indeed affect brain structure in healthy individuals suggesting that brain structure is cause rather than consequence of disease. Our preliminary GWAS results imply that healthy brain structure is a powerful intermediate phenotype for gene discovery in neuropsychiatric disorders.

78

Polymorphisms in FKBP5 define biologically distinct subtypes of PTSD: evidence from endocrine and gene expression studies. *D. Mehta[1], M. Gonik[1], M. Rex-Haffner[1], T. Klengel[1], A. Menke[1], J. Rubel[1], KB. Mercer[3], B. Puetz[1], B. Bradley[2,4], F. Holsboer[1], K. Ressler[2,5,6], B. Müller-Myhosok[1], EB. Binder[1,2,3].* 1) Max Planck Inst Psychiatry, Munich, Germany; 2) Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, USA; 3) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA; 4) VA Medical Center, Atlanta, GA, USA; 5) Yerkes National Primate Research Center, Atlanta, GA, USA; 6) Howard Hughes Medical Institute, Chevy Chase, Maryland.

Gene-environment interactions with polymorphisms in the FKBP5 gene are likely important in the etiology of post traumatic stress disorder (PTSD). FKBP5 is a part of the mature glucocorticoid receptor (GR) heterocomplex and is known to regulate GR sensitivity, which in turn has been shown to be disturbed in PTSD. We have previously shown that PTSD-related GR-hypersensitivity might be restricted to individuals carrying the FKBP5 risk allele, suggesting that these polymorphisms might define biologically distinct subtypes of PTSD (Binder et al, 2008). The aims of this study were to investigate interactions of the FKBP5 SNP rs9296158 and PTSD symptoms on baseline cortisol and low dose dexamethasone (0.5 mg) suppression (n=217) as well as on whole blood gene expression (Illumina HT12 microarray) (n=112) in a low-income highly traumatized, primarily African-American cohort. We replicated our earlier interaction between the FKBP5 polymorphisms and PTSD symptoms on low dose dexamethasone suppression (p= 0.037), with higher GR sensitivity in risk allele carriers with PTSD. In addition, we were able to show a significant interaction between rs9296158 and PTSD symptom severity on baseline cortisol levels (p=0.027), with higher PTSD symptoms correlating with low baseline cortisol levels only in carriers of the protective genotype. The importance of FKBP5 polymorphisms in the biology of PTSD was also reflected in the gene expression data. Significant effects on gene expression were only observed when rs9296158 and its interaction term with PTSD symptoms was included in the prediction model, but not with PTSD symptoms alone, with effects on the FKBP5 and RPS29 transcripts surviving correction for multiple testing. When analyzing all 18 transcripts with nominal significance in this model, genotype-dependent effects of PTSD were apparent most prominently in GR-sensitive transcripts. These findings were validated using qPCR and replicated in an independent sample (N = 40). In summary, the data presented here suggest that genetic variation in the FKBP5 gene is associated with biologically distinct subtypes of PTSD. These subtypes of PTSD are likely different on the level of GR sensitivity, which is reflected in different baseline cortisol levels, dexamethasone suppression sensitivity, and gene expression profiles of GR-sensitive genes.

79

Spinocerebellar ataxia type 31 (SCA31) common in Japan is caused by a founder insertion consists of penta-nucleotide repeats containing $(TGGAA)_n$. *K. Ishikawa[1], N. Sato[1], T. Amino[1], K. Kobayashi[2], S. Asakawa[3], T. Ishiguro[1], M. Takahashi[1], T. Matsuura[4], T. Tsunemi[1], K. Flanigan[5], S. Iwasaki[6], F. Ishino[6], Y. Saito[7], S. Murayama[7], M. Yoshida[8], Y. Hashizume[8], Y. Takahashi[9], S. Tsuji[9], N. Shimizu[3,10], T. Toda[2], H. Mizusawa[1].* 1) Dept Neurology, Tokyo Med & Dental Univ, Tokyo, Japan; 2) Dept Neurology and Neuroscience, Kobe University, Kobe, Japan; 3) Dept Molecular Biology, Keio University School of Medicine, Tokyo, Japan; 4) Division of Neurogenetics and Bioinformatics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan; 5) Dept Neurology and Eccles Institute of Human Genetics, University of Utah, Salt Lake City, UT, USA; 6) Dept Epigenetics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 7) Dept Geriatric Neuroscience, Tokyo Metropolitan Institute of Gerontology, Itabashi-ku, Tokyo, Japan; 8) Dept Neuropathology, Aichi Medical University, Aichi-gun, Japan; 9) Dept Neurology, University of Tokyo, Tokyo, Japan; 10) Advanced Research Center for Genome Super Power, Keio University, Tsukuba, Japan.

Spinocerebellar ataxia type 31 (SCA31), formerly called chromosome 16q22.1-linked autosomal dominant cerebellar ataxia (OMIM #117210), is an adult-onset neurodegenerative disorder showing progressive cerebellar ataxia mainly affecting Purkinje cells. SCA31 is one of the most common SCAs in Japan. Previous efforts to discover SCA31 mutation had set the critical region to a 900-kb interval in chromosome 16q22.1, where the disease shows a strong founder effect. By performing comprehensive Southern blot analysis and BAC/fosmid-based sequencing, we identified a founder insertion ranging in size from 2.5- to 3.8-kb, consisting of complex penta-nucleotide repeats including a long $(TGGAA)_n$ stretch. In controls, shorter (1.5- to 2.0-kb) insertions lacking $(TGGAA)_n$ were rarely encountered (<0.3% in 1000 chromosomes). We found that the length of insertion inversely correlate with age-of-onset (n=56, r= -0.375, p=0.0044). In addition, an expansion was documented in a single family clinically showing anticipation. The repeat-insertion was located in an intron shared by *TK2* (thymidine kinase 2) and *BEAN* (brain expressed, associated with Nedd4) transcribed in opposite directions in the brain. While the insertion did not appear to cause splicing abnormality in *BEAN* or *TK2*, *in situ* hybridization demonstrated formation of RNA foci in patients' Purkinje cell nuclei with *BEAN*-orientation pentanucleotide repeat transcripts. Because $(TGGAA)_n$ is a characteristic sequence of paracentromeric heterochromatin, we speculate that the insertion may have originated from heterochromatin. SCA31 is a novel human disease associated with "inserted" microsatellite repeats which can expand through transmission.

80

Exome sequencing identifies sequence variants in two genes at the SPG43 locus. *G. Landoure[1,2,3], J.O. Johnson[4], D. Hernandez[4], K.B. Meilleur[5], A. Britton[4], M. Sangare[1,3], C. Rinaldi[1], M. Traoré[3], B. Traynor[4], K.H. Fischebeck[1].* 1) Neurogenetics Branch, NINDS/NIH, Bethesda, MD; 2) Department of Medicine, University College London, London, UK; 3) Service de Neurologie, Hôpital du Point G, Bamako, Mali; 4) Laboratory of Neurogenetics, NIA/NIH, Bethesda, MD; 5) Center for Research on Genomics and Global Health, NHGRI/NIH, Bethesda, MD.

Hereditary spastic paraplegias are inherited neurological disorders characterized by progressive spasticity and weakness, sometimes associated with muscle atrophy. We previously reported a family from Mali (West Africa) with two sisters affected by spastic paraplegia with weakness of the lower limbs and marked atrophy of the distal upper extremity muscles. There was no known consanguinity, but the proportion of identity by descent by SNP analysis was higher than expected, consistent with parental inbreeding. We performed homozygosity mapping and identified a region of extended homozygosity at chromosome 19p13.11-q12 shared by the affected sisters and not by the other unaffected family members or unrelated controls (Meilleur et al, 2010). The region of interest spans 17.7 Mb, and contains about 150 annotated genes. Sequencing of candidate genes in the region, including RAB3A, GDF1, FKBP8, and εCOP, was negative. We then performed genome-wide exome sequencing in one affected individual, and found 3 homozygous single nucleotide variants in 2 genes within the previously identified locus. Two of the sequence variants are located in the same codon (first and second nucleotides). Sequencing of the rest of the family members showed that these changes segregated with the disease in the family. All the sequence variants affect residues that are conserved across species from human to fruit fly, and they are predicted to be deleterious. In addition these sequence variants were not found in 50 ethnically matched controls. Sequencing of more controls is under way to exclude possible polymorphism. Our study highlights the efficiency of exome sequencing by identifying sequence variants in 2 genes at the SPG43 locus, of which one or more may be disease-causing in the family studied here. Additional investigations including functional studies of the candidate gene products are now being done to support our findings.

81

Global Patterns of RNA Editing in Humans. *M. Li[1], Y. Li[2], V.G. Cheung[3,4].* 1) Department of Biostatistics, University of Pennsylvania, Philadelphia, PA; 2) Department of Genetics, University of North Carolina, Chapel Hill, NC; 3) Department of Pediatrics, University of Pennsylvania, Philadelphia, PA; 4) Howard Hughes Medical Institute, University of Pennsylvania, Philadelphia, PA.

RNA editing is a post-transcriptional modification process in which the base of a RNA molecule is altered. It plays an important role in increasing transcriptome diversity. However, our understanding of RNA editing is still limited since only a small number of editing sites have been identified in humans. Here with data from RNA-seq of immortalized B-cells from 27 HapMap CEU individuals, and corresponding DNA-seq data from the 1000 Genomes Project, we showed that RNA editing is very common in humans and the editing sites are not confined to those carried out by ADAR (A-to-I) and APOBEC (C-to-U). Using Illumina Genome Analyzer II, we generated over 40 million reads (50 bp single end) of RNA-seq data per individual. We detected RNA editing sites by comparing genotypes derived from DNA-seq with those from RNA-seq on the same individual. For each individual, an editing event is declared at a site if 1) the number of mapped reads in RNA-seq is ≥20, 2) the number of mapped reads in DNA-seq is ≥4, 3) the individual's genotype at the site is homozygous, and 4) >20% of the mapped reads in RNA-seq show an allele that differs from the individual's (DNA-level) genotype. Applying these criteria, we identified 102,061 autosomal editing events (median number per individual is 3,727), covering 21,289 editing sites located in 4,510 known genes. Many of the editing sites are common, e.g., 57% of the editing sites were seen in two or more individuals. We observed twelve RNA editing forms; the four most common forms are A-to-I, I-to-A, C-to-U, and U-to-C editing. Among the 102,061 editing events, 2,826 are located in the MHC region. Many of these sites likely have biological impacts; for example 30 led to a stop codon and 9 resulted in loss of a stop codon. We are currently validating these findings using Sanger sequencing. Our results also suggest that the more common editing sites (i.e., larger fraction of individuals with an editing event) tend to have higher editing levels (i.e., larger proportion of reads carrying the edited allele). In addition, besides the A-to-I and C-to-U editing of ADAR and APOBEC, we found other editing types thus it is likely that there are other RNA editing enzymes in humans. To our knowledge, our analysis represents the first whole-genome study on RNA editing in humans. The catalog of RNA editing sites will contribute to our understanding of the mechanisms of post-transcriptional modifications.

82

Identification of novel psoriasis susceptibility loci and genetic interaction between *HLA-C* and *ERAP1* provides evidence for an integrated pathogenic pathway. *J. Knight[1,2] for the Genetic Analysis of Psoriasis Consortium and Wellcome Trust Case Control Consortium II.* 1) Dept Med & Molec Gen, King's College London Sch Med, London, United Kingdom; 2) National Institute for Health Research (NIHR), Biomedical Research Centre, Guy's and St. Thomas' NHS Foundation Trust and King's College London, London, UK.

To identify new susceptibility loci for the inflammatory skin disease psoriasis, we undertook a genome-wide association study (GWAS) powered through the analysis of 594,224 SNPs in 2,622 unrelated psoriatic individuals from the British Isles and 5,667 control subjects. In addition to a strong signal at the Major Histocompatibility Complex (MHC; $P < 10^{-200}$), we confirmed association for all eight previously reported non-MHC psoriasis susceptibility loci. We discovered new associations for a further eight independent regions of the genome, of which seven contain genes with recognised immune functions (*IL28RA*, *REL*, *IFIH1*, *ERAP1*, *TRAF31P2*, *NFKBIA* and *TYK2*). We found evidence for allelic heterogeneity at one of these loci, with inclusion of both risk classes substantially increasing estimates of genetic risk for the locus. These findings were confirmed through the analysis of seven replication panels ascertained from different regions of Europe (3,174 psoriasis cases and 5,464 controls). All of the loci generated a combined $P < 5 \times 10^{-8}$. We also report strong evidence for an interaction between the *HLA-C* and *ERAP1* loci (interaction $P < 10^{-4}$ in the discovery dataset and 0.03 in the replication set), with the *ERAP1* disease associated variant only affecting psoriasis susceptibility in individuals carrying the *HLA-C* risk allele. This is one of the first compelling examples of interaction between GWAS loci. Taken together, these findings provide robust genetic evidence to implicate specific pathways that integrate epidermal barrier dysfunction with innate and adaptive immune dysregulation in the pathogenesis of psoriasis.

83

Discovery of a gene for Kabuki syndrome by exome sequencing and genotype-phenotype relationship in 110 cases. *M.C. Hannibal[1], S.B. Ng[2], A.W. Bigham[1], K.J. Buckingham[1], N. Niikawa[3], J.E. Ming[4], D. Donnai[5], C.A. Morris[6], E.H. Zackai[4], T.H. Shaikh[7], S. Banka[5], M. McMillin[1], A.E. Beck[1], H. Gildersleeve[1], H.K. Tabor[1], H.C. Mefford[1], D.A. Nickerson[2], J. Shendure[2], M.J. Bamshad[1].* 1) Dept Pediatrics, University of Washington, Seattle, WA; 2) Dept Genome Sciences, University of Washington, Seattle, WA; 3) Research Institute of Personalized Health Sciences, Health Sciences University of Hokkaido, Hokkaido, Japan; 4) Dept Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, PA; 5) Genetic Medicine, Manchester Academic Health Sciences Centre, University of Manchester, Manchester, England; 6) Dept Pediatrics, University of Nevada School of Medicine, Las Vegas, NV; 7) Dept Pediatrics, University of Colorado, Denver, CO.

Kabuki syndrome is a rare, multiple malformation disorder characterized by a distinctive facial appearance, cardiac anomalies, skeletal abnormalities, and mild to severe mental retardation. Simplex cases make up the vast majority of the 400 or so reported cases with Kabuki syndrome, but parent-to-child transmission in more than a half-dozen instances suggests that it is an autosomal dominant disorder. The relatively low number of cases, the lack of multiplex families and the phenotypic variability of Kabuki syndrome have made the identification of the gene(s) underlying Kabuki syndrome intractable to conventional approaches of gene discovery, despite aggressive efforts. We sequenced the exomes of ten unrelated individuals with Kabuki syndrome and discovered that nine of the ten of them had mutations in a gene that encodes a Trithorax-group histone methyltransferase, a protein which is important in the epigenetic control of active chromatin states. Subsequent screening of an additional 100 families with Kabuki syndrome identified mutations in ~60%. This result is compelling evidence that Kabuki syndrome is genetically heterogeneous. In about one-third of the simplex cases, DNA was available from both parents and mutations were confirmed to be de novo events in these cases. The majority of variants found to cause Kabuki syndrome were novel nonsense or frameshift mutations that are predicted to result in haploinsufficiency. Comparison of the clinical characteristics of mutation-negative and mutation-positive cases allowed us to explore both the relationship between genotype and phenotype and the phenotype of mutation-negative cases. These are important steps toward more accurate predictions about the natural history of individual cases/families as well as for identifying additional genes for Kabuki syndrome.

84

***FBN1* Genetic Variants are Associated with Ascending Aortic Aneurysms and Aortic Dissections.** *M.N. McDonald[1], S.A. LeMaire[2,3], D.G. Guo[4], L. Russell[2,3], C.C Miller, III[3], R.J. Johnson[4], A.L. Estrera[5], H.J. Safi[5], J.S. Coselli[2,3], F. Schoenhoff[6], J.E. Van Eyk[6], M. Bray[7], J.W. Belmont[1], S.M. Leal[1], D.M. Milewicz[3,4], GenTAC Consortium.* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Division of Cardiothoracic Surgery, Michael E. DeBakey Department of Surgery, Baylor College of Medicine, Houston, TX; 3) Cardiovascular Surgery and Medicine Service of the Texas Heart Institute at St. Luke's Episcopal Hospital, Houston, TX; 4) Department of Internal Medicine, University of Texas Health Science Center at Houston, Houston, TX; 5) Cardiothoracic and Vascular Surgery, University of Texas Health Science Center at Houston, Houston, TX; 6) Johns Hopkins Proteomics Center, Johns Hopkins University School of Medicine in Baltimore, Houston, TX; 7) Department of Pediatrics, Baylor College of Medicine, Houston, TX.

Acute aortic dissection is a common cause of premature death that is preventable if the ascending aorta is surgically repaired. Only 1/5 of thoracic aortic aneurysms and dissections (TAAD) are due to rare single gene mutations, such as fibrillin-1 (*FBN1*) mutations in patients with Marfan syndrome. In order to elucidate the common genetic variants predisposing the majority of patients to TAAD, a genome-wide (GW) association study was performed comprised of 765 TAAD patients and two groups of controls (Stage 1). All patients had sporadic ascending aortic aneurysms or aortic dissection. Patients under 31 years of age or who were known to have a syndromic cause of TAAD or a first-degree relative with TAAD were excluded. This analysis lead to the discovery of five SNPs demonstrating increased risk of TAAD within the *FBN1* gene region with genome-wide significance (p-value $< 5 \times 10^{-8}$). These associations were replicated in a separate sample of 448 patients and 161 controls (Stage 2) with a p-value of 1×10^{-5} observed for SNP rs2118181. The combined analysis of the Stage 1 and 2 samples produced a p-value = 3.5×10^{-14} for rs2118181 and odds ratios that ranged from 1.5 to 2.0 for SNPs that met GW significance. Further analysis revealed a haplotype consisting of 5 *FBN1* SNPs associated with increased risk of TAAD (adjusted p-value = 6.9×10^{-10}). Imputation was used to infer genotypes and 80 additional SNPs in the *FBN1* gene region were associated with TAAD with GW significance. Several SNPs in the *FBN1* gene region were also associated with TAAD in patients with and without bicuspid aortic valve. In order to gain additional biological insight, plasma TGF-β1 concentration was measured in patients undergoing ascending aortic aneurysm repair. Plasma TGF-β1 levels were higher in patients with *FBN1* risk haplotypes when compared with patients without risk haplotypes. Additionally, among patients, variants in the *FBN1* gene region were associated with aortic root aneurysms and ascending (Stanford type A) dissections, but not with descending (Stanford type B) dissections. This information may be useful for identifying individuals at risk for life-threatening aortic complications and may lead to broader application of treatment strategies and biomarkers of disease progression that have been developed for patients with Marfan syndrome.

85

DNA methylation and the evolution of gene regulation in primates. *A.A. Pai[1], J.T. Bell[1, 2], J.C. Marioni[1], J.K. Pritchard[1, 2], Y. Gilad[1].* 1) Dept Human Genetics, Univ of Chicago, Chicago, IL; 2) Howard Hughes Medical Institute, Univ of Chicago, Chicago, IL.

Modification of DNA by methylation is an important epigenetic mechanism that affects the spatial and temporal regulation of gene expression. While methylation patterns have been described in many contexts within species (for example, across different tissues), the extent to which these signatures are conserved across species has not been well characterized. To this end, we assayed genome-wide DNA methylation patterns in six livers, six hearts, and six kidneys from multiple humans and chimpanzees, using tissue samples for which genome-wide gene expression data are also available. Using the multi-species gene expression and methylation data for over 7000 genes, we were able to study the evolution of gene regulation in the context of conservation or changes in DNA methylation patterns across tissues and species. Overall, we found that inter-tissue methylation patterns are largely conserved between humans and chimpanzees, and we confirmed that hyper-methylation is correlated with decreased gene expression levels regardless of tissues or species. We identified a large number of genes that show differences in expression levels between tissues or across species, which can be explained, at least in part, by corresponding differences in methylation patterns. In particular, we observed differences in spatial expression patterns between humans and chimpanzee, which are likely to be explained by differences in tissue-specific methylation patterns across the species. By focusing on hemi-methylated genes, we were also able to identify genes that may be imprinted in only one of the species, as well as genes that likely escape X inactivation only in human or chimpanzee females. Finally, by considering genome-wide patterns, we estimated that inter-species differences in methylation patterns might underlie 5-8% of differences in gene expression levels between human and chimpanzee.

86

TBD-Please check addendum for announcement

87

Accurately assessing the risk of schizophrenia conferred by rare copy-number variation (CNVs) affecting genes with brain function. *S. Raychaudhuri[1,2,3], J.M. Korn[1,2,4], S.A. McCarroll[1,5], Intl. Schizophrenia Consortium[10], D.A. Altshuler[1,2,5,6], P. Sklar[7,8,9], S. Purcell[8,9], M.J. Daly[1,2].* 1) Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, 02142 USA; 2) Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, 02114, USA; 3) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Boston, Massachusetts, 02115, USA; 4) Harvard-MIT Health Sciences and Technology, Cambridge, Massachusetts 02139, USA; 5) Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA; 6) Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA; 7) Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts 02114, USA; 8) Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA; 9) Stanley Center for Psychiatric Research, Broad Institute, Cambridge, Massachusetts, 02142 USA; 10) http://pngu.mgh.harvard.edu/isc/.

Purpose: Investigators have linked schizophrenia and other neuropsychiatric disease to rare CNVs; in many cases application of "pathway" analyses of case CNVs have shown enrichment brain-function genes. Such analyses are suggestive, but fail to compare the rates of CNVs impacting brain function genes in cases to controls. They are therefore vulnerable to key confounders such as the large size of brain genes and overall differences in rates and sizes of CNVs. After demonstrating the impact of such confounders, we propose and apply a robust a case-control statistical test, *cnv-enrichment-test*, to compare the rate of CNVs impacting specific gene sets in cases versus controls. Our method is implemented in *PLINK*. **Results:** First, with simulated data we show that *cnv-enrichment-test* is robust to extreme case-control differences in CNV rate and size, and also to large gene size. Then, we examine rare CNV events in 2,415 unaffected meta-controls and also rare CNV events in cases and controls published by the International Schizophrenia Consortium (ISC). We examine four sets of genes with brain function : *brain-expressed* (n=2,531), *neuronal* (n=455), *learning* (n=126), and *synapse* (n=209) genes. To assess the impact of CNVs on these sets, we first apply the standard gene set enrichment approach. The standard gene set enrichment approach to events from 3,391 ISC cases reveals highly significant enrichment for each of the four gene sets. However equally significant enrichment was observed, with similar effect sizes, among rare CNV events from 2,415 meta-controls and also in 3,181 ISC controls. For each set, the observed enrichment is strictly a consequence of the large size of brain function genes. Finally, we apply *cnv-enrichment-test* to recently published rare ISC CNV events. This approach reveals subtle evidence of case-association in *neuronal* (OR=1.18; p=0.04) and the *learning* (OR=1.38; p=0.009) gene sets, but not the other two. **Conclusions:** The current and popular "pathway" analysis approach to assess whether CNVs are enriched for brain functions are confounded by the large size of brain genes and also by case-control differences in CNV size and rate. Instead, we propose a robust case-control statistical approach, the *cnv-enrichment-test*. Application to the ISC data demonstrated nominal enrichment for neuronal-activity genes in schizophrenia case events, with effect sizes that are substantially more modest than previously described.

88

Microdeletions of 3q29 confer high risk for schizophrenia. *J.G. Mulle[1], A.F. Dodd[1], J.A. McGrath[2], P.S. Wolyniec[2], A.A. Mitchell[3], A.C. Shetty[1], N.L. Sobreira[4], D. Valle[4], M.K. Rudd[1], G.A. Satten[1,5], D.J. Cutler[1], A.E. Pulver[2,6], S.T. Warren[1,7].* 1) Dept Human Gen, Emory Univ Sch Med, Atlanta, GA; 2) Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, Maryland 21231, USA; 3) Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York 10029, USA; 4) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland 21231, USA; 5) National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, Georgia 30333, USA; 6) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21231, USA; 7) Departments of Biochemistry and Pediatrics, Emory University School of Medicine, Atlanta, Georgia 30322, USA.

Schizophrenia is a severe psychiatric illness that affects ~1% of the population and has a strong genetic underpinning. Recently, genome wide analysis of copy number variation (CNV) has implicated rare and de novo events as important in schizophrenia, and a handful of specific CNV loci are associated with increased risk for schizophrenia. We report here a deletion at chromosome 3q29, which imparts a substantial risk for schizophrenia. This deletion was detected in a single case that was part of a genome wide study of 245 schizophrenia cases and 490 controls, all of Ashkenazi Jewish descent. Consistent with other studies, we observed a 2.8-fold increase in large (>500 Kb) and rare (absent in the Database of Genomic Variants; DGV) CNVs in cases relative to controls. Of these CNVs in cases, approximately 57% were de novo. We focused on one 836 kb de novo deletion at 3q29 that falls within a 1.3 - 1.6 Mb deletion previously identified in children with intellectual disability (ID) and autism. By combining our data with prior CNV studies in schizophrenia and analysis of the data of the Genetic Association Information Network (GAIN), we identified six 3q29 deletions among 7,545 schizophrenic subjects and one among 39,748 controls. This results in a statistically significant association with schizophrenia (p = 0.0097) and an odds ratio estimate of 16.98 (95% CI: 1.36 - 1198.4). Moreover this 3q29 deletion region contains two linkage peaks from prior schizophrenia family studies and the minimal deletion interval contains 19 annotated genes, including PAK2, paralogous to an X-linked ID gene and now a strong candidate for schizophrenia susceptibility.

89

Rare copy number variants in the ANXA1 gene in patients with autism spectrum disorders. *C. Correia[1, 2, 3], A.F. Segueira[1, 2, 3], J. Almeida[4], C. Café[4], S. Mouga[4], F. Duque[4], T. Magalhaes[1, 2, 3], D. Pinto[5], W. Roberts[6], S.W. Scherrer[5], G. Oliveira[4], A.M. Vicente[1, 2, 3], Autism Genome Project Consortium.* 1) DPSDC, Inst Nac Saude Dr Ricardo Jorge, Lisbon, Portugal; 2) Instituto Gulbenkian de Ciência, Oeiras, Portugal; 3) Center for Biodiversity, Functional & Integrative Genomics, Lisboa, Portugal; 4) Centro de Desenvolvimento da Criança, Hospital Pediátrico de Coimbra, Coimbra, Portugal; 5) The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children, Toronto, Ontario M5G 1L7, Canada; 6) Autism Research Unit, The Hospital for Sick Children and Bloorview Kids Rehab, University of Toronto, Toronto, Ontario M5G 1X8, Canada.

Recent studies suggest that copy number variation (CNV) in the human genome is extensive and may play an important role in susceptibility to neuropsychiatric disorders such as autism. We report here on four unrelated autistic patients with a small duplication involving the last 3-4 exons of the ANXA1 gene located at 9q21.13. The duplications were identified in the whole genome CNV analysis of 1,275 individuals diagnosed with Autism Spectrum Disorder (ASD), carried out by the Autism Genome Project Consortium using the Illumina Infinium1M SNP microarray. The identified duplications were validated by quantitative PCR, and were not present in 4964 control subjects of European ancestry with no obvious psychiatric disease history. The clinical presentation of these patients (three males and one female) was heterogeneous. The male patients had mild to moderate intellectual disability and presented with language delay, while the female patient had a normal IQ and no language impairment. Dysmorphic features were absent in all patients. All patients had a positive family history of neuropsychiatric disorder, with cases of Asperger Syndrome, schizophrenia and depression among first or higher degree relatives. In two families the CNV was maternally inherited while in the other two it was paternally inherited. In the two families where parental evaluation was performed the transmitting parent scored positive on the Social Responsiveness Scale and in the Personality Styles and Preferences Questionnaire. The ANXA1 gene encodes an effector of the mTOR (mammalian Target Of Rapamycin) pathway, a major regulator of cellular growth in mitotic cells, and mutations in several negative effectors of this pathway have been described in autistic individuals. This gene encodes annexin 1, an anti-inflammatory protein likely to play an important role in neuroinflammation. Interestingly, there is accumulating evidence implicating an active ongoing chronic neuroinflammatory process in autism, suggested also in our population by the observation of a widespread occurrence of autoreactivities to brain tissue in autistic patients. Transmission disequilibrium analysis of the 107 ANXA1 SNPs represented in the SNP microarray showed a trend towards association for two SNPs (P=0.0006 and P=0.007 for rs10512015 and rs1342022, respectively). The overall results implicate rare CNVs in the ANXA1 gene in autism risk, suggesting that this gene and the mTOR pathway should be further explored.

90

Increased gene copy number in autism is associated with specific behavioural phenotypes in the mouse. *C. Webber[1], D. Pinto[2], C.R. Marshall[2], H.J. Noh[1], H. Boulding[1], S.W. Scherer[2], C.P. Ponting[1].* 1) MRC Functional Genomics Unit, Oxford University, Oxford, United Kingdom, OX1 3QX, UK; 2) The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children and Department of Molecular Genetics, University of Toronto, Ontario, M5G 1L7, Canada.

Autism Spectrum Disorders (ASDs) are highly heritable (~90%) and characterised by impairments in reciprocal social interaction and communication, along with restricted and repetitive interests and behaviours. As only a small percentage of individuals with autism have children, the underlying genetic variants are likely to be at a lower frequency in the population than variants underlying other neuropsychiatric disorders. Thus, *de novo* copy number variants (CNVs) identified in ASD cases present a particularly good source for the genetic variation underlying ASDs. We examined 3 non-overlapping sets of *de novo* CNVs identified in individuals with autism; (*i*) a large set of 73 CNVs from 54 patients identified by the Autism Genome Project (AGP) consortium, (*ii*) a smaller set of 28 CNVs from 24 patients identified by Marshall *et al*. and (*iii*) a small set of 24 CNVs identified within 23 patients from the DECIPHER database. We extended and applied our methodology that robustly associates specific phenotypes resulting from the targeted disruption of genes in the mouse ("knock-out" phenotypes) to intervals in the human genome associated with disease. Using conservative statistics, associations to mouse phenotypes were discovered (FDR< 5%) in the powerful AGP CNV set and then either replicated (FDR < 5%) or validated (single test, $p < 0.05$) in the smaller Marshall and/or Decipher CNV sets. Among genes overlapped by both *gain* (duplication) and *loss* (deletion) ASD CNVs, we identified and replicated a significant enrichment of genes whose orthologues' disruption in the mouse yields an *abnormal CNS synaptic transmission* phenotype. However, exclusive to the *gain* CNVs, we identified many broad phenotypic enrichments, including *abnormal behavior* and *abnormal learning*. Other *gain*-specific phenotypic enrichments appear as readily comparable to primary or associated features of ASDs including (*i*) *abnormal spatial learning*, *impaired coordination* (results replicated): 60-80% of all ASD cases exhibit poor motor planning and coordination, (*ii*) *non-convulsive seizures* (result replicated):~1/3rd of all ASD cases suffer seizures and these are frequently non-convulsive, (*iii*) *impaired hearing* and *sensorineural hearing loss* (result validated): 25-40% of all ASD cases have hearing impairments. Together, these enrichments offer causal hypotheses for 19-83% of patients within each set and identify 107 largely novel candidate genes for ASDs and frequently co-occurring symptoms.

91

Rare Structural Variation of Synapse and Neurotransmission Genes in Autism. *P.S. White[1,2,3], H.M. Xie[1], J.C. Perin[1], N. Takahashi[4], K. Murphy[1], A.S. Wenocur[1], M. D'arcy[1], R.J. O'Hara[1], E. Goldmuntz[3,5], D.E. Grice[6], T.H. Shaikh[7], H. Hakonarson[3,8,9], J.D. Buxbaum[4], J. Elia[10,11], X. Gai[1].* 1) Center for Biomedical Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Oncology, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA; 4) Seaver Autism Center and Department of Psychiatry, Mt. Sinai School of Medicine, New York, NY; 5) Division of Cardiology, The Children's Hospital of Philadelphia, Philadelphia, PA; 6) Department of Child and Adolescent Psychiatry, Columbia University, New York, NY; 7) Department of Pediatrics, University of Colorado School of Medicine, Denver, CO; 8) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 9) Division of Pulmonary Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA; 10) Department of Child and Adolescent Psychiatry, The Children's Hospital of Philadelphia, Philadelphia, PA; 11) Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA.

Autism spectrum disorders (ASDs) comprise a constellation of highly heritable neuropsychiatric disorders. Genome-wide studies of autistic individuals have implicated numerous minor risk alleles but few common variants, suggesting a complex genetic model with many contributing loci. To assess commonality of biological function amongst rare risk alleles, we compared functional knowledge of genes overlapping inherited structural variants in a discovery cohort of 631 idiopathic ASD subjects and 1,162 parents relative to 1,775 matched controls. Biological processes associated with synapse function and neurotransmission were significantly enriched in ASD subjects versus controls. Independent analysis was performed on a replication cohort comprising 593 idiopathic ASD subjects and 1,109 parents relative to 2,026 matched controls. Fifteen functional attributes enriched in ASD subjects replicated with corrected significance; almost all of the replicated functions were associated with synaptic neurotransmission. Enriched functions included cell-cell signaling (p=6.14x10-7), transmission of nerve impulse (p= 4.98x10-6), and synaptic transmission (p=4.98x10-6). Replicated enriched processes were generally implicated by CNVs disrupting different subsets of genes in the discovery and replication cohorts. Analysis of phenotypes observed for mouse models of copy-variant genes established significant and replicated enrichment of observable phenotypes consistent with ASD behaviors. Most functions retained significance after removal of CNVs overlapping previously reported autism candidate genes. Twelve genes were enriched both by functional and mouse phenotype analyses. These genes, including three prior autism candidate genes (*NRXN1*, *PARK2*, *DOC2A*) and nine novel candidates, were found to be preferentially expressed in human temporal cortex (6 genes), associated with synaptic function (12 genes), and associated with glutamatergic signaling (6 genes). Taken together, these results provide for the first time statistical evidence that synaptic function and glutamate-mediated neurotransmission processes contribute to ASD pathophysiology and strongly support a rare variant model for ASD risk.

92

Exonic deletions of *NRXN1* are associated with muscular hypotonia, autism spectrum disorder and developmental delay. *S. Cheung[1,2], A. Patel[1,2], P.M. Boone[2], N. Tartaglia[3], N. Madduri[4], G. Simpson[5], J. Gibson[5], C.D. Luzzi[6], H.P. Crawford[7], P. Evans[8], J.A. Phalen[9], W. Campbell[3], C.W. Brown[4], P. Stankiewicz[2], C.P. Schaaf[2].* 1) Molecular Genetics Laboratories, Baylor College of Medicine, Houston, TX; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Department of Pediatrics, University of Colorado, Denver, CO; 4) Department of Pediatrics, Vanderbilt University, Nashville, TN; 5) 'Specially for Children, Austin, TX; 6) Memorial Medical Group, Saint Joseph, MI; 7) Division of Medical Genetics, University of Texas Medical School, Houston, TX; 8) Department of Pediatrics, University of Texas, Southwestern School of Medicine, Dallas, TX; 9) Developmental Pediatric Services WHMC, Lackland AFB, TX.

Neurexins are a family of neuronal cell adhesion molecules that have been shown to play an integral role in synaptogenesis and synaptic maintenance. Mammalian neurexins are encoded by the *NRXN1*, *NRXN2*, and *NRXN3* genes. By use of alternate promoters and extensive alternative splicing, several hundreds of different neurexin isoforms are expressed from these three genetic loci. Genetic variants and chromosomal aberrations, including copy number variants, whole gene and exonic deletions, as well as translocations disrupting the *NRXN1* gene have been associated to various developmental disorders and neuro-psychiatric phenotypes. These include autism spectrum disorders, developmental delay and intellectual disability, nicotine dependence and schizophrenia. We present detailed clinical data of nine index patients, ages 5 months to 6 years old, with exonic deletions of the *NRXN1* gene. Cases were ascertained from a total of 4500 individuals referred for exon-targeted array comparative genomic hybridization (aCGH) from June 2009 to March 2010. In our cohort, exonic deletions of *NRXN1* cause distinct clinical phenotypes with muscular hypotonia and autism spectrum disorder being recognized as the two most common clinical features. All individuals identified with heterozygous deletions involving exons 1 and 2 of *NRXN1* (coding for neurexin-1α) presented with significant infantile muscular hypotonia. Autism spectrum disorder and pervasive developmental disorder (PDD-NOS) represent the second most common finding in our cohort (83%). Developmental delay and intellectual disability are common, but vary greatly in severity (with developmental quotients ranging from 30s to 80s). Epilepsy appears to be uncommon among patients with *NRXN1* deletions (only one patient in our cohort), as are structural brain malformations (only one patient with mild subcortical hypomyelination). In summary, we report exonic deletions of *NRXN1* as one of the most common copy number variants among individuals referred for muscular hypotonia and/or autism spectrum disorder. aCGH with exon coverage has the advantage of detecting single and/or multiple exon deletions of disease relevant genes such as the *NRXN1* gene.

93

Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. *D. Moreno-De-Luca[1], J.G. Mulle[1], E.B. Kaminsky[1], S.J. Sanders[2], S.M. Myers[3], M.P. Adam[1], A.T. Pakula[4], N.J. Eisenhauer[3], K. Uhas[1], L. Weik[5], L. Guy[4], M.E. Care[6], C.F. Morel[6], C. Boni[3], B. Salbert[3], A. Chandrareddy[7], L.A. Demmer[7], E.W.C. Chow[8], U. Surti[9], S. Aradhya[10], W.G. Sanger[11], A.R. Brothman[12], E.C. Thorland[13], R. Iyer[14], J.C. Barber[15], J.A. Crolla[15], S.T. Warren[1, 16], C.L. Martin[1], D.H. Ledbetter[1], SGENE Consortium, Simons Simplex Collection Genetics Consortium, GeneSTAR.* 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, USA; 2) Child Study Center, Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA; 3) Geisinger Medical Center, Danville, Pennsylvania, USA; 4) Marcus Autism Center, Department of Pediatrics, Emory University School of Medicine, Atlanta, Georgia, USA; 5) Children's Hospital of Wisconsin, Milwaukee, Wisconsin, USA; 6) Fred A. Litwin Family Centre in Genetic Medicine, University Health Network & Mount Sinai Hospital, Toronto, Ontario, Canada; 7) Division of Genetics, Tufts Medical Center, Boston, Massachusetts, USA; 8) Centre for Addiction and Mental Health, University of Toronto, Toronto, Ontario, Canada; 9) University of Pittsburgh, School of Medicine, Pittsburgh, Pennsylvania, USA; 10) GeneDx, Gaithersburg, Maryland, USA; 11) Human Genetics Laboratory, University of Nebraska Medical Center, Omaha, Nebraska, USA; 12) ARUP Laboratories, University of Utah, Salt Lake City, Utah, USA; 13) Mayo Clinic, Rochester, Minnesota, USA; 14) Michigan Medical Genetics Laboratories, Ann Arbor, Michigan, USA; 15) Salisbury NHS Foundation Trust, Wessex Regional Genetics Laboratory, Salisbury, UK; 16) Departments of Pediatrics and Biochemistry, Emory University School of Medicine, Atlanta, Georgia, USA.

Autism spectrum disorders (ASDs) and schizophrenia are very heterogeneous neurodevelopmental disorders with a major genetic component, including a growing list of rare, de novo single gene mutations and copy number variants (CNVs). This substantial etiologic heterogeneity resembles intellectual disability, where many different rare mutations account for the majority of cases with a known genetic cause. Recent studies suggest that autism and schizophrenia share similar genetic pathways; however, the etiology remains unknown in most patients. We used whole-genome cytogenetic array analysis to search for pathogenic CNVs in a discovery sample of patients with neurodevelopmental disorders referred for clinical testing. We detected a recurrent 1.4 Mb deletion at 17q12 in 18/15,749 patients, including several with autism, but 0/4,519 controls. This region harbors the HNF1B gene, mutations in which cause maturity onset diabetes of the your type 5 (MODY5) as well as renal cysts and diabetes syndrome (RCAD). Based on these initial findings, we extended our analysis to a large follow up sample, consisting of two autism and two schizophrenia collections, to specifically assess the occurrence of the recurrent 17q12 deletion. In follow up autism samples, we found the identical deletion in 2/1,182 patients and 0/38,498 controls (Corrected OR ¥, 4.73-¥, p=1.92x10-3). In follow up schizophrenia samples, we identified this CNV in 4/6,340 patients and 0/43,076 controls (Corrected OR ¥, 1.25-¥, p=1.47x10-2). Combining the four follow up collections showed a highly significant difference in frequency of the 17q12 deletion between neurodevelopmental cases and controls (Corrected OR ¥, 5.99-¥, p=7.37x10-5). Altogether, the 17q12 deletion was present in 24/23,274 cases but was never observed in 52,449 controls, providing compelling evidence for its pathogenic role in neurodevelopmental disorders. In conclusion, deletion 17q12 is a recurrent, pathogenic CNV that confers a very high risk for autism, schizophrenia and neurodevelopmental disorders. These data indicate that one or more of the 15 genes in the deleted interval is dosage sensitive and essential for normal brain development and function.

94

Genome Wide Copy Number Variation Study Associates Metabotropic Glutamate Receptor Genes with Attention Deficit Hyperactivity Disorder. *H. Hakonarson[1], J.T. Glessner[2], K. Wang[2], N. Takahashi[3], C.J. Shtir[4], P.M.A. Sleiman[2], H. Zhang[2], C.E. Kim[2], G.J. Lyon[5], J.H. Flory[2], J.P. Bradfield[2], M. Imielinski[2], C. Hou[2], E.C Frackelton[2], F. Middleton[6], A.A. Todorov[7], A. Reif[8], B. Franke[9], K.P. Lesch[8], R. Anney[10], P. Shaw[11], M. Devoto[12,13,14,15], S.F.A. Grant[2,12,13], P. White[16,17], J.D. Buxbaum[3], J.L. Rapoport[11], N.M. Williams[18], S.F. Nelson[4], S.V. Faraone[6], J. Elia[19,20], ADHD GWAS Consortium.* 1) Pediatrics/Gen, Children's Hosp Philadelphia, Philadelphia, PA; 2) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA; 3) Laboratory of Molecular Neuropsychiatry, Department of Psychiatry, Mount Sinai School of Medicine, New York, NY 10029 USA; 4) Department of Human Genetics and Neuroscience, University of California - Los Angeles, Los Angeles, CA 90095, USA; 5) Department of Psychiatry, University of Utah, Salt Lake City, Utah, 84112; 6) Departments of Psychiatry and of Neuroscience and Physiology, State University of New York Upstate Medical University, Syracuse, New York, NY 13210, USA; 7) Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri 63110, USA; 8) ADHD Clinical Research Network, Unit of Molecular Psychiatry, Department of Psychiatry, Psychosomatics, and Psychotherapy, University of Wuerzburg, Wuerzburg, 97080 Germany; 9) Departments of Human Genetics and Psychiatry, Radboud University Nijmegen Medical Centre, Nijmegen, 6500, The Netherlands; 10) Department of Psychiatry, Trinity Centre for Health Sciences, St. James's Hospital, Dublin 8, Ireland; 11) Child Psychiatry Branch, National Institute of Mental Health, Bethesda, MD, 20892, USA; 12) Division of Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA; 13) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA, 19104, USA; 14) Dipartimento di Medicina Sperimentale, University La Sapienza, 00185 Rome, Italy; 15) Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA; 16) Center for Biomedical Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA; 17) Division of Oncology, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA; 18) Department of Psychological Medicine, School of Medicine, Cardiff University, Cardiff, Wales, CF10 3xQ, UK; 19) Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA, 19104, USA; 20) Department of Child and Adolescent Psychiatry, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA.

Attention-Deficit, Hyperactivity Disorder (ADHD) is a common, heritable neuropsychiatric disorder of unknown etiology. Recently, we reported an enrichment of rare variants in genes involved in learning, behavior, synaptic transmission and central nervous system development in autism, suggesting that rare inherited structural variants could also play a role in the etiology of ADHD, a related neuropsychiatric disorder. We performed a whole-genome CNV study in a cohort of 1,013 ADHD cases and 4,105 healthy children of European ancestry who were genotyped with 550,000 SNP markers. Positive findings were evaluated in multiple independent cohorts, totaling 2,493 ADHD cases and 9,222 controls of European ancestry, with respective case-control cohorts genotyped on matched platforms. CNVs impacting metabotropic glutamate receptor genes were significantly enriched across all independent cohorts (P= 2.1x10-9). Among them, deletions in GRM5 (glutamate receptor, metabotropic 5) occurred in ten cases across three independent cohorts and in only one control subject (P=1.36x10-6). In addition, deletions in GRM7 occurred in six cases and GRM8 in eight cases, both with a control frequency of zero. GRM1 was duplicated in eight cases, a frequency notably enriched above controls. Observed variants were experimentally validated using quantitative PCR. We have identified several rare recurrent CNVs that are overrepresented in multiple independent ADHD cohorts that impact genes involved in glutamatergic neurotransmission, an important mediator for the developing brain and normal brain function. These results suggest that variations involving glutamatergic gene networks of the brain contribute to the genetic susceptibility to ADHD.

95

Array CGH reveals pathogenic copy number changes in ~10% of patients with severe epilepsy syndromes. *H.C. Mefford[1], C. Hsu[1], S.C. Yendle[2], E.E. Eichler[3,4], S.F. Berkovic[2], I.E. Scheffer[2].* 1) Pediatrics, University of Washington, Seattle, WA; 2) Epilepsy Research Center and Department of Medicine, Univ of Melbourne, Australia; 3) Genome Sciences, University of Washington, Seattle, WA; 4) Howard Hughes Medical Institute, University of Washington, Seattle, WA.

Severe epilepsies, including the epileptic encephalopathies are often associated with deterioration in cognitive, sensory and motor function. Although a few genes, such as *SCN1A*, have been identified, the genetic etiology of most severe, early onset epilepsies is unknown. Our past work demonstrated that copy number changes, including deletions of 15q11, 15q13 and 16p13, contribute to the genetic etiology of idiopathic generalized epilepsy. To identify novel genes and regions for more severe epilepsies, we performed array comparative genomic hybridization in 300 individuals with severe epilepsy syndromes with or without brain malformations using a commercially available, exon-focused oligonucleotide array with 720,000 probes. We found 27 (9%) rare deletions (n=21) and duplications (n=6) that have not been previously reported in 2493 controls that are likely to be causative. Examples include *de novo* deletions of 1p36, 1q44, 15q11 (single exon of *UBE3A*), 15q26 and Xp22 (*CDKL5* gene). We identified a homozygous, single-exon deletion in the *CNTNAP2* gene in two affected siblings, and overlapping deletions of 7q21 (six genes in common) in two unrelated patients with similar phenotypes. Rearrangements at "hotspot" regions flanked by segmental duplications in our cohort include deletion of 15q13.3 between BP3-BP4, deletion of distal 16p11.2 (28.7-29.0 Mb) and duplication of proximal 16p11.2 (29.5-30.1 Mb). In contrast to our previous studies in which deletions of 15q13.3, 16p13.11 and 15q11.2 were frequently found in patients with idiopathic generalized epilepsies, we did not find these deletions in our cohort of patients with more severe epilepsy syndromes. In summary, we report copy number changes in the first large study of severe epilepsy syndromes. Nearly 10% of patients harbor rare rearrangements that are likely to be causative, and sequence analysis of candidate genes within these regions is likely to reveal causative mutations in additional patients.

96

Genomic and intragenic deletions of the *STXBP1* gene within the 9q34.11 chromosome region are associated with progressive early infantile epileptic encephalopathy, mental retardation and autism. *S.A. Yatsenko[1], T. Reimschisel[2], M. Thomas[3], W. Wilson[3], J.W. Wheless[4], J.A. Rosenfeld[5], B.A. Bejjani[5], L.G. Shaffer[5], S.W. Cheung[1], P. Stankiewicz[1], J.R. Lupski[1,6,7], F. Scaglia[1,7].* 1) Molec & Human Gen, Baylor Col Med, Houston, TX; 2) Pediatrics, Vanderbilt University Medical Center, Nashville, TN;; 3) Pediatrics, University of Virginia School of Medicine, Charlottesville, VA;; 4) Pediatric Neurology, University of Tennessee Health Science Center; Memphis, TN;; 5) Signature Genomic Laboratories, Spokane, WA;; 6) Pediatrics, Baylor College of Medicine; Houston, TX; 7) Texas Children's Hospital, Houston, TX.

Using microarray analysis, we have identified seven patients with overlapping interstitial microdeletions in 9q34.11. In six patients with epilepsy, developmental regression and mental retardation the common deleted region encompasses the *STXBP1* gene, including one patient with an intragenic *STXBP1* deletion involving exons 1-4. SXTBP1, syntaxin binding protein 1, is highly expressed in brain, implicated in vesicle trafficking and neurotransmitter release and has been shown recently to be associated with infantile epileptic encephalopathy type 4 (Saitsu et al. 2008, Nat Genet 40:782-8). Binding of STXBP1 regulates SNARE complex formation and subsequent synaptic vesicle fusion. Previously, nonsyndromic epilepsy and mental retardation have been observed in all of the studied patients with *de novo* mutations of the *STXBP1* gene. To date only one patient has been described with 9q34.11 deletion involving the *STXBP1* gene. To determine the size, genomic extent and gene content and to establish genomotype-phenotype correlations in our cohort, we performed custom high-resolution 9q34 array CGH in each patient. The 9q34.11 genomic deletions range from 50 kb to 2 Mb in size with unique locations for the breakpoints in each patient. Interestingly, in four of seven patients the deletion also included the *ENG* (endoglin) gene, defects in which are the cause of hereditary hemorrhagic telangiectasia type 1 (HHT1). At least one of our patients with *ENG* deletion has been diagnosed with pulmonary arteriovenous malformations. Our results provide further evidence that haploinsufficiency for *STXBP1* results in progressive neurodegenerative epileptic encephalopathy, and genomic deletions involving *ENG* may be responsible for multisystemic vascular dysplasia, thus revealing a cis genetic affect or a possible contiguous gene deletion syndrome. Patients with microdeletions in the 9q34.11 region may present with a complex phenotype due to this contiguous gene deletion.

97

A novel approach identifies new differentially methylated regions (DMRs) associated with imprinted genes. *S. Choufani[1], JS. Shapiro[1,2], M. Susiarjo[3], DT. Butcher[1], D. Grafodatskaya[1], JC. Ferreira[1,2], D. Pinto[1], SW. Scherer[1,4], LG. Shaffer[5], P. Coullin[6], I. Caniggia[7], J. Beyene[8], R. Slim[9], M. Bartolomei[3], R. Weksberg[1,2,10].* 1) Dept Genetics & Genomic Biol, Hosp Sick Children, Toronto, ON, Canada; 2) Institute of Medical Sciences, University of Toronto, ON, Canada; 3) Dept. of Cell and Developmental Biology, U. of Pennsylvania School of Medicine, Philadelphia, PA; 4) Dept. of Molecular Genetics, U. of Toronto, Toronto, ON, Canada; 5) Signature Genomic Laboratories, Spokane, DC; 6) Endocrinologie et Génétique de la Reproduction et du Developpement, Clamart, France; 7) Prog. in Fetal Health and Development, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON, Canada; 8) The Child Health Evaluative Sciences Program and the Departments of Health Policy, Management and Evaluation, The Hospital For Sick Children; and the Dala Lana School of Public Health, U. of Toronto, Toronto, ON, Canada; 9) Dept. of Human Genetics and Obstetrics-Gynecology, McGill University Health Center, Montreal, QC, Canada; 10) Division of Clinical and Metabolic Genetics, Hosp Sick Children, Toronto, ON, Canada.

Imprinted genes are critical for normal human growth and neurodevelopment. Most imprinted genes are clustered and respond to signals in cis from nearby differentially methylated regions (DMRs). Such DMRs, comprised of CpG-rich regions of DNA, demonstrate ~50% methylation because either the maternal or the paternal chromosomes (alleles) are methylated in a parent of origin specific manner. We developed a strategy whereby DNA differentially methylated regions were used to identify novel human imprinted genes. We compared genome-wide methylation profiles of bisulfite modified genomic DNA from 21 normal biparental tissues (placenta and blood) and 4 uniparental tissues and hybridized it to the Illumina methylation 27 array. This array covers ~27,600 CpG sites (~14,500 genes). Candidate DMRs were selected by identifying CpGs with putative allelic differential methylation in normal biparental tissues. In parallel, we looked for parent of origin-specific patterns in DNA methylation in paternally derived human androgenetic complete hydatidiform mole (AnCHM), and maternally derived mature cystic ovarian teratoma (MCT). Using this approach, we selected known DMRs associated with imprinted genomic regions. Furthermore, our analysis revealed previously unidentified DMRs for the known imprinted genes, NAP1L5 and ZNF597. Most importantly, several new candidate imprinted genes were identified. The DMR for one novel candidate, AXL, is validated by methylation studies in human. In mouse embryos, we showed that Axl is expressed preferentially from the maternal allele. This parent of origin -specific expression pattern is shown to be DNA methylation -dependent. This approach permitted the characterization of a new candidate imprinted gene, AXL, in humans. This is one of the few instances of a new imprinted gene being identified first in humans and then in mice.

98

Mapping Gene Regulatory Elements for Brain Expressed Genes Using Next Generation Sequencing. *C. Barr[1,2], Y. Feng[1].* 1) Genetics & Development Division, Toronto Western Hospital, UHN, Toronto, ON, Canada; 2) Program in Neurosciences and Mental Health, The Hospital for Sick Children. Toronto, ON, Canada.

Introduction: Current evidence indicates that psychiatric and neurological disorders are more likely to result from genetic variation that changes gene expression rather than variations in the protein coding region of the gene. However, the role of variation in gene regulation as a contributor to complex traits disorders has been largely ignored because of the difficulty in identifying the location of gene regulatory elements outside of the proximal promoter. Methods: We used a genome-wide approach to identify brain-relevant gene regulatory elements, in two cell lines, a neuroblastoma and a glioblastoma, and in human brain tissue. We used chromatin immunoprecipitation (ChIP) to modified histones and transcription factors combined with high throughput sequencing (ChIP-sequencing) to identify the position of brain-relevant regulatory elements across the genome. To maximize the identification of regulatory elements, we performed ChIP using antibodies to acetylated and methylated histones, (markers of active chromatin), the enhancer-associated protein p300 (a marker of enhancers), and RNA polymerase II (a marker of regions of active transcription). Results: This approach was very successful, and we now have genomic maps of putative regulatory elements for genes expressed in these cell lines including genes associated with psychiatric disorders (e.g. DISC1, NTRK3, DTNBP1, SNAP25). We also mapped putative gene regulatory elements in gene “deserts” in the region of positive markers from the genome wide association studies. We are currently screening the putative regulatory regions in associated genes for genetic variation and testing the relationship to the respective psychiatric disorder in DNA from families. Conclusions: The use of ChIP combined with high throughput sequencing is a powerful approach to map gene regulatory elements allowing gene findings for psychiatric disorders to move forward to functional studies. Further, understanding how risk genes are differentially regulated is crucial for future therapeutic interventions and the understanding of genetic variation in response to these interventions.

99

Altered promoter proximal RNAPII activity at stimulus responsive transcripts in the absence of MeCP2. *KE. Szulwach, X. Li, AC. Shetty, ME. Zwick, P. Jin.* Department of Human Genetics, Emory University School of Medicine, Atlanta, GA.

Promoter proximal regulation of transcription is thought to reflect poising of targeted genes for rapid responses to environmental cues and is widespread in mammalian genomes. Such regulation is influenced by chromatin at transcription start sites (TSSs), implicating epigenetic processes in stimulus dependent transcription. MeCP2 is an activity dependent DNA methyl-CpG binding protein in neurons, and mutations in *MECP2* cause the neurodevelopmental disorder Rett Syndrome. We have assessed the effect of MeCP2 on promoter proximal RNAPII activity by sequencing >80 million small RNA and identifying those associated with TSSs in wildtype and *MeCP2-/y* mouse neural stem cells. 304 RefSeq transcripts exhibited differential small RNA expression at TSSs in MeCP2 deficient neural stem cells. Among these genes was *Pcdh10*, a protocadherin implicated in Autism-spectrum disorders that is also activity dependent in neurons. In MeCP2 deficient neural stem cells, increased TSS associated small expression at *Pcdh10* correlated with an equivalent increase in mRNA. A disproportionate increase in RNAPII occupancy relative to expression of *Pcdh10* in the absence of MeCP2 indicated altered promoter proximal regulation of this gene. MeCP2 directly interacted near to the *Pcdh10* TSS and loss of MeCP2 correlated with an altered chromatin state reflective of *Pcdh10* expression. These results suggest a novel mechanism by which MeCP2 may influence stimulus dependent transcription of targeted genes through regulation of promoter proximal RNAPII activity.

100

Intronic methylation, microRNA and testicular germ cell tumor. *H.H. Cheung[1,2], A.J. Davis[1], T.L. Lee[1], A.L.Y. Pang[1], S. Nagrani[1], O.M. Rennert[1], W.Y. Chan[1,2].* 1) Lab Clin Devel Genomics, NICHD, NIH, Bethesda, MD; 2) School of Biomedical Sciences, Faculty of Medicine, Chinese University of Hong Kong, Shatin, Hong Kong SAR, China.

DNA methylation is an important epigenetic modification that regulates normal development and diseases, including cancer metastasis. In a genome-wide methylation profiling study, we identified a hypermethylated region in nonseminomatous TGCT cell lines. This region maps to a conserved region in intron-14 of dynamin 3, located at 1q24.3. The hypermethylated region coincides with the promoter of miR-199a-2. miR-199a was previously implicated in the progression and prognosis of gastric and ovarian cancers. Bisulfite sequencing and realtime RT-PCR studies of testicular tumor germ cell tumor (TGCT) cell lines and patient tissues revealed hypermethylation of miR-199a in malignant TGCT and correlation of its downregulation with aggressiveness of the tumors. Therefore epigenetic silencing of miR-199a may be responsible for the invasiveness of the tumors. To elucidate the role of miR-199a in TGCT, we constitutively express miR-199a in a malignant testicular embryonal carcinoma cell line Ntera2. Expression of miR-199a in Ntera2 suppresses cancer invasion as revealed by cell proliferation, metri-gel, migration, cell adhesion, and in vivo tumor growth studies. miR-199a also suppresses development of metastasized tumors in SCID mice, further suggesting an anti-metastasis role of this microRNA. To understand the action of miR-199a in regulating metastasis, we screened for its potential targets by expression microarray and confirmed the target by Western blot and luciferase reporter assay. PODXL was identified as a target of miR-199a. PODXL is an anti-adhesion protein that is upregulated in many aggressive tumors, including TGCT. Using tissue array we showed PODXL is upregulated in both malignant seminoma and nonseminoma. Expression of PODXL is negatively correlated with miR-199a level, which is indirectly controlled by DNA methylation. To demonstrate that miR-199a suppresses cancer invasion through PODXL, we knocked down PODXL by RNAi without boosting the level of miR-199a. PODXL knockdown is found to be able to suppress cancer invasion. This study reveals the role of DNA methylation in silencing of miR-199a that normally suppresses cancer invasion through inhibiting an anti-adhesion protein PODXL.

101

Genome-wide DNA methylation profiling in lupus CD4 T cells identifies differentially methylated genes and novel candidate loci for lupus. *T. Hughes[1], J. Wren[1], Y. Tang[2], Y. Fei[1,3], J. Merrill[1], R. Webb[1], A. Sawalha[1,3,4].* 1) Oklahoma Medical Research Foundation, Oklahoma City, OK; 2) Samuel Roberts Noble Foundation, Ardmore, OK; 3) University of Oklahoma Health Sciences Center, Oklahoma City, OK; 4) US Department of Veterans Affairs Medical Center, Oklahoma City, OK.

Objectives: T cell DNA methylation defect correlates with disease activity in lupus patients. We performed a whole-genome methylation scan in CD4 T cells to identify differentially methylated genes in lupus patients and determine the functional implication of differentially methylated loci upon gene expression. Methods: We used DNA immunoprecipitation with anti-5-methylcytidine antibody coupled with microarray hybridization to determine DNA methylation patterns in CD4 T cells from lupus patients and controls. Input and immunoprecipitated DNA were differentially labeled and hybridized to arrays with ~ 385,000 probes, covering all UCSC-annotated CpG islands and promoter regions for all RefSeq genes. Lupus CD4 T cell gene expression data were available from Gene Expression Omnibus. Results: We identified genes differentially methylated within the 5kb upstream to 5kb downstream of the transcription start site between CD4 T cells from lupus patients and controls. We found 624 hypermethylated and 661 hypomethylated genes in lupus patients. Functional analysis revealed that top canonical pathways shared among hypermethylated genes include Cell Cycle Regulation, Wnt/β-catenin Signaling, Dendritic Cell Maturation, Graft-versus-Host Disease Signaling, IL-10 Signaling, and p38 MAPK Signaling. Shared canonical pathways among hypomethylated genes include IL-15 Production, Dendritic Cell Maturation, p38 MAPK Signaling, Graft-versus-Host Disease Signaling, FLT3 Signaling in Hematopoietic Progenitor Cells, and Interferon Signaling. Furthermore, many differentially methylated genes in lupus CD4 T cells were also variably expressed, suggesting that the methylation changes observed, at least in a subset of genes, might be functionally relevant. Of the hypomethylated genes, 112 were found to be overexpressed, while 99 hypermethylated genes were underexpressed in lupus CD4 T cells. Some disease-relevant hypermethylated genes in lupus CD4 T cells include CASP3, BCL6, ICAM1, and GZMB. On the other hand, the genes CASP2, FCRL3, STAT1, STAT2, TYK2, IL-18, and TNFSF13 were hypomethylated. Conclusions: We performed the first whole-genome methylation analysis in lupus CD4 T cells and mapped differentially methylated regions across the genome. In addition to global T cell DNA hypomethylation previously described in lupus, our data suggest that promoter and CpG Island hypermethylation in CD4 T cells might also plays a role in the pathogenesis of lupus.

102

DNA methylation changes in monozygote twins showing discordance for psoriasis. *K. Gervin[1], M. Hamerø[1], H. Akselsen[1], H. Nygård[1], A.O. Olsen[2], J.R. Harris[3], D.E. Undlien[4], R. Lyle[1].* 1) Oslo University Hospital, Ullevål, Oslo, Norway; 2) Department of Dermatology, Oslo University Hospital, Rikshospitalet, Oslo, Norway; 3) Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway; 4) Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway.

Monozygote (MZ) twins show phenotypic discordance for several complex disorders, including autoimmune diseases. Phenotype is a result of genetics and epigentics, with the latter being prone to the influence of environmental factors. Development of autoimmune diseases is due to a dysregulation of the immune system, and identification of genes involved in deregulation of lymphocyte function is essential. Psoriasis is known to have a strong genetic component, but environmentally driven epigenetic changes are thought to be involved in the etiology as concordance rate among MZ twins is only approximately 70%. Psoriasis is a chronic inflammatory disease which mainly modulates the skin. People are affected worldwide with a prevalence ranging from 0.1%-6.5%. It is a complex disease which is mediated by T cells, dendrittic cells and inflammatory cytokines. In this study we use MZ twins discordant for psoriasis to explore differences in DNA methylation and gene expression. The study of discordant MZ twins is a good model to investigate epigenetic mechanisms and the functional aspect of it in terms of aberrant gene expression, which can explain the development of disease. Discordance can be interpreted as a result of external factors that shape the epigenetic profile and thereby the susceptibility through altered gene expression. We isolated different lymphocyte subpopulations (CD19+, CD4+, CD8+ and CD4+CD25+) and study single-cell types to overcome the issue of epigenetic heterogeneity in whole blood. In total we have so far collected 30 discordant MZ twin pairs. In order to detect gene specific DNA methylation differences associated with development of psoriasis, we used Infinium methylation assays enabling quantitatively measurement of ~27 000 CpG sites covering >14 000 genes in discordant MZ twin pairs in CD4+ and CD8+ cells. To integrate the analysis of global methylation status and gene expression of approximately all associated genes we have used HumanHT gene expression assays. Analysis of this data identifies methylation differences between twins correlated with gene expression changes and thus identifies genes which are candidates for involvement in psoriasis. We also present preliminary data on genome-wide DNA methylation (RRBS) in MZ twins discordant for disease. To our knowledge this is the first study using MZ twins discordant for psoriasis in order to reveal epigenetic alterations which potentially contributes to the development of disease.

103

Global and Gene-Specific Changes in DNA Methylation Associated with Aging. *K.N. Conneely[1], J.W. Schroeder[1,2], J.F. Cubells[1,2], Y. Tang[1], V. Kilaru[2], K.B. Mercer[1], C.F. Gillespie[2], E.J. Duncan[2,3], K.J. Ressler[2,4], A.K. Smith[2].* 1) Dept Human Genetics, Emory Univ, Atlanta, GA; 2) Dept Psychiatry and Behavioral Sciences, Emory Univ, Atlanta, GA; 3) Atlanta VA Medical Center, Decatur GA; 4) Howard Hughes Medical Institute, MD.

DNA methylation patterns change with age such that monozygotic twins can be differentiated over time, and recent studies report both global and gene-specific methylation changes with age. We examined the association between age and 27,578 CpG sites on the Illumina HumanMethylation27 BeadChip in peripheral blood samples from 147 subjects aged 18-74 recruited from either an urban public hospital or from the VA Medical Center in Atlanta, Georgia. We observed an increase in global methylation with age (P=.017), and 453 gene-specific CpG sites that were differentially methylated with respect to age ($1.7\times10^{-17} < P < 1.8\times10^{-6}$). Among sites associated with increasing methylation, 98.4% were located on CpG islands, which is greater than the 72.5% expected by chance ($P=3.2\times10^{-33}$). In contrast, only 24.0% of sites showing decreased methylation in older subjects were located on CpG islands ($P=1.8\times10^{-7}$). It was not clear what mechanism underlies this pattern, but we noted that since methylation is generally low on CpG islands and high elsewhere, methylation increasing on CpG islands and decreasing elsewhere could reflect a purely stochastic process, in which additions or subtractions of methyl groups from CpGs occur randomly over time, independent of site location. To test this, we estimated average rates of gain and loss of methylation in our data under a model where methylation of each CpG dinucleotide follows a first order Markov process. We estimated near-identical rates of methylation gain for CpG sites on or off islands, but a substantially lower rate of methylation loss for sites on vs. off islands ($P<10^{-168}$), suggesting that the observed changes in methylation with age are due to both stochastic and systematic factors. Enrichment of our 453 differentially methylated CpG sites for relevant gene networks and replication of previously associated CpG sites ($P<10^{-99}$) provided further evidence of systematic, gene-specific methylation changes with age. Similarly, the 428 CpG sites showing increased methylation with age were depleted ($P<10^{-5}$) for histone modifications associated with transcriptional activity (H2AZ, H3K4me3 and PolII) and enriched ($P<.05$) for modifications associated with inactivity (H3K9me3 and H3K27me3) in $CD4^+$ T-cells. As large-scale methylation studies become increasingly common, our data not only underscore the need for careful age-matching in these studies, but also suggest that much can be learned from analysis of age as a secondary phenotype.

104

Interaction of epigenetic and genetic changes in *COL15A1*, a target of aging and atherosclerosis. *J.J. Connelly[1], O.A. Cherepanova[1], J.R. Dungan[2], T. Lillard[1], T. Karaoli[1], Y. Wan[1], C.O. Horton[1], S. Okutsu[1], W.E. Kraus[2], E.R. Hauser[2], S.G. Gregory[2].* 1) University of Virginia, Charlottesville, VA; 2) Duke University, Durham, NC.

We have identified epigenetically regulated genes involved in changing states of aged human aortic smooth muscle cells (AoSMC) using a genome wide screen and here report the characterization of collagen type XV, alpha 1 (*COL15A1*) in aging and atherosclerosis. Bisulfite cloned sequencing of AoSMCs identified two discrete regions of *COL15A1* that are hypomethylated with age. Real Time-PCR and western analyses show *COL15A1* expression increases with methylation loss and treatment with decitabine, a demethylation reagent. Knockdown of *COL15A1* in AoSMCs leads to an increase in migration and a decrease in proliferation suggesting that changes in DNA methylation that lead to alterations in gene expression may modulate the phenotypic state of AoSMCs. COL15A1 expression was identified in the nucleus and cytoplasm of AoSMCs and in cell culture media. Immunostaining of an arterial atherosclerotic lesion from an 18 week ApoE$^{-/-}$ mouse fed a Western diet (10 weeks) showed COL15A1 expression is increased in the atherosclerotic cap and the infiltrated intima compared to the arterial media. Independently, we have identified significantly associated polymorphisms that map to one of the hypomethylated regions of *COL15A1* in aged (>55 years) individuals from CATHGEN (rs4142986, P=0.004, OR=1.7) and show that rs4142986 is a common intronic G/C polymorphism in a CpG site that can be methylated and that the presence of the C allele leads to a decrease in transcription of *COL15A1* (P=0.008). We also fitted a weighted regression model adjusted for atherosclerosis risk factors and sampling weights for 1811 CATHGEN subjects using the outcome of CADindex and the dependent variables: age, *COL15A1* genotype and an age-by-genotype interaction term. The linear regression coefficient for the age-by-genotype interaction for rs12352174 was significantly different from zero ($P=2.6x10^{-5}$) suggesting atherosclerosis risk and age can be modified by rs12352174. This SNP is in low linkage disequilibrium ($r^2=0.017$) with rs4142986 but is located 3,541 bp upstream of rs4142986. It tags two SNPs (rs2067986 and rs928522) that are differentially methylated in embryonic stem cells and fetal fibroblasts. We propose that epigenetic changes associated with aging smooth muscle cells control phenotypic state, transdifferentiation and disease. Our *COL15A1* data strongly suggests that genetically associated polymorphisms have the potential to be epigenetically regulated and therefore causal of disease pathology.

105

Hirschsprung disease illustrates a novel multifactorial mechanism of parent-of-origin effect in complex genetic diseases. *A. Jannot[1,2,3,4], A. Pelet[2,4], J. Amiel[2,4], F. Clerget-Darpoux[1,3], S. Lyonnet[2,4], the International Hirschsprung Disease consortium.* 1) Unite 669, INSERM, Villejuif, France; 2) Unite 781, INSERM, Paris, France; 3) Université Paris Sud, Faculté de médecine, Le Kremlin-Bicêtre, France; 4) Université Paris Descartes, Faculté de médecine, Paris, France.

The genetic make-up of Hirschsprung disease (HSCR) involves one major locus, encoding the RET protooncogene. A parent-of-origin effect for RET mutations has been suspected since a significant transmission distortion has been observed in affected sibships, mutations being more often maternally inherited than transmitted from the father. These data, in addition to the strong sex-ratio disequilibrium in favour of females in HSCR patients, and the low prognosis of HSCR so far, have prompted us to challenge a parent-of-origin effect. In this study, we indeed demonstrated that these two phenomena (biased sex-ratio and low prognosis) can also lead to a transmission distortion. Thus, in order to test whether these two combined observations, would suffice to explain the observed transmission distortion, we have developed the Contrasting Affected sib-pairs and Trios (CAT) method. Applying the CAT method to the International Hirschsprung disease Consortium data (16 sib-pairs and 20 trios with a coding sequence mutation on RET gene), we showed that combining both sex ratio and poor prognosis cannot explain the transmission distortion, and that a parent-of-origin effect was involved. As the RET locus is not mapping to a known imprinted region, this result prompted us to search for an alternative explanation. We showed that HSCR illustrates a novel complex mechanism of parent-of-origin effect: the large spectrum of mutation severity combined with the low prognosis of affected individuals and the high sex-ratio bias, accounts for the observed parent-of-origin effect.

106

Sex-specific differences in recombination are mediated by differences in chromatin compaction and synapsis at the onset of meiosis. *J.R. Gruhn[1], E. Cheng[2], T. Naluai-Cecchini[2], V.Y. Fujimoto[3], T. Pasternack[3], J.M. Schwartz[3], J.E. Steinauer[3], T.J. Woodruff[3], C. Rubio[4], K.W. Broman[5], P.A. Hunt[1], T. Hassold[1].* 1) School of Molecular Biosciences, Washington State University, Pullman, WA; 2) Department of OB/GYN, University of Washington, Seattle, WA; 3) Department of OB/GYN and Reproductive Sciences, University of California San Francisco, San Francisco, CA; 4) Instituto Valenciano de Infertilidad, Valencia, Spain; 5) University of Wisconsin-Madison, Madison, WI.

Sex-specific differences in recombination are common in many mammalian species, including humans; i.e., linkage analysis indicates that human females have over 1.5 fold as many exchanges as males. In the present study, we used cytological methodology to visualize meiotic recombination events in over 2,000 human oocytes and spermatocytes, allowing us to directly compare the number and location of crossovers between the sexes. Specifically, we used antibodies against the crossover-associated protein MLH1, the synaptonemal complex protein SYCP3, and centromere-associated CREST to identify the chromosomal location of exchanges. These results were combined with FISH to generate male and female recombination maps for ten representative chromosomes (1, 6, 9, 13, 14, 15, 16, 18, 21 and 22). Consistent with the results of linkage studies, the average number of MLH1 foci per cell was 1.4 fold higher in females than in males, with similar male:female ratios for each chromosome. Similarly, crossover locations recapitulated linkage data, with MLH1 foci more distally placed in males than females. The spacing of adjacent MLH1 foci also differed between sexes, with males exhibiting significantly stronger interference levels than females. Subsequently, we asked whether - as suggested from data on model organisms - these differences might be linked to variation in the structure of the meiotic axis. Thus, we analyzed chromatin compaction by measuring the lengths of the synaptonemal complexes (SC) and the DNA loops emanating from them and observed striking male:female differences: SCs were significantly shorter and DNA loops significantly longer in males than in females. Significant differences were also observed in the number and location of synaptic initiation sites, with females displaying multiple sites per chromosome and males only one distally-placed site per chromosome arm. These data provide strong evidence that sex-specific differences in recombination rates are established early in meiotic prophase, far upstream of the repair of double strand breaks by MLH1 and other proteins.

107

Deficiency in Lp-PLA2 activity due to carriage of the Asian-specific PLA2G7 gene V279F null allele is protective from coronary artery disease. *D. Waterworth[1], Y. Jang[2], J. Lee[4], K. Song[1], S. Kim[4], H. Kim[5], K.W. Park[5], H. Cho[5], I. Oh[5], J.E. Park[6], B. Lee[6], H.J. Ku[6], D. Shin[2], J.H. Lee[7], S.H. Jee[8], B. Han[9], S. Kim[4], E. Cho[4], P. Vallance[3], J. Whittaker[10], L. Cardon[1], V. Mooser[1].* 1) Genetics, Drug Discovery, GlaxoSmithKline, King of Prussia, PA; 2) From the Division of Cardiology, Cardiovascular Genome Center, Yonsei University College of Medicine, Yonsei, South Korea; 3) Drug Discovery, Research & Development, GlaxoSmithKline Greenford, UK; 4) DNA Link Inc. Seoul, South Korea; 5) Department of Internal Medicine and Cardiovascular Center, Seoul National University Hospital, Seoul, South Korea; 6) Division of Cardiovascular Disease, Samsung Medical Center, Sung Kyun Kwan University, Seoul, South Korea; 7) Department of Food and Nutrition, College of Human Ecology, Yonsei University, Seoul, South Korea; 8) Department of Epidemiology and Health Promotion, Institute for Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, South Korea; 9) Center for Genome Science, National Institute of Health, Seoul, South Korea; 10) Genetics, Drug Discovery, GlaxoSmithKline, Harlow, UK.

Innovative interventions aiming at treating or preventing coronary artery disease (CAD) are urgently needed to reduce the burden of ischemic heart disease, the major cause of death worldwide. Pharmacological inhibition of lipoprotein-associated phospholipase-A2 (Lp-PLA2), an enzyme whose plasma levels are directly associated with CAD risk, has been proposed as a way to meet this goal. Here, we took advantage of an Asia-specific V279F substitution within the Lp-PLA2 encoding PLA2G7 gene, which leads to absence of enzyme activity in plasma, to assess the role of Lp-PLA2 in the pathogenesis of CAD. Given conflicting results from mostly small studies, a large case-control study was warranted. The null allele is present in South Korea at an appreciable frequency (11%) and the study was entirely performed in that country. PLA2G7 V279F genotypes were initially compared in 2890 male cases diagnosed with CAD before age 60 with 3128 male controls without CAD at age 50 and above. This variant was subsequently examined in a second independent male dataset of 686 CAD cases and 707 controls, to evaluate the consistency of the initial result. In the first dataset, the prevalence of the 279F null allele was 11.5% in cases and 12.8% in controls. After adjustment for age, body mass index, diabetes, smoking, glucose and lipid levels, the OR (95% CI) for CAD for this allele was 0.80 (0.66-0.97, p=0.02). The results were very similar in the second dataset, despite lower power, with an allele frequency of 11.3% in cases and 12.7% in controls [OR 0.81 (0.62-1.06), p = 0.13]. Therefore the original findings were substantiated by the addition of this second dataset, as illustrated by the meta-analysis results [OR 0.80 (0.69-0.94), p=0.006]. Furthermore, there was some evidence from the meta-analysis that greater protection was afforded by the null allele in MI patients, with an overall 25% reduction in risk [OR 0.75 (0.61-0.93), p=0.007]. The magnitude and direction of this genetic effect were fully consistent with large epidemiological studies on plasma Lp-PLA2 activity and CAD or MI risk. Natural deficiency in Lp-PLA2 activity due to carriage of PLA2G7 279F allele protects from CAD in Korean men, though the findings are not likely to be restricted to this population. These results support pharmacological Lp-PLA2 inhibition as an innovative way to prevent CAD.

108

***ADAMTS7* and *ABO* are novel coronary artery disease loci that have differential associations with coronary atherosclerosis and myocardial infarction.** *M.P. Reilly[1,2], M. Li[3], J. He[3], J.F. Ferguson[1], M.S. Burnett[4], J.M. Devaney[4], C.W. Knouff[5], J.R. Thompson[6], B.D. Horne[7,8], A.F.R. Stewart[9], T.L. Assimes[10], J. Barnard[11], P.S. Wild[12], H. Allayee[13], P. Linsel-Nitschke[14], N. Martinelli[15], D. Girelli[15], A.A. Quyyumi[16], J.L. Anderson[7,17], H. Schunkert[14], T. Quertermous[10], S. Blankenberg[12], S.L. Hazen[18], R. McPherson[9], S. Kathiresan[19,20,21,22,23], N.J. Samani[24,25], V. Mooser[5], H. Hakonarson[26], S.E. Epstein[4], D.J. Rader[1,2], Wellcome Trust Case Control Consortium.* 1) The Cardiovascular Institute, University of Pennsylvania, Philadelphia, PA 19104, USA; 2) The Institute for Translational Medicine and Therapeutics, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; 3) Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; 4) Cardiovascular Research Institute, Medstar Health Research Institute, Washington Hospital Center, Washington, DC 20010, USA; 5) Genetics Division and Drug Discovery, GlaxoSmithKline, King of Prussia, Pennsylvania 19406, USA; 6) Department of Health Sciences, University of Leicester, Leicester LE1 7RH, UK; 7) Cardiovascular Department, Intermountain Medical Center, University of Utah, Salt Lake City, Utah 84112, USA; 8) Genetic Epidemiology Division, University of Utah, Salt Lake City, Utah 84112, USA; 9) Department of Medicine, University of Ottawa Heart Institute, Ottawa K1Y4W7, Canada; 10) Department of Medicine, Stanford University School of Medicine, Stanford, California 94305, USA; 11) Department of Quantitative Health Sciences, The Cleveland Clinic Foundation, Cleveland, Ohio 44195, USA; 12) Department of Medicine II, University Medical Center Mainz, Mainz 55131, Germany; 13) Department of Preventive Medicine, Institute for Genetic Medicine, USC Keck School of Medicine, Los Angeles, California 90033, USA; 14) Medizinische Klinik II, Universität zu Lübeck, Lübeck 23538, Germany; 15) Department of Clinical and Experimental Medicine, University of Verona, Verona, Italy; 16) Division of Cardiology, Emory University School of Medicine, Atlanta, Georgia 30322, USA; 17) Cardiology Division,University of Utah, Salt Lake City, Utah 84107, USA; 18) The Center for Cardiovascular Diagnostics & Prevention, Cleveland Clinic, Cleveland, Ohio 44195, USA; 19) Cardiovascular Research Center and Cardiology Division, Massachusetts General Hospital, Boston, Massachusetts 02114, USA; 20) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA, 02114; 21) Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA; 22) Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA; 23) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts02142, USA; 24) Department of Cardiovascular Sciences, University of Leicester, Leicester LE3 9QP, UK; 25) Leicester National Institute for Health Research Biomedical Research Unit in Cardiovascular Disease, Glenfield Hospital, Leicester, LE3 9QP, UK; 26) The Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA.

Background: Investigation of the genetic basis of coronary artery disease (CAD) is complicated by substantial heterogeneity in its clinical manifestations. We hypothesized that genetic factors may distinctly contribute to either development of coronary atherosclerosis or specifically to plaque rupture and myocardial infarction (MI) on the background of existing coronary atherosclerosis. Methods: We performed two genome-wide association studies (GWAS) using coronary angiographic phenotyping in European ancestry participants. In order to identify loci that predispose to angiographic CAD (AngCAD), we compared individuals with AngCAD (N=12,353) to those free of AngCAD (controls, N=7,383). In order to identify loci that predispose to MI in the setting of AngCAD, we compared patients with AngCAD and MI (AngCADMI+, N=5,783) to those with AngCAD but no MI (AngCADMI-, N= 3,644). Findings: In comparison of AngCAD versus controls, we identified a novel locus, ADAMTS7 (P = 4.98 x 10-13). In comparison of AngCADMI+ versus AngCADMI-, we identified a novel association exceeding genome-wide significance at the ABO locus (P = 7.62 x 10-9). The ABO association was attributable to ABO blood group O which protects against MI. Most previously published GWAS loci for MI had strong associations with AngCAD versus controls. However, none were related to MI in patients with AngCAD. Interpretation: Using coronary angiographic phenotypes, we identified two loci for distinct CAD manifestations, ADAMTS7 for angiographic CAD, and ABO for MI in patients with angiographic CAD. Our findings indicate that certain loci predispose to the development of coronary atherosclerosis whereas others predispose to subsequent events leading to plaque rupture and MI.

109

A common genetic architecture underlies the heterogeneous phenotypes of hypertriglyceridemia. *C. Johansen[1], J. Wang[1], M. Lanktree[1], H. Cao[1], A. McIntyre[1], M. Ban[1], R. Martins[1], B. Kennedy[1], R. Hassell[2], M. Visser[3], S. Schwartz[4], B. Voight[5], R. Elosua[6], V. Salomaa[7], C. O'Donnell[8], G. Dallinga-Thie[3], S. Anand[2], S. Yusuf[2], M. Huff[1], S. Kathiresan[5], R. Hegele[1].* 1) Department of Biochemistry, Robarts Research Institute, London, Ontario, Canada; 2) Population Health Research Institute, Hamilton Health Sciences, and Departments of Medicine and Clinical Epidemiology, McMaster University, Hamilton, Ontario, Canada; 3) Departments of Experimental Vascular Medicine and Vascular Medicine, Academic Medical Center Amsterdam, Amsterdam, Netherlands; 4) Cardiovascular Health Research Unit and Departments of Medicine and Epidemiology, University of Washington, Seattle, Washington, USA; 5) Center for Human Genetic Research, Massachusetts General Hospital, Boston, and Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 6) Cardiovascular Epidemiology and Genetics, Institut Municipal D'investigacio Medica, and CIBER Epidemiologi'a y Salud Publica, Barcelona, Spain; 7) Chronic Disease Epidemiology Unit, Department of Health Promotion and Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland; 8) Cardiovascular Research Center and Cardiology Division, Massachusetts General Hospital and Department of Medicine, Harvard Medical School, Boston, and Framingham Heart Study and National, Heart, Lung, and Blood Institute, Framingham, Massachusetts, USA.

Introduction: Plasma triglyceride (TG) concentration is re-emerging as significant risk factor for cardiovascular disease. Studying subjects with hypertriglyceridemia (HTG), a model complex polygenic disease defined by fasting plasma TG concentration >95th percentile, may provide insight into the genetic architecture that modulates plasma TG concentration and cardiovascular risk. Currently, genetic predisposition to the heterogeneous HTG phenotypes remains incompletely understood. Methods: Genome-wide genotyping was used to identify novel loci associated with HTG and replicate TG-associated loci in 504 HTG patients and 1213 population-based controls of European descent. Re-sequencing of 4 HTG-associated loci in 438 HTG patients and 327 population-based controls was used to identify rare variants potentially causative of HTG. Phenotypic heterogeneity among HTG subphenotypes (Fredrickson hyperlipoproteinemia [HLP] phenotypes 2B, 3, 4, 5) was compared using TG risk scores, mutation accumulation, explained variation and hierarchical clustering. Results: Four TG-associated loci were associated with HTG at genome-wide significance levels ($P<5X10^{-7}$): *APOA5*, *GCKR*, *LPL*, and *APOB*. Many additional TG-associated loci were also associated with HTG (P<0.05): *MLXIPL*, *TRIB1*, *ANGPTL3*, *NCAN*, and *FADS1*. Genetic risk scores composed of TG-raising alleles were higher in HTG patients vs. controls (mean score 15 vs. 13, $P=8.9X10^{-47}$); highest in patients with severe HTG (HLP type 5). Re-sequencing of the 4 HTG-associated genes revealed a significant mutation skew in HTG patients vs. controls (154 vs. 53; $P=6.2X10^{-8}$), which did not differ quantitatively or qualitatively between HLP phenotypes (P=0.76). Clinical variables, common variants and rare mutations cumulatively explained between 21-52% of variation in HTG diagnoses among HLP phenotypes, least so among subjects with dysbetalipoproteinemia (HLP type 3). Hierarchical analysis between HLP phenotypes revealed a common group of TG-associated loci underlying polygenic HLP phenotypes, whereas remaining loci were associated with HTG in a phenotype-dependent manner. Conclusions: The classical biochemically defined HLP phenotypes appear to share an overlapping genetic architecture, composed of both common and rare variants in TG-associated genes. Future studies must further assess the impact of genetic variation in remaining TG-associated loci in the context of HTG phenotypic heterogeneity.

110

Sortilin is a novel modulator of low-density lipoprotein cholesterol. *A. Strong[1], K. Musunuru[2, 3], M. Frank-Kamenetsky[4], K. Sachs[1], X. Li[1], H. Li[1], S. Lund-Katz[5], M. Phillips[5], J. Wong[4], W. Cantley[4], T. Racie[4], V. Kotelianski[4], K. Fitzgerald[4], S. Kathiresan[2, 3], D. Rader[1].* 1) University of Pennsylvania, Philadelphia, PA; 2) Massachusetts General Hospital and Harvard Medical School; 3) Broad Institute; 4) Alnylam Pharmaceuticals; 5) The Children's Hospital of Philadelphia.

Sortilin, the protein product of the SORT1 gene, is a multi-ligand sorting receptor involved in trafficking of proteins from the Golgi apparatus to the lysosome. Genome wide association studies have identified robust associations between single nucleotide polymorphisms (SNPs) in a genomic region of chromosome 1p13, encompassing the SORT1 gene, to low-density lipoprotein cholesterol levels (LDL-C) and coronary artery disease (CAD). To determine whether SORT1 is the causal gene in the 1p13 locus, we performed overexpression and knockdown experiments in mice and hepatocytes. Adeno-associated virus (AAV)-mediated hepatic sortilin overexpression in "humanized" Apobec-/-, APOB transgenic mice reduced plasma cholesterol by 70% at two weeks (n = 5 mice per group, p = 0.00004), with concomitant reduction of LDL-C. siRNA-mediated hepatic sortilin knockdown in this model increased plasma cholesterol and LDL-C by 30% at two weeks (n = 5 mice per group, p = 0.03). In vivo VLDL production studies demonstrate a reduction of the hepatic VLDL secretion rate by 50% with sortilin overexpression. Labeling studies in sortilin overexpressing and knockdown primary mouse hepatocytes show that sortilin knockdown increases apoB secretion by 30% (p = 0.002) and sortilin over-expression reduces secreted apoB by 30% (p=.01). Pre-incubation of sortilin over-expressing cells with the endolysosome cathepsin inhibitor E64d restores apoB secretion with a trend toward increased secretion (p = .07). Furthermore, preventing sortilin trafficking to the lysosome through overexpression of a sortilin trafficking mutant abolishes sortilin's affect on plasma lipids and VLDL secretion. Surface plasmon reasonance demonstrates a high affinity interaction between sortilin and apoB in LDL particles with a Kd of ~2 nM. In sum, these data are consistent with a model in which sortilin binds Golgi-localized apoB-containing lipoprotein particles and traffics them to the endolysosomal compartment for degradation prior to secretion as part of a pre-secretory degradation pathway (PERPP), resulting in decreased hepatic VLDL secretion, decreased plasma LDL-C levels, and decreased risk of CAD.

111

High-throughput omics approach provides clues for novel pathways behind low HDL-cholesterol. *P-P. Laurila[1], S. Ripatti[1], J. Naukkarinen[1], S. Söderlund[2], M. Oresic[3], M. Jauhiainen[1], M-R. Taskinen[2], L. Peltonen[1,4].* 1) Public Health Genomics Unit, Finnish Institute for Molecular Medicine, Helsinki, Finland; 2) Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland; 3) VTT Technical Research Center of Finland, Espoo, Finland; 4) The Wellcome Trust Sanger Institute, Cambridge, UK.

Low HDL-cholesterol (HDL-C) is a major risk factor for cardiovascular disease and type 2 diabetes. To elucidate potential novel pathways behind low HDL-C, we have employed 3 different high-throughput omics: genomics, transcriptomics and lipidomics. We have genotyped 450 Finnish individuals with low or high HDL-C (10th and 90th population percentiles) using Illumina370K genome-wide platform. Out of subset of these GWA genotyped individuals (n=54), we have obtained subcutaneous adipose tissue biopsies for Affymetrix genome-wide transcriptomics analysis, and isolated plasma HDL particles (n=47) subjecting them to lipidomics analysis with state-of-the-art mass spectrometry technology. We first performed a gene network analysis for the genetic loci associated with low HDL-C. We observed that allelic variants of genes within the antigen processing pathway, including HLA-DQA1, HLA-DRB1 and TNFAIP3, were significantly enriched among the low HDL-C associated genes (p=10^-6). Also the expression of this inflammatory pathway, comprising the aforementioned genes, was significantly up-regulated in the adipose tissue of subjects with low HDL-C (p= 10^-11). Moreover, individual genes of this pathway exhibited an allele specific expression patterns in the adipose tissue, inversely correlating with HDL-C levels. Consistent with the enrichment of inflammatory processes in association and expression analyses, the inflammatory nature of the HDL particle itself was evident in the lipidomics analysis of subjects with low HDL-C. For instance, arachidonic acid containing plasmalogens, precursors to potent inflammatory mediators, such as leukotriens, were elevated in HDL particles from subjects with low HDL-C. Likewise, the content of pro-inflammatory TNFα was increased in low HDL-C subjects, in line with the observed increase in expression of its low HDL-C associating target gene TNFAIP3. Our findings imply a causal role for inflammatory pathways in the regulation of HDL-C. We have demonstrated by combining GWA with transcriptomics from fat biopsies and lipidomics from the HDL particle that genetic variation within members of the inflammatory pathways influences both HDL-C levels and gene expression of that pathway. These changes are not only reflected in the HDL-C levels but also in the differential composition of the HDL particles between subjects with high and low HDL-C suggesting a more inflammatory and less vasoprotective role for HDL particles in subjects with low HDL-C.

112

Genome-wide association study identifies the Pro2712 variant of apolipoprotein B as a key determinant of elevated plasma oxidized LDL levels associated with coronary artery disease. *A.F.R. Stewart[1], M.A. Alam[1], R.W. Davies[1], M. Fan[1], O. Raitakari[2], L.-P. Lyytikäinen[3], T. Lehtimaki[3], G.A. Wells[1], R. McPherson[1], R. Roberts[1].* 1) Univ Ottawa Heart Inst, Ottawa, ON, Canada; 2) Turku University Central Hospital, Turku, Finland; 3) Tampere University Hospital, Tampere, Finland.

Background: Elevated levels of low density lipoprotein (LDL) predict coronary artery disease (CAD) risk. Genetic variants that elevate LDL also contribute to the risk of CAD. Patients with CAD also have markedly elevated levels of circulating oxidized LDL (oxLDL) when compared to controls. However, it remains controversial whether LDL is oxidized as a consequence of CAD or whether oxLDL contributes to CAD. **Methods and Results**: Here, we sought to identify genetic variants that associate with plasma levels of oxLDL in a genome-wide association study (GWAS) in early onset CAD cases (49±7.0 years) with replication in elderly asymptomatic controls (75±5.0 years). oxLDL was measured using the 4E6 antibody (Mercodia) in a competitive ELISA assay in 1628 early onset CAD cases and also in 558 asymptomatic elderly controls genotyped using Affymetrix single nucleotide polymorphism (SNP) microarrays. In CAD cases, several SNPs were highly associated with oxLDL ($p=10^{-10}$) that tagged an imputed SNP (rs676210) coding for the Pro2712Leu substitution in apolipoprotein B that showed the strongest association with oxLDL ($p=10^{-13}$). The association remained genome-wide significant after adjusting for baseline or current LDL levels or for apoB100 levels measured in separate ELISA assays, or when adjusting for triglycerides. Replication revealed a much weaker association of this locus with oxLDL levels in asymptomatic elderly controls (p=0.03). However, this association was not replicated in the Young Finns cohort of 1800 healthy young individuals for whom oxLDL levels were measured (32.0±5.0 years, p=0.89). This finding likely reflects the power to detect an association within the sample with the highest levels of oxLDL (CAD cases) than in samples with lower (elderly controls) or the lowest (young Finns) levels of oxLDL. Remarkably, this locus did not associate with CAD risk in over 12,000 individuals, suggesting that elevated plasma oxLDL levels are a consequence of CAD. **Conclusion**: The highly significant association of the Pro2712Leu APOB variant with oxLDL levels in CAD cases ($p=10^{-13}$) suggests that a structural change in the APOB protein increases its susceptibility to oxidation.

113

Uric acid is causally associated with blood pressure in a controlled setting: results from a Mendelian randomization analysis of GLUT9. *A. Parsa, E. Brown, MR. Weir, BD. Mitchel, AR. Shuldiner, PF. McArdle.* Dept Medicine, Univ Maryland Medical System, Baltimore, MD.

BACKGROUND: Elevated serum levels of uric acid (UA) are consistently associated with hypertension, but the directionality of the association remains debated. We used, in a controlled setting, a Mendelian randomization approach to demonstrate the direction of the association between UA and blood pressure (BP). Using a recently described functional single-nucleotide polymorphism, rs16890979 (Val253Ile), in the GLUT9 gene that is associated with lower UA levels, we examined the unconfounded association between genotype and BP. METHODS: Genotyping was performed on 516 participants in the HAPI Heart Study to identify variants of a genetic locus reliably associated with serum uric acid level (GLUT9). Participants were on fixed standardized high and low sodium diets (280 and 40 meq per day) for 6 days each and BP measures were based on ambulatory mean 24 hour BP measures. Using Mendelian randomization analysis, relationships between genotype and both UA and BP were assessed. RESULTS: Each copy of the minor Ile allele conferred approximately a 0.44 mg/dl reduction in uric acid (p = 3.2x10-11). On the high sodium diet each copy of the Ile allele was associated with a mean decreases in systolic BP of 2.2 mmHg (p = 0.0058). The effect of the genotype was attenuated, but still significant on the low sodium diet with a mean decreases in systolic BP of 1.48 mmHg (p = 0.038). CONCLUSION: Decreases in serum uric acid concentration, as directly mediated by genetic variants of GLUT9, are causal of decreases in BP and may be modified by dietary sodium intake.

114

Genetic variants in 22 loci are associated with cardiac ventricular conduction (QRS duration). *D.E. Arking[1], N. Sotoodehnia[2], A. Isaacs[3], P. de Bakker[4], M. Dörr[5], C. Newton-Cheh[6], I. Nolte[7], P. van der Harst[8], M. Müller[9], M. Eijgelsheim[10], A. Alonso[11], A. Hicks[12], S. Padmanabhan[13], C. Hayward[14], A. Smith[15], O. Polasek[16], S. Giovannone[17], J. Fu[7,18], I. Rudan[19,20], H. Snieder[21], J.F. Wilson[20], P. Pramstaller[12], T. Wang[6], V. Gudnason[15], C.M. van Duijn[3], S.B. Felix[5], G.I. Fishman[17], Y. Jamshidi[21,22], N.J. Samani[23], S. Kääb[24], CHARGE Consortium.* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Cardiovascular Health Research Unit, Department of Medicine, University of Washington. Seattle WA, USA; 3) Genetic Epidemiology Unit, Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 4) Program in Medical Population Genetics, Broad Institute of Harvard and MIT, Cambridge MA, USA; 5) Department of Internal Medicine B, Ernst-Moritz-Arndt-University, Greifswald, Germany; 6) Center for Human Genetic Research, Cardiovascular Research Center, Massachusetts General Hospital, Boston MA, USA; 7) Unit of Genetic Epidemiology and Bioinformatics, Department of Epidemiology, University Medical Center Groningen, Groningen, The Netherlands; 8) Department of Cardiology, University Medical Center Groningen, Groningen, The Netherlands; 9) Institute of Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 10) Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 11) Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA; 12) Institute of Genetic Medicine, European Academy Bozen/Bolzano (EURAC), Bolzano, Italy (Affiliated institute of the University of Lübeck, Germany); 13) BHF Glasgow Cardiovascular Research Centre, University of Glasgow, UK; 14) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Edinburgh, Scotland; 15) Icelandic Heart Association Research Institute, Kopavogur, Iceland; 16) Andrija Stampar School of Public Health, University of Zagreb, Zagreb, Croatia; 17) Division of Cardiology, New York University School of Medicine, New York, NY, USA; 18) Department of Genetics, University Medical Center Groningen, University of Groningen, The Netherlands; 19) Croatian Centre for Global Health, University of Split, Split, Croatia; 20) Centre for Population Health Sciences, University of Edinburgh, Edinburgh, Scotland; 21) Department of Twin Research and Genetic Epidemiology Unit, St Thomas' Campus, King's College London, St Thomas' Hospital, London, United Kingdom; 22) Division of Clinical Developmental Sciences, St George's University of London, London, UK; 23) Department of Cardiovascular Sciences, University of Leicester, Leicester, UK; 24) Department of Medicine I, Klinikum Grosshadern, Munich, Germany.

The electrocardiographic QRS interval reflects ventricular depolarization and its duration is a function of electrophysiological properties within the His-Purkinje system and the ventricular myocardium. Longer QRS is a risk factor for mortality, sudden death, and heart failure. The heritability of QRS duration has been estimated at up to 40%, but to date, few genes influencing the trait have been identified. In the current study, a large-scale meta-analysis of GWAS data was performed, comprising results from 14 studies consisting of 40,407 individuals of European ancestry, with additional validation genotyping in 7,170 individuals. Genotypes at more than 2.5 million SNPs were imputed using the HapMap CEU reference panel, and genomic control was used to correct both the test statistics from the individual studies as well as the overall meta-analytic results. We identified 22 loci associated with QRS duration ($P<5x10^{-8}$), of which only four were previously reported. These 22 loci combined explain ~5.7 ± 2.3% of the variance of the trait, and map in or near genes in cardiac conduction pathways, including sodium-channels (*SCN10A* and *SCN5A*), calcium handling (*PLN*/*SLC35F1*, *STRN*/*HEATR5B*, *CASQ2*, *TKT*/*CACNA1D*, and *PRKCA*), and transcription factors (*NFIA*, *HAND1*, *TBX3*, *TBX5*, *TBX20*, and *KLF12*), in addition to novel pathways (including kinase inhibition and growth-factor related genes). To test whether QRS-prolonging alleles, in total, increase risk of ventricular conduction defects, we calculated a risk score by summing the number of QRS prolonging alleles per individual. In an independent set of 523 individuals with bundle branch block or nonspecific prolongation of QRS interval (QRS>120 ms) and 12,804 individuals with normal conduction, each additional copy of a QRS prolonging allele was associated with a 4.2% (95% CI 1.3% to 7.6%) increase in risk of ventricular conduction defect (p=0.0014). Of note, the strongest signal was on 3p22, where we identified at least 4 SNPs independently associated with QRS duration, including 2 each in *SCN10A* and *SCN5A*. Finally, we also showed experimentally that *SCN10A* is expressed preferentially in the mouse ventricular conduction system, and that treatment with a selective *SCN10A* blocker prolongs QRS duration in the mouse. These analyses greatly expand our understanding of the genetic basis for variation in QRS duration, and provide insights into the biology of cardiac conduction.

115

Chromosome 4q25 variants are genetic modifiers of rare ion channel mutations associated with familial atrial fibrillation. *D. Darbar[1], S. Rowan[1], G. Kucera[1], T. Stubblefield[1], M. Blair[1], S. Carter[1], D.M. Roden[2], M.D. Ritchie[3].* 1) Division of Cardiovascular Medicine Vanderbilt University School of Medicine 2215B Garland Avenue Room 1285A MRB IV Nashville, TN 37323-6602; 2) Department of Medicine, Division of Clinical Pharmacology Vanderbilt University School of Medicine 2215B Garland Avenue Room 1285B MRB IV Nashville, TN 37323-6602; 3) Center for Human Genetics Research 519 Light Hall Nashville, TN 37232-0700.

Although mutations in ion channels, gap junction proteins, and signaling molecules have been described for Mendelian forms of atrial fibrillation (AF), penetrance is highly variable. Recent studies have consistently identified two common single nucleotide polymorphisms (SNPs) in the chromosome 4q25 region as independent AF susceptibility alleles. Here we tested the hypothesis that 4q25 genotypes contribute to the variable penetrance of the AF phenotype in familial AF. We studied 12 families in which AF was present in ≥2 individuals who also shared a candidate gene mutation. These mutations were identified in all subjects with AF (n=33) as well as apparently unaffected family members (age >50 yrs with no AF; n=17). Mutations were identified in SCN5A (n=6); NPPA (n=2); KCNQ1 (n=1); KCNA5 (n=1); KCNE3 (n=1) and NKX2.5 (n=1). In genetic association analyses, unstratified and stratified according to age of onset of AF and unaffected age > 50 yrs, there was a highly statistically significant association between the presence of both common (rs2200733 and rs10033464) as well as and rare variants and AF (un-stratified P<1x10-9; stratified [age of onset <50 yrs and unaffected age >50 yrs], P<0.0001). We conclude that common AF-associated 4q25 polymorphisms modify the clinical expression of latent cardiac ion channel and signaling molecule gene mutations associated with familial AF. These findings support the idea that the genetic architecture of common human phenotypes like AF includes both rare and common variants.

116

MicroRNA Regulation of the Cardiac Stress Response. *J.T.C. Shieh[1,2], D. Srivastava[2,3,4].* 1) Division of Medical Genetics, University of California, San Francisco, CA; 2) Gladstone Institute of Cardiovascular Disease, San Francisco, CA; 3) Division of Pediatric Cardiology, Department of Pediatrics, University of California, San Francisco, CA; 4) Department of Biochemistry and Biophysics, University of California, San Francisco, CA.

Hundreds of microRNAs are encoded in the human genome, and certain microRNAs are expressed in a tissue-specific manner, suggesting potential unique functions. We previously identified a highly conserved microRNA, miR-499, by expression profiling of the human heart, and recent studies have focused on its potential importance in cardiac gene regulation. Emerging data suggest that microRNA levels vary in different disease states, such as in heart failure or aortic stenosis, however it is unknown whether these changes contribute to disease or are adaptive. Altered microRNA levels could exert differential effects on tissue gene expression and subsequently change organ physiology. In this study, we tested the effect of miR-499 dosage using a murine model of cardiac disease. We utilized cardiac-specific transgenic mouse lines that express increased levels of miR-499 and performed functional analyses by echocardiography in addition to pathologic analyses for hypertrophy, fibrosis, and apoptosis. Gene expression profiling was performed using Affymetrix gene expression arrays to determine genes altered by miR-499 and affected cellular pathways. To test the role of additional cardiac stress, we used thoracic aortic banding, which mimics states of increased pressure load, and monitored cardiac function over three months. We found that increased levels of miR-499 in the heart can lead to cardiac hypertrophy. High levels of miR-499 led to spontaneous cardiac enlargement and contractile dysfunction, while more modest levels of transgenic expression conferred susceptibility to pressure load-induced dysfunction. Gene expression analyses of miR-499-altered hearts revealed select genes involved in hypertrophy and identified three unexpected genes involved in the cardiac stress response. These findings were validated in cultured cells and by inhibiting miR-499 using an antisense RNA that blocks miR-499 generation. These studies support a contribution of elevated cardiac miR-499 levels in cardiac dysfunction, particularly in the setting of pressure overload, and suggest a potential role for microRNAs in disease predisposition.

117

Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *J. Degner[1,4], R. Pique-Regi[1,2,4], A. Pai[1], D. Gaffney[1,3], Y. Gilad[1], J. Pritchard[1,3].* 1) Human Gen, Univ Chicago, Chicago, IL; 2) Committee on Genetics, Genomics and Systems Biology, University of Chicago, Chicago, IL; 3) Howard Hughes Medical Institute, University of Chicago , Chicago, IL; 4) These authors contributed equally.

A complete map of the DNA targets of transcription factors across tissue types and developmental stages will be fundamental to our understanding of the eukaryotic regulatory code. Here, we report a genome-wide map of 827,000 active transcription factor binding sites in human lymphoblastoid cell lines (LCLs) which includes TF-bound sites for 239 position weight matrices corresponding to known TFs, and for 49 novel sequence motifs. To generate this map, we have developed an alternative to ChIP that uses the characteristic DNase-I footprint of each TF given by genome-wide DNase-I cut-site data. A novel Bayesian hierarchical mixture model is used that captures differences between bound and unbound TF motifs in the distribution of DNase-I sensitivity with single-base resolution. Additionally, our model utilizes annotation information that affects our prior probability a site is bound (e.g. match to the TF position weight matrix (PWM), sequence conservation, and distance to the transcription start site). An Expectation Maximization (EM) algorithm is used fit the parameters of our model and we make our inference from an estimate of the posterior probability that a specific motif instance is active (i.e. bound by a TF). We apply our model to every occurrence of each of the PWMs in the TRANSFAC and JASPAR databases (~1000 PWMs) using data from HAPMAP cell lines. For thirteen of these PWMs (in two cell-line types), we use publicly available ChIP-seq data from the Encode Project to validate our model and find that our predictions are quite congruent with ChIP-seq results (area under curve for ROC curves > 0.95). For PWMs where publicly available ChIP data is unavailable, we use sequence conservation as validation and predict in vivo binding for 239 of these PWMs. In addition, we have applied our model to all enriched 10-mers in DNase hypersensitivity sites that show significant sequence conservation. Using this approach, we independently recover many of the PWMs in TRANSFAC and JASPAR. Furthermore, we estimate 49 PWMs with no good match in these databases and that represent binding of novel or poorly characterized TFs. We anticipate that this approach will be an invaluable tool for genome-wide studies of gene regulation in a wide variety of cell-types or tissues under diverse conditions and may help to illuminate the mechanisms through which genetic variation leads to complex variation in gene expression and disease phenotypes among individuals.

118

Ablation of Tbx20, a developmental gene, results in dilated heart failure in adult mice. *I. Aneas[1], T. Shen[2], N. Sakabe[1], S. Evans[2], M.A. Nobrega[1].* 1) Human Genetics, University of Chicago, Chicago, IL; 2) Skaggs School of Pharmacy and Pharmaceutical Sciences, UCSD, San Diego, CA.

Tbx20 is a transcription factor required for heart development. Although mutations in Tbx20 have been linked to congenital heart diseases, little is known about its function in the adult heart. To elucidate the role of TBX20 in adult heart function, we generated a conditional heart deletion of TBX20 in 4-6 week old mice, which resulted in 100% mortality at 5-16 days post tamoxifen injection. TBX20 adult mutants demonstrated impaired cardiac function and severely abnormal electrophysiological function, including severe arrhythmia, dilated cardiomyopathy and increased cell death. To characterize the molecular pathways involved in these phenotypes we used two different strategies. First, we carried out RNA-Seq in hearts of wild-type and TBX20 null mice. We observed 7,893 genes differential expressed, 5,905 upregulated and 2,288 downregulated in knockout hearts compared to control animals. Second, we used chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) to define the genome-wide binding sites of Tbx20 in adult hearts. We identified 4,012 peaks, 77% of which mapping > 6Kb from TSS. Genes that are differentially expressed and contain at least one TBX20 binding region revealed by ChIP assay are the most likely candidates to be underlying the phenotypes we observed. 1,973 genes, (1,293 overexpressed and 688 underexpressed) meet these criteria. These genes are highly enriched for molecular functions such as heart contraction, calcium transport and energy metabolism. We have validated, in vivo, over 45 of these direct downstream targets, using zebrafish reporter assays, confirming that 67%; of these Tbx20 binding regions correspond indeed to heart enhancers. These data indicate that Tbx20, a transcription factor that is critical in heart development and associated with congenital heart diseases is also a key regulator of adult cardiac function and has the potential to lead to heart failure.

119

Kinesin overexpression influences taxane resistance in breast cancer. *M.H. Tan[1], S. De[2], G. Bebek[1,6,7], M.S. Orloff[1,3], R. Wesolowski[3], E. Downs-Kelly[4], G.T. Budd[3], G.S. Stark[2,3,5,6], C. Eng[1,3,5,6].* 1) Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH; 2) Department of Molecular Genetics, Lerner Research Institute, Cleveland Clinic, Cleveland, OH; 3) Department of Solid Tumor Oncology, Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH; 4) Department of Anatomic Pathology, Pathology and Laboratory Medicine Institute, Cleveland Clinic, Cleveland, OH; 5) Department of Genetics, Case Western Reserve University School of Medicine, Cleveland, OH; 6) Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, OH; 7) Center for Proteomics Research, Case Western Reserve University School of Medicine, Cleveland, OH.

Statement of Purpose: Breast cancer is a genetically heterogenous disease with subtypes differing in prognosis and chemosensitivity. Taxane-based chemotherapy is often administered to prevent relapse or to shrink locally advanced tumors before resection. These antimitotics act through prevention of microtubule disassembly. Kinesins are a family comprising 41 motor proteins that interact with microtubules to transport cellular cargo. We thus hypothesized that specific kinesin overexpression would influence taxane resistance in cancer. Methods: Kinesin expression was studied in 3 gene expression datasets in relation to taxane resistance: (i) the diverse and well characterized NCI60 human tumor cell line panel, (ii) the MD Anderson Cancer Center (MDACC) dataset of pretreatment breast cancer samples exposed to neoadjuvant taxane-based chemotherapy (n=230) and (iii) tumor samples from residual breast cancer following neoadjuvant paclitaxel chemotherapy (n=8). Validation-based insertional mutagenesis (VBIM or “sleeping beauty”), a novel functional approach to gene modification, was used to investigate underlying mechanisms of taxane resistance by selection of kinesin-overexpressing cell line clones. Results: In the NCI60 cell line dataset, kinesin KIFC3 overexpression is correlated with resistance to both taxanes, docetaxel ($p<0.001$) and paclitaxel ($p<0.001$), but not to platinum drugs, including carboplatin ($p=0.49$) and cisplatin ($p=0.10$). KIFC1 overexpression predicted docetaxel resistance in the unselected MDACC breast cancer dataset, with KIFC3 overexpression corresponding to docetaxel resistance in basal-like breast cancer subtype cancers. KIF26B is overexpressed in taxane-resistant residual breast cancers postchemotherapy ($p=0.024$). In corresponding functional work, kinesin overexpression is associated with increased docetaxel resistance in basal-like breast cancer cell lines. This resistance is ablated upon mutation of the kinesin ATP-binding domain. Conclusion: We show that kinesin overexpression correlates with taxane resistance in both clinical and functional in vitro settings, and that the kinesin ATP-binding domain plays a key role. The critical role of genetic subtyping of breast cancer in drug response is emphasized in our analysis. Our results suggest a potential approach to overcoming taxane resistance through concurrent or sequential use of kinesin inhibitors, highlighting the ATP-binding domain as a molecular target for drug development.

120

The first transcriptional map of the human Major Histocompatibility Complex reveals new transcripts and haplotype-specific pattern of expression. *C. Vandiedonck[1,2], M.S. Taylor[1,3], H. Lockstone[1], K. Plant[1], J.M. Taylor[1], C. Durrant[1], J. Broxholme[1], B.J Fairfax[1], J.C. Knight[1].* 1) Wellcome Trust Centre for Human Genetics, Oxford University, Oxford, UK; 2) UMRS 958 INSERM, Université Paris 7, Faculté de Medecine Villemin, PARIS, France; 3) European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK.

As the most gene dense and polymorphic region in the human genome, the human Major Histocompatibility Complex (MHC) has been a paradigm for genomics. It is also the top-ranking region for the number of associations with immune and non-immune diseases. Recently, it was entirely re-sequenced for frequent haplotypes associated with common autoimmune diseases. Here, using an original hybrid microarray, including both tiling and junction probes, we draw the first strand-specific transcription map of the MHC in the context of three common haplotypes strongly associated with autoimmune disease, HLA-A3-B7-Cw7-DR15, HLA-A1-B8-Cw7-DR3 and HLA-A26-B18-Cw5-DR3-DQ2, carried by the homozygous lymphoblastoid cell lines PGF, COX and QBL. In total, 6% of the MHC is transcribed with an average density of one transcriptional active region (TAR) per 1.4kb, including new TARs in intergenic regions. Up to 11% of the TARs are haplotype-specific. The distributions of differentially expressed probes and of polymorphisms between haplotypes are significantly correlated, arguing for a role of *cis*-regulatory variants. We identified 96 differentially expressed genes and 526 differentially spliced exons between haplotypes. These patterns were validated by RT-PCR and, for the top gene, confirmed in an independent cohort of 93 healthy volunteers. We also observed that alternative splicing is significantly more extensive in the MHC than in the rest of the genome, but was independent of its immune related function. This study marks a new step towards the identification of regulatory variants involved in the control of MHC associated phenotypes.

121

A gene expression signature predictive of immunogenicity after trivalent influenza vaccination in humans. *L.M. Franco[1,2], K.L. Bucasas[4], C.A. Shaw[1], M.S. Bray[1], J.M. Wells[3], D. Nino[3], N. Arden[6], J.M. Quarles[6], R.B. Couch[2,3], J.W. Belmont[1,5].* 1) Department of Molecular and Human Genetics, Baylor College Medicine, Houston, TX; 2) Department of Medicine, Baylor College of Medicine, Houston, TX; 3) Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX; 4) Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 6) Department of Microbial and Molecular Pathogenesis, Texas A&M University System Health Science Center, College Station, TX.

Influenza viruses are highly contagious respiratory pathogens that represent one of the major public health problems worldwide. Annual vaccination has been the primary strategy for prevention of influenza virus infections. However, great inter-individual variability exists in vaccine responses, adequate correlates of protection are not available and the molecular events that take place in vivo after vaccination are poorly understood. To evaluate these events, we immunized an experimental cohort of 119 healthy human subjects with a clinically available trivalent influenza vaccine and performed a timed assessment of genome-wide gene expression patterns in peripheral blood before and at three time points after vaccination using Illumina Human HT-12v3 expression microarrays. We detected strong upregulation of several immune response genes in the initial 24 hours after immunization, particularly those in Gene Ontology categories related to antiviral response, interferon response, and in the processing and presentation of antigens. Genes involved in RNA processing, translational elongation, macromolecule biosynthesis and T-cell selection were found to be upregulated on days 3 and 14 after vaccination. The pattern of gene activation strongly suggests that the early innate immune response is directly shaped by the influenza antigens and that there is substantial individual variation in this response. General differences in the intensity of the immune response to the vaccine among individuals within our cohort allowed us to develop a robust scoring system, the Titer Response Index (TRI) based on aggregate antibody responses. Computational analyses revealed a 494-gene expression signature— including *STAT1* and *E2F2*— which strongly correlates with the antibody response measured 28 days after vaccination. Our results provide insight into the early and late molecular mechanisms that play a role in the human response to influenza vaccination. They also propose a transcriptional correlate for humoral immunogenicity and identify several potential targets for assessing and improving the efficacy of influenza vaccines.

122

mRNA-Seq transcriptome analysis of human trisomy 21 using discordant monozygotic twins. *A. Letourneau[1], S.B. Montgomery[1], E. Migliavacca[1,2], D. Gonzalez[3], C. Borel[1], D. Robyr[1], L. Farinelli[4], M. Gagnebin[1], E. Falconnet[1], S. Deutsch[1], S. Dahoun-Hadorn[1,5], J.L. Blouin[1,5], R. Guigo[3], E.T. Dermitzakis[1], S.E. Antonarakis[1].* 1) Genetic Medicine & Development, University of Geneva Medical School, 1211 Geneva, Switzerland; 2) Swiss Institute of Bioinformatics, Switzerland; 3) Center for Genomic Regulation, University Pompeu Fabra, 08003 Barcelona, Spain; 4) FASTERIS SA, 1228 Plan-les-Ouates, Switzerland; 5) University Hospitals of Geneva, 1211 Geneva, Switzerland.

Trisomy 21 (T21) is the most widely studied model phenotype of whole chromosome aneuploidy. It is likely that the majority of the T21 phenotypes are related to alterations of gene expression. Entire transcriptome sequencing now provides the opportunity to investigate in detail the perturbations of gene expression in T21 cells and tissues. In this study we used fibroblasts derived from a pair of monozygotic twins discordant for T21 (Dahoun et al. 2008). For the first time, the use of these samples eliminates the bias of genome variability and thus most of transcriptome differences observed are likely to be related to the supernumerary chromosome 21.

The transcriptome (polyA+ mRNA) was studied by RNA-Seq; 30 million 76 bp paired-end reads were generated from each sample and mapped using both MAQ (Li et al. 2008) and GEM (Ribeca et al, unpublished data). The expression level of about 15'000 protein-coding genes were compared between the samples. The differences of gene expression level, spliced isoform utilization, allelic expression, and chimeric transcripts were studied. The results were also compared with previous microarray studies of T21 fibroblasts and matched controls. As expected, we found that the majority (93%) of differentially expressed genes on chromosome 21 are upregulated in the trisomic twin. Interestingly, we found that genes involved in developmental processes are over-represented among the 200 top genes downregulated in the trisomic twin whereas genes implicated in protein binding are enriched within the 200 top genes overexpressed in the T21 sample (Gene Ontology analysis, Benjamini-Hochberg adjusted p-values < 5e-08 and 2e-10, respectively). In addition, we identified 58 genes that differentially utilized alternative spliced isoforms between the two samples.

Our study shows that RNA-Seq in monozygotic discordant twins provides the opportunity to identify biologically relevant transcriptome differences without the bias of the genetic variation among the samples.

123

Medical sequencing of 400 cases and 200 controls for 27 genes at loci associated with type 2 diabetes. *T.M. Teslovich[1], P.S. Chines[2], L.L. Bonnycastle[2], A.J. Swift[2], P. Cruz[3], NIH. Intramural Sequencing Center[3], F.S. Collins[2], J.C. Mullikin[3], M. Boehnke[1].* 1) Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA; 2) National Human Genome Research Institute, National Institute of Health, Bethesda, Maryland 20892, USA; 3) NIH Intramural Sequencing Center, Bethesda, Maryland 20892, USA.

Type 2 diabetes (T2D) is a leading cause of morbidity and mortality in the United States and worldwide, with ~200 million people affected worldwide. Recent genome-wide association and candidate gene studies have implicated nearly 40 loci associated with T2D at genome-wide significance. To understand better the allelic architecture of T2D, we performed medical sequencing of 27 genes at 17 T2D loci in 400 familial, early onset, drug-treated type 2 diabetics and 200 older normal glucose tolerant controls from the Finland-United States Investigation of NIDDM Genetics (FUSION) study. Exons, splice sites, and promoters comprising 228kb of target were sequenced by Sanger sequencing. We identified 1387 SNPs and 136 indels in regions of interest, 799 (52.5%) of which are present in dbSNP. Predicted variants include 24 frameshift, nonsense, and splice site variants, as well as 60 non-synonymous coding variants predicted by CDPred to be deleterious. Genotyping on the Sequenom MassARRAY platform confirmed 42 of 48 variants, including 16 of 17 singletons. Putative deleterious variants are being genotyped in additional samples, including 131 variants that were included on the Illumina MetaboChip.

124

Whole Exome Capture and Parallel DNA Resequencing of Familial ALS Cases. *H. Daoud, S. Girard, J. Gauthier, P. Dion, G. Rouleau.* Centre of Excellence in Neuromics, CHUM Research Center and the Department of Medicine, Université de Montréal, Montreal, Canada.

Amyotrophic lateral sclerosis (ALS) is an adult-onset rapidly progressive neurodegenerative disease that leads to a progressive paralysis due to the death of motor neurons (MN) in the central nervous system. The identification of ALS causative genes, notably SOD1, TARDBP and FUS, is already providing promising insights for a better understanding of the disease but both the basic pathogenic mechanism and the genetic etiology of most ALS cases remain unknown. This emphasizes our need to identify additional ALS-causing genes and converging evidence suggests that a large number of individually rare, highly pathogenic, mutations underlie a substantial fraction of the familial cases of the disease. In order to identify these rare ALS-causing mutations, we performed a whole exome capture and parallel DNA resequencing in a cohort of 60 unrelated SOD1, TARDBP and FUS negative patients with familial amyotrophic lateral sclerosis (FALS). We first selected 60 unrelated FALS patients from our largest and clinically well-defined familial ALS pedigrees (affected and unaffected relatives), that are SOD1, TARDBP and FUS negative. We then proceeded with the exome capture of 26 individuals (18 affected and 8 unaffected), altogether eight families, using the Agilent Technology, SureSelect Human All Exon kit, and resequence the products of these captures using an ABI SOLiD3 machine. Short reads alignment was performed using two different algorithms (BFast and Burrow-Wheeler transform) and SNP variations were filtered using stringent quality controls (SNP coverage, SNP frequency, etc). For each individual sequenced, a list of genes with rare missense or protein truncating variants was generated. Genes with rare variants present in all the affected individuals, but not the unaffected or control individuals or in public SNP databases, were then identified for each family. Therefore, by gathering the information for each of the eight families together, we were able to establish a list of candidate genes for FALS, which will undergo a second stage of genetic validation. We believe this study will enable us to identify novel gene(s) for ALS which would open new avenues for research into the pathogenesis of this disease and offer leads to the development of new treatment strategies.

125

Loss-of-function mutations in healthy human genomes: implications for clinical genome sequencing. *D.G. MacArthur[1], S. Balasubramanian[2], N. Huang[1], A. Frankish[1], Z. Zhang[2], C.A. Albers[1], R. Handsaker[3], L. Habegger[2], X. Mu[2], E. Khurana[2], B. Yngvadottir[1], M. Bainbridge[4], J. Harrow[1], R.A. Gibbs[4], S.A. McCarroll[5], M.E. Hurles[1], M.B. Gerstein[2], C. Tyler-Smith[1], 1000 Genomes Consortium.* 1) Wellcome Trust Sanger Institute, Hinxton, UK; 2) Yale University, New Haven, CT; 3) Broad Institute of MIT and Harvard, Cambridge, MA; 4) Baylor College of Medicine, Houston, TX; 5) Harvard Medical School, Boston, MA.

For the clinical benefits of high-throughput sequencing technologies to be fully realised it must be possible to distinguish medically relevant genetic variants from benign polymorphisms in individual human genomes.

Here we present a comprehensive analysis of variants predicted to severely affect gene function, using data from all three pilots of the 1000 Genomes Project (low-coverage sequencing of ~180 individuals, high-coverage sequencing of six individuals and sequencing of ~1,000 genes in over 600 individuals), including all nonsense and splice site-disrupting SNPs, frame-shift-inducing indels and structural variants predicted to disrupt one or more protein-coding exons.

We show that predicted loss-of-function (LOF) variants are highly enriched for sequencing and annotation artefacts, posing challenges for clinical sequencing projects. We have thus developed a series of stringent filters to reduce false positive rates for LOF variants, including analysis of overlapping copy number variation and multi-nucleotide substitutions, and manual reannotation of gene models.

Even after filtering, the number of LOF variants in the genomes of healthy individuals is surprisingly high. LOF variants display a derived allele frequency spectrum consistent with many being mildly deleterious to human health. We have explored the functional impact of these variants through comparison with RNA-seq data, genome-wide association study signals, and Mendelian disease genes.

Finally, we have produced catalogues of LOF variants and LOF-tolerant genes for use in clinical genome sequencing projects. We demonstrate marked differences in functional, evolutionary, expression and interaction properties of LOF-tolerant genes and severe recessive Mendelian disease genes, potentially allowing prioritisation of novel variants for functional and clinical follow-up.

126

A Genome-wide Copy-number Polymorphism Scan for Diabetic Nephropathy: The Family Investigation of Nephropathy and Diabetes (FIND) Study. *R.P. Igo[1], H. Abboud[2], S. Adler[3], B. Freedman[4], X. Guo[5], R. Hanson[6], W.H.L. Kao[7], C. Langefeld[8], J. Sedor[9], S.K. Iyengar[1, 10], The FIND Consortium.* 1) Epidemiology & Biostatistics, Case Western Reserve Univ, Cleveland, OH; 2) Medicine/Nephrology, University of Texas Health Science Center, San Antonio, TX; 3) Medicine/Nephrology and Hypertension, Harbor-UCLA Medical Center, Torrance, CA; 4) Nephrology, Wake Forest University, Winston-Salem, NC; 5) Medicine/Medical Genetics, Cedars-Sinai Medical Center, Los Angeles, CA; 6) National Institute of Diabetes and Digestive and Kidney Diseases, Phoenix, AZ; 7) Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 8) Biostatistical Sciences, Wake Forest University, Winston-Salem, NC; 9) Physiology and Biophysics, Case Western Reserve University, Cleveland, OH; 10) Genetics, Case Western Reserve University, Cleveland, OH.

Diabetic nephropathy (DN) is the leading cause of end-stage kidney disease in developed countries. The multi-center FIND Consortium aims to characterize genetic variants involved in the pathogenesis of DN and has collected qualitative and quantitative DN traits in four populations: African American (AA), American Indian (AI), European American (EA) and Mexican American (MA). A genomewide scan for common DNA copy-number polymorphisms (CNPs) was conducted on 2666 DN cases and 1655 controls with either type I or type II diabetes. DN was defined using stringent criteria, and controls had diabetes for at least 10 years and normal albumin excretion and estimated glomerular filtration rates $\geq$ 60 ml/min at recruitment. DNA samples were derived either from lymphoblastoid cell lines or directly from buffy coats from centrifuged blood samples. The overall distribution of genomewide CNV load was not significantly different between the two DNA types. DNA copy number (CN) was evaluated at 1320 locations across the human genome from single-nucleotide polymorphism (SNP) and monomorphic CN markers on the Affymetrix 6.0 chip, using the Canary and Birdseye programs in Birdsuite. A total of 826, 508, 618 and 774 CNPs passed quality control criteria in AA, AI, EA and MA, respectively. Association between DN and CN was evaluated via logistic regression, coding copy number as a linear predictor (on the logarithmic scale) and adjusting for age, sex and FIND recruitment center.

Logistic regression without adjustment for source of DNA revealed strong associations ($p < 0.001$) for CNP10227 (chrom. 1q, 244 Mb), CNP10119 (6p, 9 Mb) and CNP11328 (7q, 150 Mb) for AA, and for CNP2014 (14q, 44 Mb) for MA, with odds ratios ranging from 0.22 to 7.2 per DNA copy. When the analysis was restricted to DNA samples derived from cell lines, which account for about half of AA and 90% of AI, EA and MA, strong association was found instead for CNP1837 (12p, 33 Mb) for AI; CNP10572 (3q, 121 Mb), CNP898 (5q, 178 Mb) and CNP1648 (10q, 107 Mb) for EA; and CNP12350 (15q, 84 Mb) for MA. Association peaks from the entire sample coincided with strong single-SNP association peaks, whereas the strongest peaks from the cell-line DNA samples usually did not. These findings underscore the importance of DNA source in analyses on CN, and may elucidate the role of CNPs in the etiology of DN.

127

Imputation-based genotype calling in a worldwide sample of 15,000 individuals. *Y.Y. Lo[1], C. Sidore[1], Y. Li[3], B. Li[1], J. Li[4], C. Verzilli[2], K. Nangle[2], S.L. Chissoe[2], M.R. Nelson[2], M.G. Ehm[2], G. Abecasis[1], S. Zöllner[1,5].* 1) Department of Biostatistics, University of Michigan; 2) GlaxoSmithKline, Deptartment of Genetics, RTP, NC; 3) Department of Genetics, Department of Biostatistics, University of North Carolina; 4) Department of Human Genetics, University of Michigan; 5) Department of Psychiatry, University of Michigan.

Precise genotype calls are required to understand the contribution of rare polymorphisms to phenotypic variation. Generating high quality genotypes from next generation sequencing reads requires aggregating information across individuals in a sample. As sample sizes increase rapidly, methods need to be assessed on their ability to evaluate large sample sizes. Here we perform a comparative analysis on a dataset consisting of the exons of 202 genes sequenced to an average depth of 30x in a sample of 15,000 individuals. Eight of the 202 genes were sequenced in 993 individuals using conventional Sanger-sequencing as well as 150 sample duplicates, 30 trios, and 1 CEPH trio (duplicated on each plate) will facilitate assessment of genotype call quality. In addition, genome-wide SNP data was available for 10,500 individuals. After careful data cleaning and read mapping, we applied several methods for calling genotypes; We used both population based calling algorithms and linkage disequilibrium (LD) based methods. For calling genotypes based on LD patterns, we combined exon sequences from target genes with flanking genotype data in a 1Mb window. We compare the performance of LD-based calling with population based calling by calculating error statistics for each calling method dependent on multiple covariates such as the underlying coverage, GC content and sample size. Moreover, we imputed genotypes in an additional 11,878 individuals with genotype data but without sequence data at the target regions. By masking sequence information and applying the imputation procedure, we estimate imputation quality in all individuals. The results of our study provide important guidelines for the design and interpretation of resequencing based association studies.

128

Tests for association in next-generation sequence data with uncertain genotypes. *M.A. Schmidt, D.D. Kinnamon, E.H. Powell, G.W. Beecham, R.-H. Chung, E.R. Martin.* John P. Hussman Institute of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL.

With the EM algorithm-based SeqEM, we have a tool for calling bi-allelic genotypes based on next-generation sequence (NGS) data from unrelated individuals. These genotype calls are given based on estimated posterior probabilities for the 3 possible genotypes in an individual. In variants with higher read depth, these calls are usually nearly certain, with one genotype having nearly all probability mass. The situation is not as clear when the read depth is low and genotype calls are uncertain. Without having a certain genotype for each individual, it is not immediately clear how to analyze genotype-phenotype associations with commonly used methods. Standard approaches would treat the genotype with the largest posterior probability as certain and then compare the frequency of these genotypes in cases to controls. We compared this approach based on the most likely genotype (MLG) to two novel approaches: a likelihood ratio test (LRT) and permutation test (PT). The LRT is based on the SeqEM likelihood for the observed NGS data, a standard mixture of binomials. This test compares the maximized likelihood in which genotype frequencies can vary by affection status to the maximized likelihood in which genotype frequencies are constrained to be the same regardless of affection status using the asymptotic chi-square distribution of the appropriately transformed likelihood ratio. The PT assumes only that the NGS data are identically distributed in cases and controls under the null. The test statistic is the average number of variant calls among all reads in cases and in controls, the null distribution of which is obtained by permuting affection status multiple times. We conducted simulations with MAF equal or greater 0.05 to evaluate the performance of the MLG approach compared to the LRT and PT approaches. We found that all 3 methods produced the correct type I error when simulating a read depth of 5 or greater. In general, both the PT approach and LRT had power as high as or higher than the MLG approach. Differences were more apparent at low read depths. At higher read depths, individual genotypes become increasingly certain, and the power of the LRT and MLG methods is similar. Incorporating genotype call uncertainty using the aligned read data or genotype posterior probabilities in tests for disease association can provide greater power than treating called genotypes based on the maximum posterior probabilities as certain.

129

Powerful strategy to assess the effect on disease of rare variants in large genetic regions. *I. Ionita-Laza[1], J. Buxbaum[2], N. Laird[3], C. Lange[3].* 1) Dept Biostatistics, Columbia Univ, New York, NY; 2) Mount Sinai School of Medicine, New York, NY; 3) Dept Biostatistics, Harvard Univ, Boston, MA.

Rapid advances in sequencing technologies set the stage for the large-scale medical sequencing efforts to be performed in the near future, with the goal of assessing the importance of rare variants in complex diseases. The discovery of new disease susceptibility genes requires powerful statistical methods for rare variant analysis. The low frequency and the expected large number of such variants pose great difficulties for the analysis of these data. We develop here a robust and powerful likelihood-based approach to study the role rare variants may play in affecting susceptibility to complex traits. The method is based on a new statistic to assess whether rare variants occur in cases at significantly higher frequency than in controls (or vice versa). For large genetic regions such as gene-sets we describe a backward exclusion procedure whereby each gene in the set is evaluated one at a time by excluding it from the set. A unique feature of the proposed methodology is that, although it is an overall test assessing a possibly large number of rare variants simultaneously, the disease variants can be both protective and risk variants, with small decreases in statistical power when both types of variants are present. Using coalescent-based simulations, we show that this approach can be powerful under complex and general disease models, and in large genetic regions such as pathways, where the proportion of disease susceptibility variants may be very small (less than 1%). Comparisons with previously published tests (Li and Leal 2008; Madsen and Browning 2009) on simulated data show that the proposed approach has improved power over these existing methods. We show applications to two real datasets. A first application to the ENCODE III data shows a significant enrichment in rare variants for the YRI population compared with the CEU population, as expected. A second application to a recently published study on Type 1 Diabetes (Nejentsev et al. 2009) finds rare variants in a set of ten candidate genes to be protective against Type 1 Diabetes.

130

Testing for an unusual distribution of rare variants. *B.M. Neale[1,2], M.A. Rivas[1,2], B.F. Voight[1,2], D. Altshuler[2,3,4], B. Devlin[5], M. Orho-Melander[6], S. Katherisan[1,2,7,8], S. Purcell[2,9], K. Roeder[10], M.J. Daly[1,2].* 1) Richard Simches Bldg CPZN 6818, Ctr Human Gen Res, Boston, MA, 02114; 2) The Broad Institute of Harvard and MIT, Cambridge, MA, 02142; 3) Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA; 4) Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA; 5) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, 15213; 6) Department of Clinical Sciences Malmö, Diabetes and Cardiovascular Diseases, Genetic Epidemiology CRC, University Hospital Malmö SE-205 02 Malmö, Sweden; 7) Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA 02114; 8) Department of Medicine, Harvard Medical School, Boston, MA 02114; 9) Psychiatric and Neurodevelopmental Unit, Massachusetts General Hospital, Boston, MA, 02114; 10) Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania, 15213.

Background: Technological advances make it possible to use high-throughput sequencing as a primary discovery tool of medical genetics, specifically for assaying rare variation. A key analytic challenge of the sequencing approach is that the influence of very rare variants on complex phenotypes can only be evaluated effectively as a group. A further complication is that any given rare variant may be phenotypically neutral, risk, or protective. Methodology/Principal Finding: We propose here the C-alpha test statistic as a novel approach for testing the presence of this mixture of effect across the set of rare variants. Unlike existing proposed burden tests, C-alpha, by testing the variance rather than the mean, maintains consistent power when the target set contains both risk and protective variants. Additionally, C-alpha can be applied to summary allele counts such as might be obtained from meta-analysis and pooled experiments. Conclusion/Signficance: Through simulations and analysis of case/control data, we demonstrate good power relative to existing methods that assess the burden of rare variants in individuals. Because this statistic is closed form, it enables rapid calculation essential for impending whole exome and whole genome sequencing studies of large numbers of individuals.

131

Association Studies for Next-generation Sequencing. *L. Luo, E. Boerwinkle, M. Xiong.* Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, 1200 Herman Pressler St, Houston, TX 77030.

Genome-wide association studies (GWAS) have become the primary approach for identifying genes with common variants influencing complex diseases. Despite considerable progress that GWAS have made in identifying hundreds of putative disease loci, these newly discovered loci account for only a small proportion of disease heritability. This implies that the common variations identified by GWAS are unlikely to explain the majority of phenotypic variations of common diseases. A potential source of the missing heritability is the contribution of rare variants. Next-generation sequencing technologies will detect millions of novel rare variants, but these technologies have three defining features: identification of large number of rare variants, a high proportion of sequence errors, and large proportion of missing data. These features raise challenges for testing the association of rare variants with phenotypes of interest. To meet the great challenges raised by next-generation sequencing, instead of modeling the genome as a few separated individual loci, we use a genome continuum model and functional principal components as a general principle for developing novel and powerful association analysis methods designed for resequencing data. To evaluate the performance of functional principal component analysis (FPCA)-based statistics, we use simulations to calculate the type I error rates and the power of five alternative statistics: FPCA-based statistic, the generalized T2, Collapsing method, CMC method and individual test. We also examine the impact of sequence errors on the type I error rates of the five test statistics based on 1000 genome project data. We also apply the five statistics to published resequencing dataset from ANGPTL4 in the Dallas Heart Study. We report that FPCA-based statistics have higher power to detect the association of rare variants and stronger ability to filter sequence errors than the other four methods. Finally, comprehensive simulations studies based on 1000 genome project data showed that a large proportion of the loci being tested will yield wrong type I error rates, when we apply the individual test to testing the association of rare variants. Our work represents a paradigm shift from the current single marker association analysis toward collectively testing for association of multiple rare variants in sequence-based association analysis.

132

Strengths and limitations of follow-up imputation to investigate variants originally identified through deep-sequencing experiments. *M.G. Ehm[1], L. Li[1], K. Song[2], S.-A Bacanu[1], C. Cox[3], J. Aponte[1], J.K. Mitchell[1], S.L. Chissoe[1], D. Fraser[1], D. Briley[1], X. Yuan[2], C. Verzilli[3], J. Shen[1], K. Nangle[1], P. Vollenweider[5], G. Waeber[5], L.R. Cardon[2], V.E. Mooser[2], D.M. Waterworth[2], J.C. Whittaker[3], M.R. Nelson[1], Y. Li[4].* 1) Genetics, GlaxoSmithKline, Res Triangle Park, NC; 2) Genetics, GlaxoSmithKline, Swedeland, PA; 3) Genetics, GlaxoSmithKline, Harlow, UK; 4) University of North Carolina, Chapel Hill, NC; 5) CHUV,Lausanne, Switzerland.

Many deep resequencing experiments have been initiated to understand the contribution of rare functional variation in human traits. While many identified gene or variant trait associations will require follow-up sequencing, the limits of investigating these initial results utilizing genotype imputation is unclear. We investigated the characteristics of genotype imputation for exon sequence data using a reference set of thousands of individuals, Affymetrix 500K GWAS data and follow up genotype and sequence data. Whilst imputation into samples with GWAS data from rare variation will not work for private mutations, it could work for low frequency variants. Current imputation methods have been shown to be robust for variants with minor allele frequencies (MAF)>5%. As the number of available reference samples with sequence data increases due to the addition of 1000 Genomes Project data and deep resequencing studies, the lower limit of resolution and optimal sample selection for imputation are unclear. As the density of the variants changes and becomes more variable, the characteristics of variants that can be imputed and the required back-bone genotype data to produce reliable imputed genotypes are unknown. We performed Sanger sequencing to investigate exonic variants within 8 genes (including 10248 coding bp) in 2000 subjects from the CoLaus population-based study (Lausanne, Switzerland). We identified 515 variants, including 292 (56%) singletons, 437 (85%) variants with MAF<0.5%, and 41 (8%) with MAF >5%. All coding and selected non-coding variants were typed using a FRET-based genotyping methodology in the 2000 sequenced and an additional 4000 CoLaus samples resulting in 130 markers with reliable data. Affy 500K data was also available in these 6000 subjects. Furthermore, single molecule sequencing of captured DNA is in progress for these 8 genes for 1447 of 6000 CoLaus samples. To determine the limitations of imputing variants derived from exon sequencing, we used MaCH to impute variants identified in the 2000 resequenced samples into the 4000 additional samples. Comparisons with genotypes showed heterozygote accuracy of 86% for well-imputed variants (r^2>0.3) when at least 20 copies were observed in the training sample (MAF≥0.5%). Newly available resequence data will enable more accurate estimates of imputation accuracy - especially for MAF<0.5%. Additionally, we will discuss the required back-bone genotype data and optimal sample selection.

133

Comparing total and allelic expression for mapping *cis*-acting polymorphims. *M.F. Santibanez-Koref, M.S. Cunnington, B. Keavney, H.J. Cordell.* Institute of Human Genetics, University of Newcastle, Newcastle upon Tyne, United Kingdom.

There is increasing evidence that modulation of transcription, in particular through changes acting in *cis*, plays an important role in modifying disease susceptibility. The effects of sequence variation on expression can be assessed by comparing the total levels of a transcript, without differentiating the contributions of each allele, in individuals with different genotypes at the polymorphism of interest, and treating expression levels as a quantitative trait (expression quantitative trait locus; eQTL). Proximity to the target gene is seen as evidence for a *cis* effect. Expression is also affected by influences acting through diffusible factors in *trans*. These include environmental and genetic factors and affects transcripts from both alleles. An alternative approach for mapping *cis* acting loci on autosomes is to compare the relative expression of the two alleles within an individual. This can be quantified as the allelic expression ratio (AER). We compared both approaches using published data on allelic and total expression of CDKN2B and for antisense transcripts encompassing this locus (CDKN2BAS exons 1 to 2). We estimated the proportion of *cis* acting variance to be higher for CDKN2BAS (point estimate 20%) than for CDKN2B (5%). One of the drawbacks of mapping using AER is that it requires individuals that are heterozygous for a transcribed polymorphism. We present an approach that can use several transcribed SNPs even in the absence of linkage disequilibrium. In general, several sites can be independently associated with differences in transcription. In eQTL analysis the effects of several SNPs can be considered and the association with particular polymorphisms can be assessed by adjusting for the effects of others. We show that a similar approach can be used for AER analysis. Mapping cis acting sites by eQTL and AER analysis, does not need to produce the same results. Tight posttranscriptional regulation could keep total mRNA levels constant even when the contributions of each allele vary. SNPs surrounding CDKN2B show a correlation between estimated effect sizes using eQTL and AER analysis (R2= 0.87). The significance of the p-values is higher for AER analysis. Simulations show that increased disequilibrium, transcribed marker heterozygosity and extent of trans acting variation increases the power of AER compared to that of eQTL analysis. AER and total expression data can be integrated allowing to estimate *cis* and *trans* effects separately.

134

An Efficient LD Based Variant Calling and Phasing Method for Next Generation Sequencing in Trios. *W. Chen[1], B. Li[1], Y. Li[2], G. Abecasis[1].* 1) Dept Biostatistics, Univ Michigan, Ann Arbor, MI; 2) Dept Genetics, Dept Biostatistics, Univ North Carolina, Chapel Hill, NC.

Following successful genome-wide association studies, next generation sequencing provides a deeper catalog of human genetic variation and possibilities for direct detection of causal variants. Accurate and efficient methods for variant calling and haplotype inference are essential to follow up analysis. However, to date, there are no published methods on variant calling from sequence data of family members. We propose an efficient and accurate variant calling method for shotgun sequencing data of trios. Our method incorporates both linkage disequilibrium (LD) patterns and constrains due to the trio family structure through a hidden Markov model (HMM). Our model is built upon a HMM previously proposed for genotype imputation, which takes advantage of similar stretches of chromosomes shared among apparently unrelated individuals. Our method will facilitate variant calling and haplotype inference for many ongoing sequencing projects including trios. In addition, it allows us to systematically explore the potential advantages of sequencing family members over sequencing unrelated individuals. We simulate shotgun sequencing data in genotype likelihood format (GLF) for 60 trios, 120 and 180 unrelated samples at various depths with different sequencing error rates. Our simulations show that sequencing trios can have similar or even higher variant calling rate than the same number of unrelated individuals at low depth 1X and 2X. For depth 4X and 8X, although unrelated samples has more power to detect variants, trios have advantage of higher sensitivities and specificities at heterozygous sites, which is crucial for individual genotype call of rare variants. Generally, trios have higher calling accuracy across different frequency spectrum. In addition, trios can greatly increase haplotyping accuracy. For example, the average number of flips is ~20% of number based on a design of unrelated individuals. At depth 8X and above, the gains of the trio design are limited and the design of unrelated samples are more preferable. The method can be naturally extended to the design including parent offspring and has potential applications in general pedigree structure. We are applying our method to real data, combined from the HapMap Project and the 1000 Genomes Project. We will also apply our method to ongoing SardiNIA sequencing project. Our method is implemented in C++ and will be available for public soon.

135

Sampling strategies for rare variant tests. *S. Zöllner[1,2].* 1) Dept Biostatistics, Univ Michigan, Ann Arbor, MI; 2) Dept Psychiatry, Univ Michigan, Ann Arbor, MI.

Recent advances in sequencing technology allow assessing the contribution of rare variation to the heritable risk of common disorders. Methods have been developed to combine functional rare variants across a gene and compare this aggregate statistic between cases and controls. Existing samples provide a large number of possible cases. However, as it is presently still costly to sequence many individuals, it is vital to identify samples that result in powerful tests. Testing randomly selected cases has low power, especially after Bonferroni correction. Power can be increased by select cases that are highly likely to carry risk variants. Here we consider the strategy of identifying cases likely to carry risk variants based on the affection status of family members. Rare variants that contribute to the heritability of a disease co-segregate among affected family members. Hence selecting cases that have affected family members will increase the probability of cases carrying a rare risk variant and will thus increase the power of rare variant tests considerably. We quantify this power gain and provide criteria for sample selection under different models of inheritance based on realistic effect sizes estimated from the absence of linkage signals for common diseases. We compare the sample sizes required to achieve 80% power between samples of cases with one affected first degree family member and samples of random cases. To achieve the same power the random case sample has to be 2-5 times the sample size of cases ascertained to have an affected family member. We also consider re-sequencing studies that target candidate regions. For single regions, a family-informed sample can be further improved by selecting cases that share the target segment with an affected family member. Selecting cases conditional on sharing 2 chromosomes with an affected family member can result in an increase in power equivalent to sequencing 10 times as many random cases. In summary, while deep sequencing large number of individuals is expensive, considerable gains in power can be achieved by carefully selecting case samples based on their evidence of carrying risk variants. Disease inheritance patterns provide such evidence and should thus inform sample selection for re-sequencing.

136

Sequence similarity-based association tests under genetic heterogeneity and population stratification. *X. Qin, S. Schmidt, E.R. Hauser.* Ctr Human Gen, Duke Univ, Durham, NC.

Sequence similarity measures have been used to test trait association or to evaluate and adjust for population stratification. Here we develop a sequence similarity measure for case-control data, and then apply it as covariate in our previously published method Ordered Subset Analysis for Case-Control data (OSACC) (Qin et al, Genetic Epidemiology, 2010). The fundamental model is that sequence variants should be more similar around a trait-related variant in an evolutionarily-similar case subgroup within a heterogeneous dataset. First to develop the sequence similarity measure we regress all sequence variants in a defined window around the potential trait-related variant (test variant) on the test variant itself, followed by in-phase calculation of regression residuals for each allele (a calculation similar to estimating linkage disequilibrium coefficients). We do this for each sequence variant in turn to obtain residuals for each variant. Second, we perform principal components analysis using the residuals calculated in step one to get coordinates for all individuals using a small number of principal components defined by the scree plot. The Mahalanobis distance weighted by the variance is used when more than one principal component used. The coordinates calculated in this way are the measures of similarity of the sequence. Sequences that are very similar will have similar values of the coordinates. Third, the coordinates for each individual are used as a covariate in OSACC to evaluate the association in the presence of genetic heterogeneity and to identify the OSACC subset that maximizes the evidence for association. We evaluated the performance of the sequence similarity method under two null hypotheses (with and without marker-marker LD) and several alternate hypotheses. The type I error was not inflated and the power to detect association depends on how well the sequence similarity distinguishes the subset with the trait-related variant, with increases in power of up to 25%; relative to the standard trend test. As in our previous studies the OSACC covariate test is robust to population stratification. We demonstrate the application of the method to existing sequence data sets.

137

Identification of 2 novel genes involved in Treacher Collins syndrome. *J.G. Dauwerse[1], J. Dixon[9], S. Seland[2], C.A. Ruivenkamp[1], A. van Haeringen[1], L.H. Hoefsloot[3], C. Daumer-Haas[4], R. Maiwald[5], C. Zweier[6], A.J.M. Hoogeboom[7], U. Hehr[8], D.R. Lohmann[2], M.J. Dixon[9], M.H. Breuning[1], D. Wieczorek[2].* 1) Centre for Human and Clinical Genetics, Leiden Univiversity Medical Centre, Leiden, The Netherlands; 2) Institut für Humangenetik, Universitätsklinikum, Essen, Germany; 3) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 4) Pränatal-Medizin München, München, Germany; 5) Medizinische Versorgungszentrum für Laboratoriumsmedizin, Mikrobiologie und Humangenetik, Mönchengladbach, Germany; 6) Humangenetisches Institut, Universitätsklinikum Erlangen, Erlangen, Germany; 7) Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands; 8) Zentrum für Humangenetik am Universitätsklinikum Regensburg, Regensburg, Germany; 9) Faculty of Medical and Human Sciences, Manchester Academic Health Sciences Centre, University of Manchester, Manchester, United Kingdom.

Treacher Collins Syndrome (TCS) is an autosomal dominant disorder of craniofacial development which is characterized by bilateral downward slanting palpebral fissures, colobomas of the lower eyelids with a paucity of eyelashes medial to the defect, hypoplasia of the facial bones, particularly the mandible, cleft palate, deformity of the external ear, atresia of the external auditory canal and bilateral conductive hearing loss. Although mutation detection rates as high as 93% have been reported a subset of patients in which the causative mutation has not been identified remains. To investigate the genetic basis of TCS in a 3-year-old boy who was negative for a *TCOF1* mutation, we performed genome-wide, micro-array analysis using an Affymetrix GeneChip 262K *Nsp*I SNP array. We identified a *de novo* 156 kb deletion within chromosome 13q12 that resulted in the deletion of 2 genes. One of the two TCS candidate genes (*TCSC2*) was analyzed in 259 TCS patients who did not exhibit a *TCOF1* mutation and 18 heterozygous mutations were detected. In 5 families the identified mutation co-segregated with the phenotype; however in 5 families, cases of non-penetrance were found. Non-penetrance has been previously reported in TCS families with *TCOF1* mutations. Subsequently, we sequenced another *TCSC2*-like gene, *TCSC3*, in the group of 259 patients. In one patient with typical TCS, we detected a stop mutation in exon 9; however, a missense mutation in exon 8 of the same gene was also detected in the same patient. Similarly, a second unrelated patient showed mutations in 2 different exons of *TCSC3*. In both these cases, the independent mutations were inherited from different parents, both of whom were unaffected, indicating autosomal recessive inheritance of the syndrome. These findings confirm genetic heterogeneity of TCS and identify two additional genes involved in the pathogenesis.

138

X- Linked Genetic Factors in DMD associated with Oral Clefts. *P.J. Patel[1], T.H. Beaty[1], J.C. Murray[2], M.L. Marazita[3], R.G. Munger[4], I. Ruczinski[1], J.B. Hetmanski[1], K.Y. Liang[1], T. Wu[1], M. Rose[1], R.A. Redett[5], S.C. Jin[1], A. Wilcox[6], R.T. Lie[7], Y.H. Wu-Chou[9], H. Wang[10], X. Ye[8,11], S. Huang[12], V. Yeow[13], S.S. Chong[14], S.H. Jee[15], B. Shi[16], K. Christensen[17], K. Doheny[18], E.W. Pugh[18], H. Ling[18], A.F. Scott[5].* 1) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 2) Department of Pediatrics, University of Iowa, Iowa City, IA; 3) University of Pittsburgh, School of Dental Medicine, Pittsburgh, PA; 4) Utah State University, Logan, UT; 5) Johns Hopkins University, School of Medicine, Baltimore, MD; 6) NIEHS/NIH, Durham, NC; 7) University of Bergen, Bergen, Norway; 8) Mount Sinai Medical School, New York, NY; 9) Chang Gung Memorial Hospital, Taoyuan, Taiwan; 10) Peking University Health Science Center, Beijing, China; 11) Wuhan University, Wuhan, China; 12) Peking Union Medical College, Beijing, China; 13) KK Women's & Children's Hospital, Singapore; 14) National University of Singapore, Singapore; 15) Yonsei University, Epidemiology & Health Promotion, Seoul, Korea; 16) West China School of Stomatology, Chengdu, China; 17) University of Southern Denmark, Odense, Denmark; 18) Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD.

Oral clefts are known to be influenced by both genetic and environmental risk factors, and the sex ratio differs between CP and CL/P. As part of an international consortium, 2,463 case-parent trios were collected from 9 countries (13 collection sites) for a genome wide association study of nonsyndromic oral clefts. Among male probands, 16.4% had cleft palate (CP), 21.5% had cleft lip (CL), 61.8% had cleft lip and palate (CLP), and 0.3% had an unknown cleft type. Among female probands, 31.3% had CP, 22.9% had CL, 45.6% had CLP, and 0.1% had an unknown cleft type. Using 14,486 SNPs on the X chromosome, transmission disequilibrium tests (TDT) for individual SNPs was conducted, stratified by cleft type and racial group (Asian and European ancestry). The analysis in the total CL/P group yielded nominally significant results but none reaching conventional genome wide levels of significance. Four SNPs in the Duchenne muscular dystrophy (DMD) gene, the largest known gene in the human genome, gave p-values ranging from 2x10-5 to 3.09x10-6 where the most significant SNP was rs5928207. When samples were stratified by race, it became clear this statistical signal was strongest among Asian CL/P trios. In the Asian group the most significant SNP was rs6631759 giving a p-value of 7.84X10-6 which had a p-value of 9.29x10-6 in the total CL/P trio group. Sliding window analysis using 14 SNPs adjacent to rs6631759 failed to increase the statistical significance. Amongst European CL/P trios, the most significant SNP in DMD was rs1800280, which gave a p-value of 0.0016. To further investigate these results, a joint analysis of the most significant 25 SNPs in DMD in the Asian CL/P group was conducted. The initial analysis using all combinations of these significant SNPs in DMD yielded a 4-SNP "haplotype" with a p-value=4.99x10-8, but these four SNPs spanned almost the entire length of the DMD gene (~2 Mb). A permutation test (n=5000) was then conducted, which yielded an adjusted p-value=0.0002. Sliding window analyses using SNPs adjacent to those 4 SNPs, respectively, did not improve their single SNP p-values. The DMD gene has not been implicated in any previous study of oral clefts. Our results suggest that some SNPs in DMD gene maybe associated with CL/P in Asian populations, but they appear less important in populations of European ancestry.

139

FAF1, the First Gene Associated with Cleft Palate. *M. Ghassibe[1], L. Desmyter[1], T. Langenberg[2,3], F. Claes[2,3], O. Boute[4], B. Bayet[5], Ph. Pellerin[6], L. Backx[7], K. Hermans[2,3], P. Brouillard[1], J. Murray[8], M.A. Mansilla[8], S. Imoehl[8], A. Mangold[9], M. Noethen[9,10], S. Nowak[9], K.U. Ludwig[10], M. Rubini[11], C. Baluardo[11], M. Ferrian[11], P. Mossey[12], J. Hecht[13], G. François[5], N. Revencu[1,14], R. Vanwijck[5], J.R. Vermeesch[7], H.A. Poirel[1,14], P. Carmeliet[2,3], M. Vikkula[1].* 1) Laboratory of Human Molecular Genetics, de Duve Institute, Université catholique de Louvain, Brussels, Belgium; 2) Vesalius Research Center, K.U.Leuven, Leuven, Belgium; 3) Vesalius Research Center, VIB, Leuven, Belgium; 4) Centre de Génétique, CHU de Lille, Lille, France; 5) Centre Labiopalatin, Service de Chirurgie Plastique, Cliniques universitaires Saint-Luc, Brussels, Belgium; 6) Service de Chirurgie Plastique et Reconstructive, CHU de Lille, Lille, France; 7) Center for Human Genetics, K.U.Leuven, Leuven, Belgium; 8) Department of Pediatrics, University of Iowa, Iowa City, IA, USA; 9) Institute of Human Genetics, University of Bonn, Bonn, Germany; 10) Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany; 11) Medical Genetics Unit, Department of Experimental and Diagnostic Medicine, University of Ferrara, Ferrara, Italy; 12) Orthodontic Unit, Dental Hospital & School, University of Dundee, Dundee, UK; 13) University of Texas Medical School at Houston; 14) Center for Human Genetics, Cliniques universitaires Saint-Luc, Université catholique de Louvain, Brussels, Belgium.

Cranial neural crest (CNC) is a multipotent migratory cell population that gives rise to most of the craniofacial bones. An intricate regulatory network mediates CNC formation, epithelial-mesenchymal transition, migration along distinct paths and differentiation. Errors in these processes lead to craniofacial abnormalities, including cleft lip and palate. In this study, the breakpoint of a 46,XY,t(1;2)(p34;q33) translocation present in a family with hereditary Pierre Robin sequence (PRS), is characterized. Fas-Associated Factor-1 (FAF1) is shown to be disrupted and its expression is decreased in a PRS family with an inherited balanced translocation. Association and relative risk calculations on 7597 individuals demonstrate the locus is strongly associated with CPO but not with CL/P, and shows an increased relative risk. In situ hybridizations on murine embryos show that Faf1 is highly expressed in mouse and zebrafish cartilages during embryogenesis. Moreover, knockdown of zebrafish faf1 leads to pharyngeal cartilage defects, due to a failure of CNC to differentiate into and express cartilage specific markers, such as sox9 and col2a1. This phenotype is rescued by administration of faf1 mRNA. In conclusion, FAF1 is the first gene shown to predispose to cleft palate. It is a newly recognized regulator of CNC differentiation, necessary for lower jaw development. Its disruption results in craniofacial anomalies across species.

140

Mutation of *Rubie*, a novel non-coding RNA located upstream of *Bmp4*, causes vestibular malformation in mice. *K.A. Roberts[1,2], V.E. Abraira[3], A.F. Tucker[3], L.V. Goodrich[3], N.C. Andrews[1,2].* 1) Department of Pharmacology and Cancer Biology, Duke University, Durham, NC; 2) Graduate Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA; 3) Department of Neurobiology, Harvard Medical School, Boston, MA.

The vestibular apparatus of the vertebrate inner ear is an intricate structure responsible for sensing motion and position of the head. Within the inner ear, angular head movements are detected by three fluid-filled semicircular canals. Malformation of these canals disrupts the sense of balance and frequently causes circling behavior in mice. The *Epistatic circler* (*Ecl*) is a complex mouse mutant originally discovered in a multi-generation intercross population derived from wildtype SWR/J and C57L/J mice. In *Ecl* mice, circling is caused by the epistatic interaction of a recessive, SWR-derived locus on chromosome 14 and a dominant, C57L-derived locus on chromosome 4. Prior to this study, neither of the underlying genes had been identified. We developed a mouse chromosome substitution strain (CSS-14) that inadvertently recapitulated *Ecl*, and used it to identify one of the *Ecl* genes. CSS-14 carries an SWR/J chromosome 14 on a C57BL/10J genetic background and, like *Ecl*, exhibits circling behavior due to bilateral malformation of the lateral semicircular canal. We confirmed that CSS-14 circling is correlated with SWR homozygosity at the recessive *Ecl* locus on chromosome 14 and utilized CSS-14 mice to identify the causative gene by positional cloning. Our 445 Kb candidate interval is located upstream of bone morphogenetic protein 4 (*Bmp4*) and contains an inner ear-specific, long non-coding RNA (ncRNA) that we have designated *Rubie* (RNA upstream of *Bmp4* expressed in inner ear). *Rubie* is a spliced, polyadenylated ncRNA transcript that is normally expressed in developing semicircular canals. However, sequence and expression analysis revealed that the SWR/J allele of *Rubie* is disrupted by an intronic endogenous retrovirus that causes aberrant splicing and premature polyadenylation of the transcript. *Rubie* lies in the conserved gene desert upstream of the *Bmp4* transcriptional start site, within a region previously shown to contain a distant *cis*-regulatory element important for inner ear expression of *Bmp4*. The expression patterns of *Rubie* and *Bmp4* are nearly identical in developing inner ears and heterozygous deletion of *Bmp4* produces an inner ear phenotype indistinguishable from that of *Ecl* mice and CSS-14 circlers. Based on these results, we propose that transcription of *Rubie* regulates inner ear expression of *Bmp4*, and that aberrant *Bmp4* expression contributes to the *Ecl* phenotype.

141

HPPD: An Apparent New Autosomal Dominant Disorder Involving Hypertelorism, Preauricular sinus, Puntal Pits and Deafness Mapping to Chromosome 14q31. *Y. Lacassie[1], S. Sampath[2], B.J. Keats[2].* 1) Dept Ped/Div Clin Gen, LSU Hlth Sci Ctr, and Children's Hospital New Orleans, LA; 2) Dept Genetics, LSU Hlth Sci Ctr New Orleans, LA.

We report a three-generation family with an apparent novel autosomal dominant disorder whose major features include hypertelorism, preauricular sinus, deafness, and punctal pits with lacrimal-duct obstruction. The family history revealed multiple affected members with variable phenotypic expression. We presented this family at the 57th ASHG Annual Meeting in 2007 with a possible differential diagnosis of Branchio-oto-renal syndrome. At the ACMG Annual Clinical Genetics Meeting in 2008 we presented the data on this family indicating the elimination of linkage to known candidate genes EYA1, SIX1, and SIX5 and the identification of a novel locus at 14q31.1-3 (LOD = 3.14) using SNP based whole-genome parametric linkage analysis. In this meeting we present the progress we have made following the linkage mapping. The SNP based linkage mapping results were confirmed using microsatellite markers identifying the proximal and distal breakpoint inteval of 8.3 Mb. This large interval contained very few genes (6 RefSeq genes and two unknown transcripts). Since none of the genes within the critical interval were biological candidates, we sequenced all exons of the genes and transcripts. We found no mutations within the exons and the conserved splice-donor and splice-acceptor sites. A custom aCGH experiment was performed with oligos spanning the 20 Mb across the critical interval to identify any segregating deletions or duplications. No segregating copy-number changes were found. Since variation in promoter sequence could affect gene expression levels we evaluated 3 Kb of promoter region for each of the genes within the critical interval. Sequencing a total 24 Kb of promoter regions revealed a novel polymorphism 1249 bp upstream of the SELIL start codon. This polymorphism segregated with the disease haplotype resulting in a 20% reduction in the expression of reporter constructs. The identification of this family with a distinctive clinical presentation from BOR and other reported syndromes and linkage to a novel locus at 14q31 support the existence of a new syndrome of the brachial cleft.

142

TARGETED 'NEXT GENERATION' SEQUENCING IN ANOPHTHALMIA PATIENTS REVEALS AN *OTX2* MUTATION. *A. Slavotinek[1], J. Li[2], T. Bardakjian[3], L. Tonkin[4], A. Schneider[3], E. Sherr[2], N. LopezJimenez[1].* 1) Dept Pediatrics, U585P, Univ California, San Francisco, San Francisco, CA; 2) Dept. of Neurology, UCSF, San Francisco, CA; 3) Clinical Genetics Division, Albert Einstein Medical Center, Philadelphia, PA; 4) Vincent J. Coates Genomic Sequencing Laboratory, University of California, Berkeley, CA.

Anophthalmia is found in 1 in 5,000 to 10,000 individuals and is a devastating birth defect because of severe visual impairment. *SOX2* is mutated in 10-20% of patients with bilateral anophthalmia or microphthalmia (A/M). However, the remaining genes implicated in A/M, such as *OTX2*, are each mutated in only 2-3% of patients, meaning that over 60% of patients do not receive a genetic diagnosis after molecular testing. In addition, screening of the known A/M genes is rarely completed because of the costs and labor involved in sequencing multiple genes. 'Next generation' sequencing is a novel strategy for identifying sequence variations in birth defects. We used next generation sequencing as a cost and time saving methodology for sequencing genes that are known to cause A/M in humans or mice. We used 22 PCR fragments to amplify 57 coding exons from 11 genes (*SOX2*, *OTX2*, *PAX2*, *CHX10*, *FOXE3*, *CRYBA4*, *PAX6*, *SIX3*, *MAF1*, *BMP4* and *VAX1*). We used the Illumina GA2 sequencer with paired end module for next generation sequencing and analyzed the resultant sequence with Maq software and SAMtools. Using DNA from a normal control of Hispanic ethnicity as a test, we identified 90 sequence variants: 2 known exonic SNPs, 1 unreported, synonymous exonic SNP in *CRYBA4* (c.6348A>T predicting p.Thr108Thr; verified by Sanger sequencing and found in 5% of controls), 38 known, non-coding SNPs and 49 non-coding sequence variants of unknown significance. We then sequenced DNA samples from 8 A/M patients in one experiment and verified exonic alterations by Sanger sequencing. In a patient with bilateral anophthalmia and feeding difficulties, we found c.8175C>T in *OTX2*, predicting p.Gln112X and a premature stop codon just after the homeodomain of the 297 amino acid protein. This mutation is consistent with the loss of function mutations observed in *OTX2* in A/M patients, and was found independently by another researcher sharing the same DNA sample. We were therefore able to use next generation sequencing to screen multiple genes efficiently in A/M patients and were able to identify an *OTX2* mutation. However, we did not find causative mutations in the other patients, emphasizing the genetic heterogeneity of A/M and the need for further gene discovery after the exclusion of known candidate genes.

143

A novel cerebello-ocular syndrome with abnormal glycosylation due to abnormalities in dolichol metabolism. *E. Morava[1], A. Brouwers[2], R. Wevers[3], V. Cantagrel[4], M. Jongmans[2], C. Ravenswaaij-Arts[6], B. Ng[5], H. van Bokhoven[2], D. Babovic-Vuksanovic[7], J. Gleeson[4], D. Lefeber[3], L. Al-Gazali[8].* 1) Dept Pediatrics, IGMD, Radboud University Nijmegen Medical Center, Nijmegen, Netherlands; 2) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Nederlands; 3) 3Department of Laboratory Medicine, Laboratory for Genetic, Endocrine and Metabolic Disease, Nijmegen, The Netherlands; 4) Howard Hughes Medical Institute, Department of Neurosciences and Pediatrics, University of California, San Diego, USA; 5) Sanford Children's Health Research Center, Burnham Institute for Medical Research, La Jolla, USA; 6) Department of Genetics, Groningen University, Groningen, The Netherlands; 7) Department of Medical Genetics, Mayo Clinic College of Medicine, Rochester, USA; 8) 6Departments of Pediatrics and Pathology, United Arab Emirates University, Faculty of Medicine and Health Sciences, Al Ain, United Arab Emirates.

Cerebellar hypoplasia and slowly progressive ophthalmological symptoms are common features in patients with Congenital Disorders of Glycosylation type I. In a group of CDG-I patients with unknown etiology we have previously described a distinct phenotype with severe, early visual loss and variable eye malformations, including optic nerve hypoplasia, retinal coloboma, congenital cataract and glaucoma. Some of the symptoms overlapped with the phenotype in other CDG-I subtypes, like vermis hypoplasia, anemia, ichtyosiform dermatitis, liver dysfunction and coagulation abnormalities. We recently identified pathogenic mutations in the SRD5A3 gene, encoding steroid 5α-reductase type 3, in a group of patients who presented with this particular phenotype and a common metabolic pattern. Here we report on the clinical, genetic and metabolic features of 12 patients from nine families with cerebellar ataxia and congenital eye malformations due to steroid 5α-reductase type 3 defect. This enzyme is necessary for the reduction of polyprenol to dolichol, the lipid anchor for N-glycosylation in the endoplasmatic reticulum. Dolichol synthesis is an essential metabolic step in protein glycosylation. The current defect leads to a severely abnormal glycosylation state already in the early phase of the N-glycan biosynthesis pathway in the endoplasmic reticulum. We detected high expression of SRD5A3 in fetal brain tissue, especially in cerebellum, coherent with the finding of the congenital cerebellar malformations. Based on the overlapping clinical, biochemical and genetic data in this large group of patients with Congenital Disorders of Glycosylation, we define a novel syndrome of cerebellar ataxia associated with congenital eye malformations due to a defect in dolichol metabolism.

144

Phenotypic Features of Carbohydrate Sulfotransferase 3 (CHST3) Deficiency in 27 Patients: Congenital Dislocations and Vertebral Changes as Principal Diagnostic Features. *A. Superti-Furga[1], E. Lausch[1], A. Rossi[2], A. Megarbane[3], D. Sillence[4], A. Aytes[5], R. Mendoza-Londono[6], S. Nampoothiri[7], B. Afroze[8], B. Hall[9], I. Lo[10], S. Lam[11], J. Hoefele[12], I. Rost[12], E. Wakeling[13], E. Mangold[14], K. Godbole[15], N. Vatanavicharn[16], L. Franco[17], K. Chandler[18], E. Faqeih[19], S. Hollander[1], T. Velten[1], K. Reicherter[1], J. Spranger[1], S. Robertson[20], L. Bonafe[21], B. Zabel[1,22], S. Unger[1,22].* 1) Ctr Pediatrics, Univ Freiburg, Freiburg, Germany; 2) Department of Biochemistry, University of Pavia, Italy; 3) Unité de Génétique Médicale, Faculté de Médecine, Université Saint Joseph, Beirut, Lebanon; 4) Children's Hospital Westmead, The University of Sydney, NSW 2006 Australia; 5) Hospital Infantil "La Fe", 46000, Valencia, Spain; 6) Division of Clinical and Metabolic Genetics, Hospital for Sick Children, 555 University Ave.,Toronto, Canada; 7) Amrita Institute Of Medical Sciences & Research Centre, Cochin, India; 8) Department of Paediatrics and Child Health, Aga Khan University Hospital,Karachi, Pakistan; 9) Department of Pediatrics, University of Kentucky, Lexington, Kentucky, USA; 10) Clinical Genetic Service, Department of Health, Hong Kong, China; 11) Department of Pediatrics, Chinese University of Hong Kong, Hong Kong, China; 12) Center for Human Genetics and Laboratory Medicine, Martinsried, Germany; 13) North West London Regional Genetics Centre (Kennedy-Galton Centre), North West London Hospitals NHS Trust, Middlesex, HA1 3UJ, UK; 14) Institute for Human Genetics, University Hospital, Bonn, Germany; 15) Department of Genetic Medicine, Deenanath Mangeshkar Hospital and Research Center, Pune, Maharashtra, India; 16) Division of Medical Genetics, Department of Pediatrics, Siriraj Hospital, Bangkok, Thailand; 17) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA; 18) Clinical Genetics, St Mary's Hospital Manchester, Hathersage Road, Manchester, UK; 19) Children's hospital, King Fahad Medical City, Riyadh, Saudi Arabia; 20) Department of Paediatrics and Child Health, Dunedin School of Medicine, Otago University, Dunedin, New Zealand; 21) Division of Molecular Pediatrics, CHUV, 1011 Lausanne, Switzerland; 22) Institute of Human Genetics, University of Freiburg, 79106 Freiburg, Germany.

We recently reported deficiency of carbohydrate sulfotransferase 3 (CHST3; chondroitin-6-sulfotransferase) in six subjects diagnosed with recessive Larsen syndrome or Humero-Spinal Dysostosis [Hermanns et al, AJHG 82:1368-74, 2008]. Since then, we have identified eighteen additional families with CHST3 mutations and we report here on a series of 27 patients. The diagnostic hypothesis prior to molecular analysis had been: Larsen syndrome (sixteen families), Humero-Spinal Dysostosis (four cases), chondrodysplasia with multiple dislocations (CDMD "Megarbane type"; two cases), Desbuquois syndrome (one case) and spondylo-epiphyseal dysplasia (one case). In spite of the different diagnostic labels, the clinical features in these patients were similar and included dislocation of the knees and/or hips at birth, clubfoot, elbow joint dysplasia with subluxation and limited extension, short stature, and progressive kyphosis developing in late childhood. The most useful radiographic clues were the changes of the lumbar vertebrae. We conclude that CHST3 deficiency presents with congenital dislocations of knees, hips, and elbows, and is often diagnosed initially as Larsen syndrome, humero-spinal dysostosis, or chondrodysplasia with dislocations. The incidence of CHST3 deficiency seems to be higher than assumed so far. The clinical and radiographic pattern (joint dislocations, vertebral changes, normal carpal age, lack of facial flattening, and recessive inheritance) is characteristic and distinguishes CHST3 deficiency from other disorders with congenital dislocations such as filamin B-associated dominant Larsen syndrome and Desbuquois syndrome.

145

A LARGE SCREENING OF ADAMTSL2 MUTATIONS IN 36 PATIENTS REVEALS THE GENETIC HETEROGENEITY OF GELEOPHYSIC DYSPLASIA. *C. Le Goff[1], S. Allali[1], G. Pfenning[1], N. Dagoneau[1], Y. Alanay[2], A. Brady[3], Y.J. Crow[4], K. Devriendt[5], V. Drouin-Garraud[6], E. Flori[7], D. Genevieve[8], R.C. Hennekam[9], J. Hurst[10], D. Krakow[11], K.D. Lichtenbelt[12], S.A. Lynch[13], K. MacDermot[14], S. Mansour[15], A. Magarbané[16], H.G. Santos[17], M. Splitt[18], A. Superti-Furga[19], S. Unger[19], D. Williams[20], M. Le Merrer[1], A. Munnich[1], V. Cormier-Daire[1].* 1) Dept Gen, INSERM U781, Hopital Necker, Paris, France; 2) Genetics Unit, Department of Pediatrics Hacettepe, University School of Medicine, Ankara, Turkey; 3) North West Thames Genetic Service, Imperial College, London, UK; 4) Genetic Medicine, Manchester Academic Health Sciences Centre, University of Manchester, UK; 5) Department of Medical Genetics, Ghent University Hospital, Ghent, Belgium; 6) Department of Medical Genetics, Hôpital Charles Nicolle, Rouen, France; 7) Department of genetics, Strasbourg hospital, Strasbourg, France; 8) Department of Genetics, Hôpital Arnaud de Villeneuve, Montpellier, France; 9) Department of Pediatrics, Academic Medical Center, University of Amsterdam, Netherlands; 10) Departement of Clinical Genetics, Oxford Radcliffe Hospitals, UK; 11) Cedars Sinai Medical Center, Los Angeles, USA; 12) Department of Clinical Genetics, University Medical Center, Utrecht, Netherlands; 13) National Center for Medical Genetics, Dublin, Ireland; 14) North West Thames Genetic Service, Imperial College, London , UK; 15) Medical Genetics Unit, St George's university of London, London, UK; 16) Unité de génétique médicale, Université Saint Joseph, Beyrouth, Liban; 17) Medical Genetics Service Hospital S. Maria , Lisboa , Portugal; 18) Institute of Human Genetics, Newcastle, UK; 19) Department of Pediatrics, University of Freiburg, Freiburg, Germany; 20) Department of clinical genetics, Birmingham women's hospital, Birmingam, UK.

Geleophysic dysplasia (OMIM 231050, GD) is an autosomal recessive disorder characterized by short stature, small hands and feet, stiff joints and thick skin. Patients present with a progressive cardiac disease often leading to an early death. Studying six GD families, we have mapped the disease gene on chromosome 9q34.2 and identified 6 mutations in the A Disintegrin And Metalloproteinase with Thrombospondin repeats-like 2 gene (ADAMTSL2). We identified Latent TGFβ Binding Protein 1 (LTBP1) as a partner of ADAMTSL2. We also found an enhanced TGFβ signalling in GD fibroblasts. Following this study, we collected the samples of 30 additional GD families. All cases fulfil the following diagnostic criteria: 1) short stature (< -2 SD) 2) short hands and feet 3) stiff joints 4) no internal notch of the femoral head to exclude acromicric dysplasia cases. Cardiac valve disease and thick skin were not considered as mandatory criteria based on their absence in some ADAMTSL2 mutated patients. We identified ADAMTSL2 mutations in 13/30 patients comprising 9 novel mutations. The absence of mutations in 17 patients prompted us to compare the two groups of GD patients, namely group 1, patients with ADAMTSL2 mutation (n=19) and group 2, patients without ADAMTSL2 mutation (n=17). No consanguinity was found in group 2 whereas half of the families of group 1 were consanguineous. The antenatal history showed hydramnios and prenatal growth retardation in 20% of group 2, versus 50% in group 1. We found that the main discriminating features were facial dysmorphism and tip-toe walking, almost constantly observed in group 1. Pseudo-muscular hypertrophy, high pitched voice and myopia/strabismus were seen more frequently in group 1 whereas liver enlargement was seen more frequently in group 2. No differences were found concerning heart involvement and skin thickness (which were observed in 75 % of both groups), the severity of the short stature, respiratory disorders, laryngo-tracheal stenosis, deafness and radiographic features. Two mutated patients and five non mutated patients died of cardiorespiratory failure. Finally, we found a significant increase of TGFβ level in fibroblasts of group 2, as previously observed in group 1. We conclude that GD is a genetically heterogeneous condition. Ongoing studies will hopefully lead to the identification of another disease gene presumably also involved in TGFβ bioavailability.

146

SHOX anormalies in children with Idiopathic Short Stature or Leri-Weill dyschondrosteosis in the French population. *V. Cormier-Daire[1], C. Huber[1], M. Rosilio[2], H. Sapin[2], J.-C. Carel[3], W.F. Blum[4].* 1) Dept of Medical Genetics, Paris Descartes University, INSERM U781, Necker Enfants Malades Hospital, Paris, France; 2) Lilly France, Medical Department Endocrinology and Diabetes, Suresnes, France; 3) Department of Pediatric Endocrinology, Hopital Robert Debré, France; 4) Eli Lilly and company, Lilly Research Laboratories, 6352 Bad Homburg, Germany.

The prevalence of SHOX deficiency in children evaluated for idiopathic short stature (ISS) has been variable in different studies (2-15%) due to the selection of patient for genetic screening, the molecular techniques used and geographical variance regarding prevalence of SHOX anomalies. We describe the results of SHOX analysis in a series of children with SS included in the SHOX Deficiency module of the GeNeSIS observational study. Genetic screening was performed in a central laboratory after patients had been routinely evaluated for other causes of SS at pediatric endocrinology centers throughout France. The molecular study included segregation of polymorphic microsatellite markers in the SHOX gene and the PseudoAutosomal region, PAR 1, and direct sequencing of the gene if no deletion was found. Statistical analyses were mostly descriptive. Clinical data were available for 216 patients (62% females) screened at 38 clinical centers (Jan 2003 - Oct 2007). Mean (SD) age was 10.6 (3.5) yrs and mean height SDS -2.4 (0.9). Patients were clinically categorized by the physician prior to genetic testing as Leri-Weill dyschondrosteosis (LWD, n = 119) or Idiopathic Short Stature (ISS, n = 97), that is short stature with no obvious dysmorphic signs such as Madelung deformity. 87 patients (40.3%) were diagnosed with SHOX deficiency due to partial or complete SHOX deletion (42 %), deletion of enhancer elements downstream of the SHOX gene (32.2%) or intragenic point mutations (19.5%). The frequency of SHOX gene point mutations/ deletions or downstream enhancer deletions was 60.5% in the LWD group and 15.5% in the ISS group. Our study confirms previous reports on the high frequency of SHOX gene/ region anomalies in LWD patients but also in patients with apparent ISS. Downstream deletions were relatively more frequent in ISS than LWD patients, suggesting a less severe phenotype with such genetic defect. Given the difficulty of adequately classifying patients, especially young children when phenotype is mild, SHOX analysis should be discussed in the workup of patients referred for short stature, after exclusion of the other usual causes.

147

Comparison of family health history to personal genomic screening for risk assessment of breast, colon and prostate cancer. *B. Leach, C. Eng.* Genomic Med Inst, Cleveland Clinic, Cleveland, OH.

Family history risk assessment (FHRA) has been shown to be 1 of the most effective tools for predicting what diseases an individual may be at risk for developing. With increased access and decreasing price-point for SNP profile-based risk assessment (RA) for common diseases, healthcare providers and consumers are questioning how personal genomic screening (PGS) can be integrated into preventative medicine, and how this type of genetic RA compares to standard FHRA. To date, there has been limited study of the accuracy and concordance of these 2 methods of risk assessment for common diseases. We compared FHRA with Navigenics PGS for breast (women), prostate (men), and colon cancer in 44 subjects ascertained from clinics at the Genomic Medicine Institute at Cleveland Clinic. We categorized each subject into the following groups based on FHRA and PGS results: general population (genpop), moderate (mod), or high risk. We assessed each subject's hereditary risk based on clinical criteria and/or validated gene test results. Half the subjects were male, 42 White, 1 Asian, and 1 Pacific Islander. We compared FHRA and PGS RA. Both FHRA and PGS put 59%, 41%, and 39% of subjects into the same risk categories for breast, prostate, and colon cancer, respectively. Overall concordance between FHRA and PGS was low for all 3 cancers (K<0.15). For prostate cancer, the PGS predicted a mod/high risk for 8 subjects while FHRA was genpop risk (K=0.145). We also observed 12 subjects with mod colon cancer risk on PGS but genpop risk on FHRA (K=-0.049). None of the 3 hereditary prostate cancer subjects were assessed as high risk on PGS. Based on FHRA, 10 subjects had hereditary breast cancer risk and PGS only identified 1 as high risk (K=0.12). None of the 9 hereditary colon cancer subjects were high risk on PGS. For cancer RA FHRA and PGS agreed on the subjects' risk categories 46% of the time. For the 22 with a hereditary risk, the PGS only identified 1 subject as high risk. This is especially concerning for patients at hereditary risk that may elect a PGS approach over validated FHRA and testing. Based on our study, FHRA and PGS may be complementary tools for cancer RA. However, evaluation of family history is still the proven gold standard, and this should be used to clinically evaluate an individual's risk of developing cancer until further research is done to prove that PGS can predict an individual's risk or can be integrated with FHRA to increase sensitivity.

148

Components of family history associated with women's disease perceptions for cancer: A report from the Family Healthware Impact Trial (FHITr). *W.S. Rubinstein[1,2], S.M. O'Neill[1,3], N. Rothrock[3], E.J. Starzyk[4], J.L. Beaumont[3], R. Gramling[5], J.M. Galliher[6], L.S. Acheson[7], C. Wang[8], M.T. Ruffin IV[9] for the FHITr group.* 1) Medical Genetics, NorthShore University HealthSystem, Evanston, IL; 2) University of Chicago Pritzker School of Medicine, Chicago, IL; 3) Northwestern University, Chicago, IL; 4) University of Illinois at Chicago, Chicago, IL; 5) University of Rochester, Rochester, NY; 6) American Academy of Family Physicians' National Research Network; 7) Case Western Reserve University, Cleveland, OH; 8) Boston University School of Public Health, Boston, MA; 9) University of Michigan, Ann Arbor, MI.

Purpose: To determine the elements of family history (FHx) and personal characteristics most closely related to women's disease perceptions about breast (BC), colon (CC) and ovarian (OC) cancer. Methods: Baseline, cross-sectional data on 2505 healthy women age 35-65 enrolled from 41 primary care practices in the cluster-randomized Family HealthwareTM Impact Trial (FHITr), assessed for perceived risk, perceived severity, worry, and perceived control over getting coronary heart disease, stroke, diabetes, BC, CC and OC, and detailed FHx. Results: Participants provided details on 41,841 total relatives. Excluding their parents, significantly more women reported having cancer in their maternal vs. paternal lineage. Additionally, FHx of cancer on the paternal vs. the maternal side was associated with a lower perceived risk for BC. Cancer-specific perceived risks were seen for all three cancers for the total number of relatives or having any first-degree relative with that cancer. There were also novel "spillover" effects: a positive FHx of BC or CC was associated with a higher perceived risk of OC. Age and parenthood were associated with disease perceptions in ways that ran counter to empiric risks. Conclusions: Understanding the elements that contribute to individuals' disease perceptions has important implications for communicating health risk through automated family history tools, by healthcare providers, and via public health campaigns. Our findings suggest that female primary care patients have limited awareness and understanding of the relevance of their paternal FHx, which creates an opportunity for a public health education approach. Our discovery that disease risk perceptions may not be congruent with objective familial risk estimates raises the specter of providing computerized health advice that runs counter to or ignores the patients' perceptions and their personal sense of vulnerability. Some have suggested that public health should place a stronger emphasis on patient-centered approaches aimed at helping patients achieve informed choices instead of promoting in general the positive value of screening, preventive surgeries, or genetic testing. The integration of patient- and provider-centered approaches (including provider competencies) has been advocated by the American Society of Preventive Oncology as a key research focus. Our work supports this approach and identifies key areas of patients' disease perceptions for consideration.

149

Are physicians prepared to effectively use genetic testing for at-risk relatives of cancer patients? *S.E. Plon[1,2,3], B. Parks[1], P. Cooper[4], T. Wang[2,4], S. Dhar[3], P.A. Kelly[4], S. Gonzalez[5], A. Weinberg[4], S. Hilsenbeck[2,4].* 1) Dept Pediatrics, Baylor Col Med, Houston, TX; 2) Dan L Duncan Cancer Center, Baylor Col Med, Houston, TX; 3) Dept Mol & Human Genetics, Baylor Col Med, Houston, TX; 4) Dept Medicine, Baylor Col Med, Houston, TX; 5) Texas Medical Association, Austin, TX.

Background: DNA sequence-based testing is used for prediction of cancer risk with results described as negative, deleterious mutation or variant of uncertain significance (VUS). There is little knowledge of how physicians utilize different types of genetic test results in clinical decision making. **Methods:** Practicing physicians in Texas from five specialties (general surgery, obstetrics-gynecology, hematology-oncology, family medicine and internal medicine) completed an online case-based survey which asked for genetic testing and management recommendations for the healthy relatives of cancer patients who had no testing, were found to carry a *BRCA1* deleterious mutation or VUS. **Results:** Surveys were obtained from a random sample of 225 physicians. When the cancer patient carried a deleterious mutation or VUS, 98% and 82% of physicians, respectively, recommended testing of healthy relatives (p<0.0001). In both situations comprehensive *BRCA1/2* analysis was selected most often (compared with the appropriate "single-site test" for a deleterious mutation or "testing not appropriate" for VUS) with no difference (p=0.3869) between physicians who had (n=81) or had not (n=134) ordered *BRCA1/2* testing in their own practice. There were no significant differences among specialties for testing recommendations when a VUS was identified in the cancer patient (Chi-Square test, p=0.1011). Post-test cancer risk was assessed as higher by 70% of responders for a healthy relative found to carry the deleterious mutation versus only 27% for VUS carriers (p<0.0001). Similarly, a cancer risk management intensity score was highest for the healthy relative with a deleterious mutation compared with VUS, negative or no testing and 63% of physicians recommended oopherectomy for the healthy relative with a deleterious mutation compared with 13% for the VUS and 2.2% for negative (McNemar's test, p<0.0001). **Conclusions:** Independent of experience, physicians chose comprehensive sequencing for healthy relatives compared with guidelines which recommend only testing for the specific deleterious mutation identified in the cancer patient. In contrast to testing decisions, appreciation for the uncertainty associated with a VUS result was demonstrated by cancer risk and management decisions. Utilization of genetic professionals and education of physicians on family-based testing may improve efficacy and substantially reduce costs associated with genetic testing. Supported by grant 5R01HG004064.

150

Exploring the effect of gene-gene and gene-environment interaction in breast-cancer risk prediction. *H. Aschard[1], J. Chen[2], P. Kraft[1], Program in Molecular and Genetic Epidemiology, Harvard School of Public Health.* 1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) Department of Biostatistics, University of Pennsylvania, Philadelphia, PA.

Genome-wide association scans have identified scores of common genetic variants associated with the risk of complex diseases in the last years. However their aggregate effects on risk beyond traditional factors remain uncertain. A recent paper reported that the addition of information on genetic variants associated to breast cancer (BrCa) to the Gail model, a commonly used tool to assess the risk of BrCa, can moderately improve discrimination and prediction in estimating that risk. Here we explore the extent to which the consideration of interaction between genetic variants (GxG) and interaction between genetic variants and these established clinical factor (GxE) can add to these risk-assessment models.

Using data from the Nurses' Health Study, we derive the predictive ability of 17 independent published risk alleles in 1145 BrCa cases and 1142 matched controls. We evaluate the extent to which these single nucleotide polymorphisms (SNPs) can discriminate BrCa cases from controls, in the total sample and in strata defined by age and Gail Score. We also conduct a simulation study to explore the potential improvement of discrimination if more complex GxG and GxE interactions exist and we know them. Disease status is generated using the real genotypes across a broad range of SNP-SNP and SNP-environment interaction effects. The interaction effects were defined such that the final marginal effect of each SNP was equal to its previously reported estimated effect.

Stratification of genetic risk by other risk factors shows a significant difference of C statistic by age quartile, with a decrease of discrimination ability of the SNP model with age (from 64.42% for the women of the first quartile to 57.77% for the women of the last quartile, p=0.04). The C statistic was also significantly different when stratifying the sample on the Gail score (62.37% for the women having below median risk and 57.31% for the women above median risk; p=0.04). Adding the simulated GxG and GxE interactions effect in the SNP model shows a potential improvement in the discrimination ability. However this gain is quite small (<5%), except for rare extreme scenarios that are inconsistent with empirical results on gene-gene and gene-environment interactions for these SNPs.

151

Why is genetic screening for autosomal dominant disorders underutilized in families? The case of hereditary hemorrhagic telangiectasia (HHT). *C. Zayac, G. Griffin, R.E. Pyeritz, B.A. Bernhardt.* Medical Genetics, University of Pennsylvania, Philadelphia, PA.

HHT is an autosomal dominant disorder of vascular development resulting in epistaxis, pulmonary and cerebral AVMs, telangiectases, and GI bleeding. Appropriate management reduces morbidity and mortality, but relies on identifying affected relatives. Because the most easily identified signs of HHT (telangiectases and epistaxis) may not appear until young adulthood, many young affected relatives cannot be diagnosed clinically. Genetic testing identifies a mutation in 85% of probands, and testing for the familial mutation can identify which at-risk relatives need intensive management. Despite its clinical utility and cost effectiveness, genetic testing for HHT is underutilized. To investigate barriers to genetic testing, we conducted 13 on-line discussion groups, 11 with people with HHT and family members (119 participants), and 2 with physicians, genetic counselors, nurses and clinical laboratory personnel (26 participants). Multiple barriers to HHT genetic testing were identified including lack of knowledge and awareness of genetic testing, problems with access, and emotional barriers. People from HHT families often fail to understand the rationale for HHT testing, believing that testing is performed in a proband only to confirm the diagnosis, rather than to facilitate testing of at-risk relatives. Both providers and patients/family members incorrectly assume genetic testing is not covered by insurance. The majority of participants believe that primary care providers do not know how to order HHT testing or interpret results. Access to testing is limited by distance from an HHT Center or a genetics clinic. Emotional barriers to genetic testing include fear of insurance discrimination; fear of being diagnosed with HHT; denial of having HHT or being at risk; fear of stigmatization; and feelings of guilt. These findings show that there is misunderstanding of genetic testing among both HHT families and healthcare providers. Although we focused on HHT, our findings have implications for genetic testing for other genetic conditions as well. We recommend the development and distribution of brief educational materials for both patients and providers that clearly describe the rationale for genetic testing, and emphasize the benefits of early detection and treatment. In addition, laboratories offering genetic testing should provide support for primary care physicians to facilitate the ordering process and to help interpret genetic test results.

152

Challenges with increased microarray density: increased frequency of unclear results and reconsideration of parental testing utility. *S.A. Morton, A.M. Bandholz, N. Neill, B.C. Ballif, A.N. Lamb, L.G. Shaffer, B.A. Bejjani.* Signature Genomic Laboratories, Spokane, WA.

Evolution of microarray technology has yielded progressively dense, high resolution clinical aCGH testing. The ability to identify smaller gains and losses increases diagnostic power, but also presents a challenge as more copy number variations (CNVs) of unclear clinical significance are identified. Parental testing is routinely used to clarify the significance of CNVs in children. Recent awareness of the influence of reduced penetrance and variable expressivity and de novo mutation rates of CNVs demands reconsideration of dependence on parental testing. Microarray data were compiled on 2,203 consecutive individuals with CNVs by whole genome oligonucleotide arrays. The increase in array resolution was compared with the detection rate of CNVs. The rate of inheritance was correlated with both gene content and CNV size. Increasing array resolution to detect CNVs smaller than 500 kb yielded an increase in clinically significant results by 14% and resulted in a 174% increase in CNVs of unclear significance. Ninety-six percent of copy gains were found to be inherited regardless of size and number of genes. The rate of de novo deletions correlated with gene content, ranging from 14.3% for deletions of £ 2 genes to 45.0% for deletions with greater than 11 genes. Size demonstrated a similar correlation, increasing from a de novo rate of 10.1% for deletions smaller than 250 kb to 47.4% for those over 1 Mb. Our results show CNVs of unclear significance are commonly inherited and their detection increases with higher resolution aCGH. When inherited, they could represent a benign finding, a predisposition locus with incomplete penetrance, or a deleterious aberration in a subclinically or overtly affected parent. Apparent de novo alterations may represent a benign novel structural mutation, nonpaternity, or a disease-causing alteration. Parental testing for results of unclear clinical significance cannot solely determine phenotypic effect or provide recurrence risks of the abnormal phenotype. A thoughtful approach evaluating the gene content, family history, and clinical phenotype is critical in discerning when parental testing may be useful for an individual or family. Ultimately, determination of significance at a given locus is dependent on comparison to large control populations, additional functional studies, or the identification of specific point mutations and may not be resolved by parental testing in an individual family.

153

Interpretation and Genetic Counseling for the newly characterized recurrent genomic disorders. *S. Mulchandani[1], L. Conlin[1], B. Thiel[1], E. Zackai[3], N. Spinner[1,2].* 1) CytoGenomics lab, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) The University of Pennsylvania School of Medicine Philadelphia, PA; 3) Clinical Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA.

Disease causing genomic deletions (del) and duplications (dup) can occur anywhere in the genome, although the presence of segmental dups predisposes to these events. While genomic disorders were first recognized as clinical syndromes with subsequent recognition of the etiology, the use of chromosomal microarray (CMA) in patients with non specific clinical features has lead to the identification of several new segmental dup mediated disorders. Some of these new syndromes appear to have distinctive phenotypes, while others are highly variable despite their identical genomic content. We have analyzed 3,267 patients using CMA (Illumina Quad610 SNP array) in our diagnostic lab and reported 833 patients (25.5%) with genomic alterations not seen in control individuals. Eighteen percent of the reported events occurred in regions that are flanked by segmental dups. Of the segmental dup mediated regions, 58% (88/152) resulted in well-described, clinically recognizable syndromes such as deletions in 7q11.23 (Williams syndrome), 8p23.1 (deletion involving GATA4), 17p11.2 (Smith Magenis syndrome), Xp22.31 (X linked ichthyosis) and deletion /duplication in 15q11q13 (Prader-Willi/Angelman), 22q11.2, 17p12 (HNPP/CMT1). The remaining 42% are newly described syndromes including del and dup of 16p11.2, 1q21.1, 16p13, 15q13, 16p12, and deletion 17q21.31 and reciprocal dups of the previously described syndromes. We recognize four groups of alterations. The first group is most likely to be pathogenic, with clear phenotype and de novo origin (17q21.31 deletion). The second group also had a consistent phenotype, but was consistently inherited from an affected parent (7q11.23 duplication). The third group had a variable phenotype, but showed consistent de novo origin (del/dup 16p11.2). The final group showed striking variable expressivity and was seen inherited in 17/21 cases (del/dup 1q21.1, 16p12, 16p13 and dups of Xp22.31). For these last two groups, the variable expressivity coupled with inheritance made interpretation of the alteration very challenging. Our experience highlights the difficulties in counseling the families with the recently described microdeletion/microduplication syndromes. Continued follow up, thorough analysis of controls, and research into factors controlling variable expressivity will be crucial to better understand the significance of the genomic alterations.

154

The Use of High Resolution SNP Arrays in the Diagnostic Work-Up for Apparent Non-Syndromic Bilateral Sensorineural Hearing Loss. *D. Clark[1], L. Francey[1], L. Conlin[3], M. Berman[1], A. Wilkens[1], H. Fetting[1], H. Kadesch[1], M. Kenna[4], H. Hakonarson[2], N. Spinner[3], H. Rehm[4], I. Krantz[1].* 1) Dept of Human Genetics, 1007 ARC, Children's Hosp Philadelphia, Philadelphia, PA; 2) The Center for Applied Genomics, Children's Hosp Philadelphia, Philadelphia, PA; 3) Div of Clinical Labs, Dept of Pathology, Children's Hosp Philadelphia, Philadelphia, PA; 4) Dept of Otolaryngology and Communication Disorders, Children's Hospital Boston, Boston, MA.

The availability of high-resolution genome-wide SNP arrays has greatly improved the detection of genomic rearrangements in patients with syndromic presentations, yet there is little published on the value of these platforms in a population with isolated structural or sensory deficits. We evaluated the utility of SNP array genotyping for the detection of pathogenic copy number variants (CNVs) in a large population of children with apparent isolated non-syndromic bilateral sensorineural hearing loss (BLSNHL). BLSNHL is genetically heterogeneous however despite the plethora of known BLSNHL genes, greater than 50% of children with a genetic form of BLSNHL have no identifiable cause. At CHOP, children with BLSNHL are evaluated by clinical geneticists who offer targeted genetic testing. We enrolled 341 *GJB2* mutation negative probands in an IRB approved study and performed Illumina SNP array analysis to identify pathogenic copy number variations (CNVs). We anticipated identifying: 1) cryptic syndromic forms of HL, 2) new dominantly acting BLSNHL loci, 3) recessive forms of BLSNHL through detection of i) homozygous deletions or ii) heterozygous deletion that unmasks a point mutation on the non-deleted allele. The SNP arrays were able to detect: all 3 known *GJB6* deletions (left in for validation purposes), a homozygous *OTOF* deletion, an *USH2A* deletion unmasking a trans allele mutation, and 3 *STRC* deletions. We subsequently tested 111 consecutive individuals referred to the Genetics of Hearing Loss Clinic with non-syndromic BLSNHL uncovering unsuspecting syndromic diagnoses in 4 (3.6%) patients, a *CHD23* deletion unmasking a mutation on the trans allele, and 3 patients with *STRC* deletions. Several additional heterozygous deletions in known HL genes were identified. Given the genetic heterogeneity of BLSNHL, the use of newborn screening enabling early detection of HL, the cost of screening all known HL genes and the inability of molecular sequencing to detect larger gene deletions, SNP array testing may be a valuable diagnostic tool in this population. A non-syndromic BLSNHL cohort from Boston Children's, where the patients were not evaluated by a clinical geneticist, found 6 out of 301 to have large CNVs. In our cohort 4% (17/452) of individuals with BLSNHL were molecularly diagnosed using SNP arrays. These findings highlight the evolving utility of high-density, genome-wide SNP arrays in uncovering diagnoses in patients with a specific clinical finding.

155

Comprehensive Genetic Diagnosis for Hereditary Hearing Loss Using Next-Generation Sequencing. *E. Shearer[1, 2], A.P. DeLuca[3, 4], M.S. Hildebrand[1], K.R. Taylor[3, 4], J.G. Gurrola[1], S.E. Scherer[5], T.E. Scheetz[3, 4, 6], R.J. Smith[1, 2, 7].* 1) Department of Otolaryngology - Head and Neck Surgery, Carver College of Medicine, University of Iowa, Iowa City, IA; 2) Department of Molecular Physiology and Biophysics, Carver College of Medicine, University of Iowa, Iowa City, IA; 3) Department of Biomedical Engineering, University of Iowa, Iowa City, IA; 4) Center for Bioinformatics and Computational Biology, University of Iowa, Iowa City, IA; 5) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 6) Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, IA; 7) Interdepartmental PhD Program in Genetics, University of Iowa, Iowa City, Iowa City, IA.

Background: Hearing loss is the most common human sensory impairment, with several hundred mutations in more than 50 genes identified. This extreme genetic heterogeneity makes genetic diagnosis expensive and time-consuming using currently available methods. Next-generation sequencing technologies have recently become available that may provide the throughput and efficiency required for a truly comprehensive genetic test for hearing loss but validation is required before implementation. Methods: We performed sequence capture on DNA from 10 subjects for the exonic regions and flanking sequence of all genes implicated in nonsyndromic hearing loss and Usher syndrome using both solid-phase and solution-based sequence capture methods. We sequenced the enriched libraries using the Illumina and 454 next-generation sequencing platforms, respectively at saturating depths of coverage, mapped sequencing reads with BFAST and BOWTIE, and developed a custom variant calling and annotation pipeline (ASAP) in order to prioritize variants as most likely to be disease-causing. To measure the validity of these methods, we used positive and negative controls, biological replicates, and we genotyped 537 single nucleotide polymorphisms (an average of 54 per sample) using Sanger sequencing. Results: Our results show a sensitivity of >99% and specificity of >99% for both solution-based sequence capture paired with Illumina sequencing and solid-phase sequence capture paired with 454 sequencing when saturating sequence coverage is applied. We determined the causative positive control mutation in three samples, and found no disease-causing variants in the negative control. We determined a known hearing loss missense mutation in one patient and discovered a novel stop mutation paired with a large deletion in another patient. We also determined that one patient was a compound heterozygote for known mutations causing Usher syndrome. Conclusions: Next-generation sequencing technologies show high sensitivity, specificity, and reproducibility in genetic testing when saturating coverage is applied and appropriate data analysis is used. This technology can be applied to the genetic diagnosis of hearing loss as a more comprehensive and efficient method than currently available.

156

Chromosome 22q11.2 "B-D" Deletion - New Diagnosis or More of the Same? *D.M. McDonald-McGinn[1], A. Kohut[1], S. Saitta[1], R. Laskin[1], A. Hacker[1], S. Bale[3], S. Aradhya[3], J. Rosenfeld[4], L. Shaffer[4], K. Sullivan[2], B. Emanuel[1], E.H. Zackai[1].* 1) Div Human Gen, Children's Hosp Philadelphia and U of Penn School of Medicine, Philadelphia, PA; 2) Div All, Immun & Infect Dis, Children's Hosp Philadelphia and U of Penn School of Medicine, Philadelphia, PA; 3) Gene Dx, Gaithersburg, MD; 4) Signature Genomics, Spokane, WA.

Historically most patients with a 22q11.2 deletion were found to have A-D hemizygous deletions including loss of TBX1 positioned between the A-B LCR blocks & considered a major phenotypic contributor based on knockout mice. However, arrays are now identifying atypical/nested deletions often not including TBX1 raising important management & genetic counseling questions. To provide such anticipatory guidance we reviewed our database & queried 2 clinical labs comparing frequencies of A-D v. atypical deletions & clinical features. Of our 904 patients 436 had deletion sizing by enhanced FISH or array studies. Of these 63 (15%) had atypical deletions including 28 A-B (44%), 10 A-C (16%), 17 B-D (27%), 2 C-D (3%), & 6 "other" (10%) (not flanked by LCR blocks). The reference lab data was similar with 44/231(19%) & 11/43 (26%) atypical deletions found including 38% & 18% with B-D deletions. Of note, B-D & the less common C-D deletions do not include TBX1 & are not detectable using standard FISH probes. Findings in our patients with B-D deletions included: developmental delay (10); heart defects (5); GERD (5); short stature (3); palatal anomalies (3); hypocalcemia (2); iris coloboma; laryngeal web; polymicrogyria; preauricular tags & unilateral renal agenesis. Conversely 8/8 patients tested had normal T cells. Interestingly of the 3 cohorts 5/11, 7/10 & 1/1 families tested respectively were familial - a much higher frequency than we previously reported using FISH studies (6%) (McDonald-McGinn, 2001). So in summary, B-D is the most frequent atypical/nested deletion not including TBX1. In general structural anomalies may be less common but 29% of our patients had classically associated conotruncal anomalies such as TOF, IAA & TA even with TBX1 present, perhaps suggesting a downstream effect. Moreover, with the exception of the lack of immunodeficiency, the systems involved (e.g. palate, gastrointestinal tract, CNS, etc.) appear similar to the standard A-D deletion. Therefore, while our studies delineate an important new subset of patients it appears clinical management including prenatal and postnatal genetic counseling should mirror that which is currently employed for patients with the standard A-D deletion. Furthermore, in light of the high incidence of familial B-D deletions, adult practitioners must be vigilant in their history taking in order to ascertain affected persons with a 50% recurrence risk. Lastly, perhaps the term atypical should be revisited.

157

Pharmacogenomic Tests: The Good, the Bad and the Unknown. *D.J. Allingham-Hawkins, A. Lea, L. Spock, S.A. Levine.* Hayes Inc, Lansdale, PA.

Pharmacogenomics, also known as pharmacogenetics, is the study of how an individual's genetic make-up affects his or her response to a medication. Pharmacogenomic tests have the potential to revolutionize personalized healthcare by allowing the accurate prediction of drug efficacy and/or the risk of adverse side effects. Unfortunately, many pharmacogenomic tests are made clinically available before sufficient evidence exists to support their routine use. The Genetic Test Evaluation (GTE) Program at Hayes, Inc., a private health technology research and consulting company, has performed evidence-based reviews of 20 widely available pharmacogenomic tests. Methods: All reviews are based on the ACCE model for assessing genetic tests that was developed by the Centers for Disease Control and Prevention (CDC). Comprehensive literature reviews were performed for each test evaluated and the relevant gray literature (conference proceedings, websites, etc.) was reviewed. A Hayes GTE Rating that reflects the strength and direction of the evidence was assigned for each possible unique application of a test. Results: Some tests, such as *BCR-ABL* testing to predict response to imatinib in chronic myelogenous leukemia patients and *HLA-B*5701* testing to detect those individuals most likely to experience a potentially fatal hypersensitivity reaction to abacavir, have been sufficiently studied and clinical utility can be demonstrated. For other tests, such as *KRAS* testing to predict response to the anti-epidermal growth factor receptor antibodies cetuximab and panitumumab used to treat metastatic colorectal cancer, there is consistent evidence of potential utility, although the number of patients studied to date is relatively small. The majority of tests reviewed (16/20; 80%), however, currently have either conflicting or inconsistent evidence, or the evidence is insufficient to evaluate their clinical utility. Of the 28 applications of the 20 tests evaluated, 22 (78.6%) were not supported by current evidence. As such, the use of these tests, which include *CYP2C19* and *CYP2D6* testing to predict response to clopidogrel or tamoxifen, respectively, and *CYP2C9* and *VKORC1* testing for management of warfarin dosing, among others, appears to be premature. Conclusions: Although a small number of pharmacogenomic tests have demonstrated clinical utility based on published evidence, for the majority, the current evidence does not support their use in routine clinical practice.

158

Efficacy and Safety of Clopidogrel compared with Placebo according to CYP2C19 Genotype in over 6000 patients with Non-ST-elevation Acute Coronary Syndromes (CURE trial) and atrial fibrillation (ACTIVE trial). *G. Pare[1], S. Mehta[1], S. Yusuf[1], S. Anand[1], S. Connolly[1], J. Hirsh[1], K. Simonsen[2], D. Bhatt[3], J. Eikelboom[1].* 1) McMaster University, Hamilton, ON, Canada; 2) Bristol-Myers Squibb, New York, NY, USA; 3) Harvard Medical School, Boston, MA, USA.

Background: It has been suggested that clopidogrel may be ineffective in preventing major cardiovascular (CV) events in those who are carriers of a loss-of-function CYP2C19 alleles that are associated with reduced metabolism of clopidogrel to its active metabolite. However, none of these reports examined the efficacy and safety of clopidogrel compared with a randomized placebo group according to CYP2C19 genotype. Methods: We genotyped patients from two large randomized trials that demonstrated the efficacy of clopidogrel versus placebo in preventing vascular events (5059 acute coronary syndrome patients from the CURE trial and 1156 atrial fibrillation patients from the ACTIVE trial) for three single nucleotide polymorphisms (*2, *3, *17) that define the major CYP2C19 alleles. We used Cox proportional hazard regression to examine the efficacy and safety of clopidogrel compared with placebo according to genetically determined CYP2C19 metabolizer phenotype and functional allele carrier status. The primary efficacy outcome was the composite of myocardial infarction, stroke or death and the primary safety outcome was major bleeding events. Results: In ACS patients, clopidogrel compared to placebo reduced the primary efficacy outcome irrespective of genetically determined metabolizer phenotype (P for heterogeneity = 0.12). Carriers of loss-of-function alleles derived a similar protection against CV events from clopidogrel (8.0% vs 11.6%, HR=0.69, 95%CI 0.49-0.98) as compared to noncarriers (9.5% vs 13.0%, HR=0.72, 95%CI 0.59-0.87). By contrast, a significant interaction (P=0.016) was observed between gain-of-function allele carrier status and treatment allocation such that gain-of-function carriers had a more pronounced benefit with clopidogrel (7.7% vs 13.0%, HR=0.55, 95% CI 0.42-0.73) as compared to noncarriers (10.0% vs 12.2%, HR=0.85, 95%CI 0.68-1.05, P interaction= 0.016). The effects on bleeding did not vary by genotypic subgroups. The results were similar in AF patients, with no evidence of interaction either on efficacy or bleeding between treatment and metabolizer phenotype, loss-of-function carrier status or gain-of-function carrier status. Conclusions: Our study shows that CYP2C19 genetic variants do not have a discernible effect on clinical responses to clopidogrel. Our data suggest that genotype status should not preclude the use of clopidogrel at currently recommended doses in non-invasively managed ACS patients and AF patients.

159

MEK inhibitor rescues growth of peewee tibia in culture, a therapeutic model for acromesomelic dysplasia Maroteaux type. *K.A. Geister, M.L. Brinkmeier, S.A. Camper.* Human Genetics, Univ Michigan Medical Sch, Ann Arbor , MI.

Skeletal disorders are an important contributor to short stature, occurring 1/4000 births. Half of the 372 known subtypes can be explained by mutations in about 140 genes. There are no current treatments other than controversial leg lengthening surgeries. Mouse models of human skeletal dysplasias provide a tool for understanding the mechanisms of disease pathology and for exploring therapies. We characterized a new recessive, spontaneous mouse mutant with skeletal dysplasia and fertility problems, peewee (*pwe*) and mapped it to a region of mouse Chr 4 with over 100 genes. We identified a 4 bp deletion in exon 3 of the natriuretic peptide receptor 2 gene (*Npr2*) that produces a loss of function allele. Peewee mutants exhibit severe, disproportionate growth insufficiency and reduction in the hypertrophic zone of the epiphyseal growth plate. Humans homozygous for *NPR2* mutations have severe short stature, Acromesomelic Dysplasia Maroteaux type. NPR2 is a guanylyl cyclase that is activated by C-type natriuretic peptide (CNP) to regulate a kinase signaling pathway in chondrocytes. NPR2 activation inhibits fibroblast growth factor signaling by inhibiting the MAP kinase pathway (MEK 1/2), resulting in balanced regulation of bone growth. We treated fetal tibiae from normal and peewee mutants with a MEK inhibitor in culture and rescued the deficiency in bone growth. We are testing the efficacy of the drug to correct the growth defect in intact fetal and neonatal mice and evaluating the quality of bone produced in response to the therapy. Peewee females are infertile and exhibit ovarian hypoplasia. *Npr2* is expressed in multiple tissues of the hypothalamic-pituitary-gonadal axis, including the brain, pituitary, ovary, and uterus, which may contribute to the ovarian hypoplasia and infertility of peewee females. We are exploring the role of NPR2 in the pituitary response to gonadotropin releasing hormone and in regulation of ovarian follicular atresia. In summary, we present a mouse model of human skeletal dysplasia that is responsive to pharmacotherapy in culture.

160

Inhibition of TGFβ Signaling Prevents Axial Myopia in a Mouse Model of Marfan Syndrome. *H. Dietz[1,2], J. Doyle[1], C. van Erp[1], H. Aziz[1].* 1) Johns Hopkins Univ Sch Med, Baltimore, MD; 2) Howard Hughes Medical Institute (HHMI), Baltimore, MD.

Myopia is a highly prevalent ocular condition, affecting over 30% of the U.S. population. High myopia is associated with a number of sight-threatening complications, including retinal detachment. Myopia typically results from excessive axial elongation of the globe, although the molecular mechanisms driving this process are poorly understood. TGFβ signaling has been implicated in myopia in visual deprivation-based rodent models and human association studies, but its role remains highly controversial, with many studies suggesting a protective effect. Early-onset, rapidly progressive and severe myopia is highly prevalent in Marfan syndrome (MFS), a connective tissue disorder caused by mutations in *FBN1*. Enhanced TGFβ signaling has been implicated in the pathogenesis of multiple manifestations of MFS, including aortic aneurysm, emphysema, myxomatous valve disease and skeletal myopathy. TGFβ neutralizing antibody (TGFβNAb) or the angiotensin II type 1 receptor blocker losartan ameliorate these phenotypes in MFS mice, in association with blunted TGFβ signaling. We analyzed the role of TGFβ signaling and the potential of its antagonism as a therapeutic strategy for MFS-associated myopia. The axial globe length of MFS and wild-type (WT) mice was measured using microscopy of OCT-embedded eyes sectioned through the axial midline. MFS mice had progressively greater axial globe length compared to WT littermates at 2, 4 and 6 months of age. Whereas eye growth was complete in WT mice by 4 months of age, it progressed rapidly in MFS mice through 6 months of age, suggesting an opportunity for postnatal intervention. To investigate whether TGFβ signaling was causal in this process, we treated mice with TGFβNAb or placebo from 2 to 4 months of age. At 4 months, TGFβNAb-treated MFS mice had axial globe lengths that were significantly less than placebo-treated MFS littermates ($p<0.005$), being indistinguishable from WT ($p=0.33$). Losartan-treated MFS mice also showed shorter globe lengths compared to placebo-treated MFS littermates at 6 months of age ($p<0.001$), again being indistinguishable from WT ($p=0.57$). Intraocular pressure was indistinguishable between WT and placebo- or losartan-treated MFS mice, eliminating this as a contributing variable. These data suggest that enhanced TGFβ signaling drives elongation of the globe and myopia in MFS, and provide the first evidence for potential medical treatments for myopia in MFS and perhaps in other nonsyndromic presentations.

161

A genome-wide association study on response to chemotherapy in ovarian cancer. *S. Macgregor[1], Y. Lu[1], S.E. Johnatty[1], J. Beesley[1], X. Chen[1], B.L. Fridley[2], P.M. Webb[1], H. Gurney[3], P.R. Harnett[4], M. Haber[5], E.L. Goode[2], A. DeFazio[4], G. Chenevix-Trench[1].* 1) Genetics and Population Health Division, Queensland Institute of Medical Research, Brisbane, Australia; 2) Mayo Clinic College of Medicine, Rochester, MN, USA; 3) Department of Medical Oncology, Faculty of Medicine, University of Sydney, Australia; 4) Westmead Institute for Cancer Research, University of Sydney at the Westmead Millennium Institute; 5) Children's Cancer Institute Australia, Sydney, Australia.

We performed a multi-stage genome-wide association study (GWAS) for drug response in ovarian cancer. In our first stage we took 313 chemotherapy treated patients from the Australian Ovarian Cancer Study (AOCS). Patients were treated with paclitaxel/carboplatin and the trait of interest was Progression Free Survival (PFS). We maximized power for a fixed genotyping budget by genotyping 94 patients with good survival and 94 with poor survival, taking into account extent of residual disease following surgery. Patients were genotyped on Illumina HumanOmni1-Quad arrays (1140419 loci). We sought replication in two further cohorts with measures of PFS; 148 individuals from the Mayo clinic (genotyped on Illumina HumanHap 610 arrays, 620901 loci) and 242 individuals from The Cancer Genome Atlas ovarian cancer set (genotyped on Illumina Human1M arrays, 1199187 loci). A SNP on chromosome 14 which ranked in the top 10 (p=2.6e-6) in the AOCS samples achieved nominal replication in both the Mayo and TCGA samples (meta-analysis p=3.2e-7). Two further snps on chromosome 13 (meta-analysis p=2.8e-6) and chromosome 14 (distinct from other chromosome 14 locus, meta-analysis p=3.2e-6) were identified in a meta-analysis of all samples. Imputation is underway to allow more direct comparison of SNPs not common to all of the arrays used. For the AOCS data, restricting analysis to a defined, homogeneous treatment subgroup (a minimum of 4 cycles of Carboplatin AUC 5 or 6 and Paclitaxel 135 or 175 mg/m2) revealed strong evidence for association at the previously identified ABCB1 (MDR) locus (p=2.0e-6, rs1211152) but this effect was diminished in analysis of the full AOCS data set (p=0.02). Analysis of the restricted chemotherapy subgroup in the Mayo and TCGA samples is underway. We are currently following up the top findings from our GWAS in a large independent replication cohort (N=2000, with samples from multiple sites). In summary, we have identified genes that underlie individual differences in chemoresponse. In the future, typing variations in these genes will be important for personalizing drug regimes for patients.

162

Susceptibility to amoxicillin-clavulanate-induced liver injury is influenced by multiple HLA class I and class II alleles. *Y. Shen[1,5], M.L. Lucena[2], M. Molokhia[3], T.J. Urban[4], A. Floratos[1], I. Pe'er[1,5], M. Daly[6], D. Goldstein[4], J. Dillon[7], M.R. Nelson[8], P.B. Watkins[9], G.P Aithal[10], R.J. Andrade[2], C.P Day[11], F. Ruiz-Cabello[12], P.T. Donaldson[11], C. Stephens[2], M. Pirmohamed[13], M. Eichelbaum[14], M.R. Romero-Gomez[15], J.M. Navarro[16], R.J. Fontana[17], M. Miller[7], M. Groom[7], E. Bondon-Guitton[18], A. Conforti[19], B.H.C. Stricker[20], A. Carvajal[21], L. Ibanez[22], A.K. Daly[11], Spanish DILI Registry, EUDRAGENE, DILIN, DILIGEN and International SAEC.* 1) Center for Computational Biology and Bioinformatics, Columbia University, New York, NY; 2) Facultad de Medicina, Hospital Universitario Virgen de la Victoria, Campus Universitario s/n, Málaga, Spain; 3) London School of Hygiene & Tropical Medicine, London, UK; 4) Duke University, Durham, NC; 5) Department of Computer Science, Columbia University, New York, NY; 6) Harvard Medical School, Boston, MA; 7) University of Dundee, Scotland; 8) GlaxoSmithKline, Research Triangle Park, NC; 9) Hamner-UNC Institute for Drug Safety Sciences, Research Triangle Park, NC; 10) Nottingham Digestive Diseases Centre: NIHR Biomedical Research Unit, Nottingham, UK; 11) Newcastle University, Newcastle upon Tyne, UK; 12) Hospital Universitario Virgen de las Nieves, Facultad de Medicina, Granada and Red Genómica del Cáncer, Spain; 13) University of Liverpool, Liverpool, UK; 14) Dr. Margarete Fischer-Bosch Institute for Clinical Pharmacology, Stuttgart, Germany; 15) Hospital Universitario de Valme, Sevilla, Spain; 16) Hospital Costa del Sol, Marbella, Málaga, Spain; 17) University of Michigan Medical School, Ann Arbor, Michigan; 18) Université de Toulouse, Toulouse, France; 19) University Hospital, Verona, Italy; 20) Erasmus Medical Center, Rotterdam, Netherlands; 21) Universidad de Valladolid, Valladolid, Spain; 22) FICF (Catalan Institute of Pharmacology Foundation).

Background Drug-induced liver injury (DILI) is an important cause of serious liver disease. Amoxicillin-clavulanate is a leading cause of DILI worldwide, but genetic susceptibility to this adverse reaction is poorly understood.

Methods We performed a genomewide association study (GWAS) with 201 White European and US cases of amoxicillin-clavulanate-associated DILI and 532 population controls matched for genetic background.

Results Extensive genome-wide significant associations in the major-histocompatibility-complex (MHC) were observed. The strongest effect was with a human leukocyte antigen (HLA) class II SNP (rs9274407, $p=4.8\times10^{-14}$) related to *HLA-DRB1*1501-DQB1*0602*, a previously reported risk factor. A novel independent association was observed in a class I region SNP (rs2523822, $p=1.8\times10^{-10}$) related to *HLA-A*0201*. There was significant statistical interaction between these two SNPs (rs9274407 and rs2523822; p= 0.0015). High resolution HLA genotyping in 177 cases and 219 controls confirmed the associations of *HLA-A*0201* ($p=2\times10^{-6}$), *HLA-DQB1*0602* (p= 5×10^{-10}) and their interaction (p=0.005). Additional population-dependent effects from the MHC region were observed with an association of *HLA-DQB1*0402* in cases of Northwestern European origin ($p=4.9\times10^{-3}$) and *HLA-B*1801* in cases from Spain (p=0.004). Finally, there was evidence of the association of rs2476601 from gene *PTPN22* (1858C>T) ($p=1.3\times10^{-4}$; p=0.02 after Bonferroni correction considering all reported GWAS hits of autoimmune diseases).

Conclusions We have shown the influence of class I and II HLA genotypes in susceptibility to amoxicillin-clavulanate DILI. The effect of *HLA-DRB1*1501-DQB1*0602* was confirmed but the novel class I association points to an independent role for *HLA-A*0201* and a possible role for *HLA-B*1801*. The population-dependent effects may be the result of different patterns of linkage disequilibrium with causal alleles in nearby genes. The associations from both the MHC region and *PTPN22* affirm the importance of the adaptive immune response in amoxicillin-clavulanate DILI.

163

Genetic Variations in the APOA5 locus are associated with a positive response to the combined therapy of statins and fenofibric acid in a randomized clinical trial of individuals with atherogenic dyslipidemia. *A. Brautbar[1,2], J. Belmont[1], D. Covarrubias[1], LG. Fremiet[1], SM. Leal[1], CM. Ballantyne[2].* 1) Department of Human Genetics, Baylor College of Medicine, Houston, TX;; 2) Center for Cardiovascular Disease Prevention and Section of Atherosclerosis and Vascular Medicine, Methodist DeBakey Heart and Vascular Center and Baylor College of Medicine, Houston, TX.

Background: Optimizing therapy based on individual genetic variation has clear clinical benefits especially in improving treatment efficacy. Atherogenic dyslipidemia is a phenotype characterized by elevated triglycerides (TG) combined with low high density lipoprotein cholesterol (HDL-C) and is highly associated with coronary heart disease (CHD). Fenofibric acid (FA), usually combined with statins, is a major modality to treat individuals with atherogenic dyslipidemia and can attenuate the risk for future CHD events. Increased HDL-C is considered to constitute a positive response to FA therapy. We hypothesized that genetic variation in candidate genes affect HDL-C response to FA therapy. Methods: A multicenter, randomized, double-blind, active-controlled study, included patients with atherogenic dyslipidemia that were randomized to either FA (135 mg), FA in combination with a statin, or statin alone for a period of 12 weeks. Participants were genotyped for 294 SNPs in genes related to HDL and TG metabolism. Study participants treated with FA only (n=314), combination of FA and statins (n=794), and statin only (n=812) were examined. Results: A multivariate linear regression analysis for percent change in HDL-C adjusted for age, gender, and baseline HDL-C levels was performed. SNP's in the apolipoprotein (APO) A5 locus rs3741298 (P=3.1x10-8), rs964184 (P=1.6x10-7,), and rs10750097 (P= 4.5x10-6) were significantly associated with increase in HDL-C levels in the combined treatment group. The minor allele frequency of the three SNP's was 26.3, 19.6 and 26 percent respectively. The association remained significant after Bonferroni correction for number of SNP's tested. Ten additional SNP's in the APOA5 gene region demonstrated borderline significance. While no significant associations were demonstrated for the statin alone group, the effect of APOA5 SNP's was consistent in the FA only group. Conclusion: We identified three novel associations between SNP's within the APOA5 locus and FA therapy response. At least 20% of individuals with atherogenic dyslipidemia are carriers of one of three minor alleles in the identified SNP's. Further study is required to examine the clinical applicability of APOA5 region genotyping for individuals with atherogenic dyslipidemia receiving combined statin and FA therapy.

164

Association of polymorphisms in *PEAR1* with platelet aggregation and cardiovascular outcomes in patients on aspirin therapy. *J. Lewis[1], J. O'Connell[1], K. Ryan[1], R. Horenstein[1], C. Damcott[1], Q. Gibson[1], T. Pollin[1], B. Mitchell[1], A. Beitelshees[1], R. Pakzy[1], A. Parsa[2], U. Tantry[3], K. Bliden[3], W. Post[4], W. Herzog[4], N. Faraday[5], Y. Gong[6], T. Langaee[6], C. Pepine[7], J. Johnson[6], P. Gurbel[3], A. Shuldiner[1].* 1) Dept. of Medicine, Division of Endocrinology, Diabetes, and Nutrition, University of Maryland, Baltimore, MD; 2) Dept. of Medicine, Division of Nephrology, University of Maryland, Baltimore, MD; 3) Sinai Center for Thrombosis Research, Sinai Hospital of Baltimore, Baltimore, MD; 4) Dept. of Medicine, Division of Cardiology, Johns Hopkins University, Baltimore, MD; 5) Dept. of Medicine, Division of Anesthesiology and Critical Care Medicine, Johns Hopkins University, Baltimore, MD; 6) Center for Pharmacogenomics, Dept. of Pharmacotherapy, University of Florida, Gainesville, FL; 7) Dept. of Medicine, Division of Cardiovascular Medicine, University of Florida, Gainesville, FL.

Dual antiplatelet therapy (DAPT) using aspirin and clopidogrel improves cardiovascular outcomes in patients with acute coronary syndromes and after percutaneous coronary intervention (PCI) by reducing platelet aggregation. However, inter-individual variation in response to these medications is widely recognized and is associated with recurrent cardiovascular events. We measured ex-vivo platelet aggregation before and after DAPT in healthy Amish individuals (n=565) from the Pharmacogenomics of Antiplatelet Intervention (PAPI) study and conducted a genome-wide association study of drug response. We detected strong association between SNPs on chromosome 1q23 near the platelet endothelial aggregation receptor-1 (*PEAR1*) gene and post-aspirin collagen-induced platelet aggregation. Further genotyping revealed rs12041331 in intron 1 of *PEAR1* to be most strongly associated with aspirin response ($p=3.11x10^{-9}$, additive model). These findings were extended by examining the relationship between rs12041331 genotype and cardiovascular outcomes in two independent samples: 227 patients on aspirin undergoing PCI at Sinai Hospital in Baltimore and 361 aspirin-treated subjects of the International VErapamil SR/trandolapril Study (INVEST) GENEtic Substudy (INVEST-GENES). Sinai patients who carried the A-allele of rs12041331 were more likely (25.3% vs. 9.0%) to experience a cardiovascular event or death compared to GG homozygotes (hazard ratio= 3.14, 95%CI 1.48-6.66, p=0.02). Similarly, INVEST-GENES patients who were A-allele carriers had significantly increased risk of fatal or nonfatal myocardial infarction compared to individuals with the GG genotype (OR= 2.19, 95%CI 1.12-4.30, p=0.02). Taken together, genetic variants in *PEAR1* may be important determinants of platelet response to aspirin and poorer cardiovascular outcomes.

165

Using Electronic Medical Records Linked to a Biorepository for Pharmacogenomics: replication of the genetic predictors of warfarin maintenance dose *in the community*. *M.T. Oetjens[1], A.H. Ramirez[2], J. Denny[2,3], J. Schildcrout[4], J.M. Jeff[1], R. Zuvich[1], J.D. Cowan[5], M. Basford[6], E. Bowton[6], J. Pulley[6], R.A. Wilke[2], J. Delaney[2], M. Ritchie[1], D.M. Roden[2,6,7], D. Masys[3], D.C. Crawford[1].* 1) Center for Human Genetics Research; 2) Department of Medicine, Division of Clinical Pharmacology; 3) Department of Biomedical Informatics; 4) Department of Biostatistics; 5) Institute for Clinical and Translational Research; 6) Office of Personalized Medicine; 7) Department of Pharmacology, Vanderbilt University, Nashville, TN.

There is great interest in the use of electronic medical records (EMRs) for large-scale pharmacogenetic studies within a clinical practice-based setting. For a variety of reasons (common clinical indication, high prescription frequency, narrow therapeutic index, and severe adverse drug reactions), warfarin outcome has been one of the most rigorously characterized clinical traits interpreted within a genetic framework. Common promoter variants in warfarin's target, *VKORC1* in the vitamin K pathway, and its primary metabolic pathway in the liver, *CYP2C9*, are strongly associated with dose requirements in European-descent populations. Herein, we replicate these findings within BioVU, the dynamic Vanderbilt DNA Databank containing de-identified EMRs on >86,000 patients to date. EMR-defined steady state warfarin dose was defined for patients having two notes from a Coumadin clinic between three and 12 weeks apart, two or more consecutive INRs between 1.9 and 3.2 (separated by at least one day) in this time frame, with no INR outside the therapeutic range in the same period. Eight SNPs were genotyped using Sequenom in 1,007 European Americans and 23 patients of unknown race/ethnicity that our previous studies have shown are presumably European American. Our quality control criteria excluded SNPs based on minor allele frequency (<1%), tests of HWE ($p<0.05$), and call rates (<95%). Using linear regression assuming an additive genetic model, we performed unadjusted tests of association with four SNPs that passed quality control. The median warfarin dose was 37.0 mg/week. Log transformed median steady-state warfarin dosage was associated with *VKORC1* SNPs rs2359612 ($p = 7.07$ x10-56, $\beta = -0.35$), rs9934438 ($p = 2.49$ x 10-56, $\beta = -0.35$), rs992323 ($p = 11.06$ x 10-55, $\beta = -0.35$) and CYP2C9 rs4917639 ($p = 8.49$x10-31, $\beta = -0.32$). The three *VKORC1* SNPs are in linkage disequilibrium (LD) with one another, and any one of these SNPs explained ~22% of the variability observed in this dataset for warfarin dosing. *CYP2C9* rs4917639, which is in LD with both the *2 and *3 alleles, explained ~12% of the variability observed in this dataset, and combined with *VKORC1*, our genetic model explained ~34% of the variability. We conclude that DNA repositories linked to drug outcome data in an EMR such as BioVU can replicate previously known pharmacogenetic associations and are appropriate for future discovery efforts in the field.

166

Sustained therapeutic hFIX levels in rhesus macaques without neutralizing anti-hFIX antibody development following catheter-based delivery of HDAd. *N. Brunetti-Pierri[1,2], N. Grove[3], P.M. Patel[4], D. Palmer[3], A. Beaudet[3], C. Mullins[4], P. Ng[3].* 1) Telethon Institute of Genetics and Medicine, Naples, Italy; 2) Department of Pediatrics, Federico II University, Naples, Italy; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 4) Department of Cardiology, Baylor College of Medicine, Houston, TX, USA.

Hemophilia B is an excellent candidate for gene therapy because factor IX (FIX) activity ≥1% results in clinically significant improvement of the bleeding diathesis. The rhesus macaque is an excellent animal model for preclinical testing of hemophilia B gene therapy, and helper-dependent adenoviral (HDAd) vectors hold tremendous potential as gene transfer vehicles because they mediate long-term transgene expression without chronic toxicity. Previous studies have shown that administration of first generation adenoviral (FGAd) vectors expressing human FIX (hFIX) in rhesus macaques resulted in short term hFIX expression and the induction of high-titer, neutralizing anti-hFIX antibodies (Lozier et al., 1999. Blood 94:3968). The development of neutralizing anti-hFIX antibodies following administration of FGAd expressing hFIX is a serious adverse event and therefore must be carefully evaluated in preclinical studies for HDAd-mediated hemophilia B gene therapy. To determine the safety and efficacy of HDAd-mediated hemophilia B gene therapy, we have administered an HDAd expressing hFIX into rhesus macaques through a balloon catheter strategy that we have previously developed and shown to result in greater hepatocyte transduction than systemic intravenous administration (Brunetti-Pierri et al., 2009 Mol Ther 17:327). In this approach a balloon catheter is percutaneously positioned in the inferior vena cava to occlude hepatic venous outflow and $1x10^{10}$, $3x10^{10}$, $1x10^{11}$, or $1x10^{12}$ vp/kg of an HDAd expressing hFIX was injected into the liver via a hepatic artery catheter. Animals injected with $1x10^{11}$ and $1x10^{12}$ vp/kg exhibited therapeutic levels of hFIX (>10% of normal FIX activity) for at least 270 days. Sub-therapeutic levels corresponding to »0.5% and »0.1% were achieved in the animal injected with $3x10^{10}$ vp/kg and $1x10^{10}$ vp/kg, respectively. Importantly, neutralizing anti-hFIX antibodies were not detected in any of the injected animals. These results suggest that in contrast to FGAd, HDAd vectors can drive long-term FIX expression, do not result in neutralizing anti-hFIX antibody formation, and $1x10^{11}$ vp/kg is the minimal dose to achieve clinically beneficial levels of FIX. Therefore, HDAd are attractive vectors for hemophilia B gene therapy.

167

A novel protein Lztfl1 regulates ciliary trafficking of the BBSome. *S. Seo[1,2], K. Bugge[1,2], C.C. Searby[1,2], M.V. Nachury[3], V.C. Sheffield[1,2].* 1) Dept Pediatrics, Univ Iowa, Iowa City, IA; 2) Howard Hughes Medical Institute; 3) Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA.

Primary cilia function as cellular antennae sensing and transducing extracellular signals. Malfunction or loss of cilia is linked to various developmental defects and human genetic disorders. Bardet-Biedl syndrome (BBS) is a heterogeneous, autosomal recessive disorder resulting in obesity, retinal degeneration, polydactyly, cardiac and renal defects, and cognitive abnormalities. BBS proteins are involved, at least in part, in maintaining normal ciliary function by mediating protein trafficking to the cilia. However, the mechanisms governing ciliary trafficking by BBS proteins are not well understood. Here, we show that a novel protein Leucine-zipper transcription factor-like 1 (Lztfl1) interacts with the BBSome, a complex composed of seven BBS proteins and BBIP10, and inhibits ciliary trafficking of the BBSome. We performed tandem affinity purification using transgenic mouse testis expressing LAP-BBS4 recombinant protein and isolated the BBSome and its associated proteins. Lztfl1 was co-purified with the BBSome in the mouse testis. It interacts with BBS9 among the BBSome subunits. Depletion of Lztfl1 increases ciliary localization of the BBSome, whereas over-expression of Lztfl1 inhibits BBSome localization to cilia, while general intraflagellar transport (IFT) is not affected. We also show that most BBSome subunits are essential for BBSome assembly except for BBS5, which is not required for the assembly of the remaining subunits, and that all BBSome subunits and BBS3 (also known as Arl6) are required for BBSome ciliary entry. Depletion of Lztfl1 restores ciliary trafficking of the BBSome caused by loss of BBS3 and BBS5. These results uncover Lztfl1 as a negative regulator of BBSome ciliary trafficking and provide a target for therapeutic intervention for BBS.

168

A Bobtail ties CCDC39 to human ciliary defects and the assembly of inner dynein arms and dynein regulatory complexes. *E.E. Davis[1], A.-C. Merveille[2], A. Becker-Heck[3,4], M. Legendre[5], G. Bataille[2], F. Billen[6], C. Clercx[6], S. Deleuze[6], P. Duquesnoy[5], D. Escalier[5], E. Escudier[5], M. Fliegauf[3], J. Horvath[3], M. Jorissen[7], M. Lathrop[8], N.T. Loges[3,9], Y. Momozawa[2], G. Montantin[5], H. Olbrich[3,9], J.-F. Papon[5,10], M. Schmidts[3], H. Tenreiro[5], D. Zelenika[8], M. Georges[2], A.-S. Lequarré[2], N. Katsanis[1], H. Omran[3,9], S. Amselem[5].* 1) Center for Human Disease Modeling, Duke University Medical Center, Durham, NC, USA; 2) Unit of Animal Genomics, GIGA-R & Faculty of Veterinary Medicine, University of Liège, Liège, Belgium; 3) Department of Pediatrics and Adolescent Medicine, University Hospital Freiburg, Freiburg, Germany; 4) Faculty of Biology, Albert-Ludwigs-University, Freiburg, Germany; 5) Institut National de la Santé et de la Recherche Médicale (INSERM) U.933, Université Pierre et Marie Curie-Paris 6 and Assistance Publique-Hôpitaux de Paris (AP-HP), Hôpital Armand-Trousseau, Paris Cedex, France; 6) Department of Clinical Sciences, Faculty of Veterinary Medicine. University of Liège, Belgium; 7) Department of Otorhinolaryngology, Head and Neck Surgery, University Hospitals Leuven, Leuven, Belgium; 8) Centre National de Génotypage, Evry Cedex, France; 9) University Children's Hospital Muenster; Department of General Pediatrics, Münster, Germany; 10) AP-HP, Hôpital Intercommunal et Groupe Hospitalier Henri Mondor-Albert Chenevier, Service d'ORL et de chirurgie cervico-faciale, Créteil, France.

Primary Ciliary Dyskinesia (PCD) is a rare hereditary disorder characterized by recurrent infections of the upper and lower respiratory tract, reduced fertility in males, and situs inversus in ~50% of patients, referred to as Kartagener syndrome. The complex phenotype is caused by motility defects of respiratory cilia responsible for airway clearance, flagella propelling sperm cells, and nodal monocilia essential for correct determination of left-right asymmetry. Recessive mutations in several genes encoding components of the outer dynein arms, radial spokes and cytoplasmic pre-assembly factors of axonemal dynein have been identified but only account for ~50% of cases. In this work, we exploit the unique properties of dog populations to positionally clone a novel PCD gene: *CCDC39*. We demonstrate that loss-of-function mutations in the human ortholog underlie a significant fraction of PCD cases with axonemal disorganization and abnormal ciliary beating. Functional analyses indicate that CCDC39 localizes to ciliary axonemes, and is essential for assembly of inner dynein arms as well as the dynein regulatory complex.

169

Candidate Exome Capture Reveals Mutations in SDCCAG8 as Causing a Retinal-Renal Ciliopathy. *E.A. Otto[1], T. Hurd[1], R. Airik[1], M. Chaki[1], W. Zhou[1], A.K. Ghosh[1], R.H. Giles[2], P. Nürnberg[3], E. Pierce[4], P. Jackson[5], C. Antignac[6], S. Saunier[6], R. Roepman[7], H. Khanna[8], F. Hildebrandt[1,9,10].* 1) Dept Ped, Univ Michigan, Ann Arbor, MI; 2) Dept of Medical Oncology, University Medical Center, Utrecht, The Netherlands; 3) Cologne Center for Genomics, Univ of Cologne, Cologne, Germany; 4) F.M. Kirby Center for Molecular Ophthal, Univ of Pennsylvania School of Medicine, PA; 5) Dept of Cell Regulation, Genentech Inc., South San Francisco, CA; 6) Dept of Genetics, Hopital Necker-Enfants Malades, Paris, France; 7) Dept of Human Genetics, Radboud Univ Nijmegen, Nijmegen, The Netherlands; 8) Dept of Ophthal and Visual Sciences, Ann Arbor, MI; 9) Dept of Human Genetics, Univ of Michigan, Ann Arbor, MI; 10) Howard Hughes Medical Institute.

Nephronophthisis-related ciliopathies (NPHP-RC) are recessive disorders featuring dysplasia or degeneration preferentially in kidney, retina, and cerebellum. We combined homozygosity mapping with candidate gene analysis by performing "ciliopathy candidate exome capture" of 828 candidate genes followed by next-generation sequencing on a Illumina platform. Candidates were derived from ciliopathy animal models, the centrosome proteome, and from the photoreceptor sensory cilia proteome. Altogether, we detected 6 different homozygous null mutations in the gene *SDCCAG8* (serologically defined colon cancer antigen 8) in 5 NPHP-RC families. Subcellular localization studies revealed that SDCCAG8 occurs at centrosomes throughout the cell cycle and is located e.g. in mouse photoreceptor basal bodies and connecting cilia transition zone together with other NPHP-RC proteins (e.g. NPHP5). We identified OFD1 (Oral-facial-digital syndrome 1), another cystic kidney disease causing NPHP-RC protein, as a direct interaction partner of SDCCAG8 by a yeast 2 hybrid screening. Depletion of sdccag8 in zebrafish resulted in kidney cysts and body axis defects. Furthermore, siRNA knockdown of Sdccag8 disturbed lumen formation of renal epithelial cells in 3D spheroid cultures, indicative for a cell polarity defect. In addition to the centrosomal localization we observed SDCCAG8 at cell-cell junctions. Interestingly, SDCCAG8 abandons these cell-cell junctions in response to increased intracellular cAMP levels. This is especially of interest, because in the nphp3 loss of function mouse model ("*pcy*"), pharmacological reduction of intracellular cAMP levels successfully delayed the development of renal cystic disease. In summary, this work identifies SDCCAG8 loss of function as a novel cause of a retinal-renal ciliopathy and validates targeted candidate exome capture analysis for broadly heterogeneous single-gene disorders.

170

Exome sequencing identifies WDR35, encoding an IFT-A protein, as a novel gene involved in Sensenbrenner syndrome. *C. Gilissen[1], H.H. Arts[1], A. Hoischen[1], L. Spruijt[1], D.A. Mans[1], P. Arts[1], B. van Lier[1], M. Steehouwer[1], J. van Reeuwijk[1], S.G. Kant[2], R. Roepman[1], N.V.A.M. Knoers[1], J.A. Veltman[1], H.G. Brunner[1].* 1) Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands; 2) Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands.

Sensenbrenner syndrome/cranioectodermal dysplasia (CED) (MIM %218330) is an autosomal recessive disease which is characterized by craniosynostosis, ectodermal and skeletal abnormalities. This phenotype shows remarkable overlap with the ciliopathies, a spectrum of disorders which is associated with dysfunction of the cilium, an antenna-like organel that occurs in almost all vertebrate cells. Therefore, it has been suggested that the genetic defects that cause CED also involve ciliary genes. We enriched the exomes (targeting ~18,000 genes) of two unrelated CED patients using the SureSelect human exome kit (Agilent, Santa Clara, CA, USA) and subsequently sequenced each exome on one quarter of a SOLiD sequencing slide (Life Technologies, Carlsbad, CA, USA). More than 89% of the targeted exons was covered more than 10 times, which is sufficient for calling heterozygous variants. On average 12,736 genetic variants were identified per patient in the coding regions or the canonical dinucleotide of the splice sites, including 5,657 non-synonymous changes. For each patient we excluded known variants and selected variants consistent with the recessive inheritance pattern of the disease. We identified three and four candidate genes with compound heterozygous variants for patient one and two respectively. The inheritance of the variants in these genes was subsequently determined by Sanger sequencing which yielded WDR35 as the single candidate gene in patient 1, and as one of only two candidate genes in patient 2. The only other gene that appeared to have bi-allelic variants in patient 2, was USH2A but the change on one allele was in a non-conserved amino acid. In contrast all four WDR35 variants detected in the two Cranioectodermal Dysplassia patients affected strictly conserved amino acids. None of the four identified variants in WDR35 were detected in 180 control alleles. The Chlamydomonas orthologue of WDR35 (IFT122B) is part of the intraflagellar transport complex A. Independently, another member of this complex, IFT112A has recently been shown to be involved in CED as well. Our genetic data confirm that CED is a ciliary disorder. Furthermore, our data suggest that for a rare recessive condition, it is possible to find the causative gene by sequencing the exome of a single sporadic patient.

171

Utilization of whole-exome sequencing to identify genetic causes of primary ciliary dyskinesia. *MA. Zariwala[1], MW. Leigh[2], SD. Davis[2], AS. Cutting[3], L. Huang[3], SL. Minnix[3], EH. Turner[4], A. MacKenzie[4], J. Shendure[4], MR. Knowles[3].* 1) Dept. Pathology and Laboratory Medicine, University of North Carolina, Chapel Hill, NC; 2) Dept. Pediatrics, University of North Carolina, Chapel Hill, NC; 3) UNC School of Medicine, University of North Carolina, Chapel Hill, NC; 4) Dept. Genome Sciences, University of Washington, Seattle, WA.

Primary ciliary dyskinesia (PCD)/Kartagener syndrome is a genetically heterogeneous, usually autosomal recessive disorder that affects structure and function of cilia and flagella. Clinical manifestations include: neonatal respiratory distress despite full term birth, recurrent otitis media, chronic sinusitis, chronic bronchitis leading to bronchiectasis, and male infertility. Approximately half of the PCD patients have situs inversus and at least ~6% have situs ambiguus (heterotaxy). At present, the "gold standard" diagnostic test for PCD is ultrastructural analysis of the cilia. Most PCD patients (80-90%) have defective outer dynein arms (ODA), and/or inner dynein arms (IDA). A few patients have defective radial spokes (RS) or central apparatus (CA). Ultrastructural studies are challenging. In some patients, PCD is confirmed by genetic studies despite normal ciliary ultrastructure. Thus far, disease-causing mutations have been identified in 9 different genes that account for only 1/3rd of PCD subjects tested. These genes encode ODA proteins (*DNAH5*, *DNAI1*, *DNAI2*, *DNAH11* and *TXNDC3*), RS proteins (*RSPH9*, *RSPH4A*) or cytoplasmic proteins involved in DA assembly (*KTU* and *LRRC50*). Due to the extensive locus heterogeneity, we utilized whole-exome capture coupled with massively parallel sequencing to discover novel mutations in 24 unrelated, well-characterized PCD patients who were negative for the most common PCD-causing genes (*DNAH5*, *DNAI1* and *DNAH11*). These patients were sub-grouped based on the ultrastructural findings: isolated ODA defects (n=6), isolated IDA defects (n=6), ODA+IDA defects (n=3), RS+CA defects (n=3), acilia on multiple biopsies (n=1) and normal DA with compatible clinical phenotype and low nasal nitric oxide (n=5). To date, we have identified a novel homozygous stop mutation in a relatively rare PCD-causing dynein gene in three unrelated subjects with PCD and ciliary ODA defects. All three had Jewish ethnicity; haplotype analysis suggested a founder effect. Identification of this novel founder mutation suggests that mutations in this gene may be more prevalent, particularly in subjects of Jewish ethnicity. In conclusion, we demonstrate that exome-sequencing can successfully define novel disease-causing genes/mutations for genetically heterogeneous disorders such as PCD. This abstract was funded by MO1RR00046, UL 1 RR025747, 5 R01HL071798, 5 U54 HL096458-06. Exome-sequencing was provided by NHLBI/NHGRI Exome project 5 R01HL094976-02.

172

Downregulation of Hedgehog Signaling Prevents Renal Cystogenesis in Mouse Models of Ciliopathies. *D.R. Beier, P.V. Tran.* Genetics Division, Brigham & Women's Hospital, Harvard Medical School, Boston, MA.

Cystic kidney disease (CKD), a leading cause of renal failure, is proposed to originate from an underlying ciliary defect. There are no proven therapies for CKD and molecular mechanisms remain unclear. In our characterization of the ENU mutant mouse alien *(aln)*, we identified a novel ciliary protein, THM1 (Tetratricopeptide Repeat Containing Hedgehog Modulator 1, also IFT139), which negatively regulates Hedgehog (Hh) signaling. Mutations in IFT139 have been identified in patients with several ciliopathies that feature renal cysts as a major clinical component, including nephronophthisis, Bardet-Biedl Syndrome and Meckel-Gruber Syndrome. This prompted us to examine the kidney phenotype in aln. By E16.5, aln develops renal cysts in the proximal tubules and ascending loops of Henle. Importantly, cyst formation was markedly reduced in *aln, Gli2-/-* double mutant mice, demonstrating that genetic deletion of *Gli2*, the main transcriptional activator of Hh signaling, prevents *aln* cyst formation, implicating a role for increased Hh signaling in the etiology of *aln* renal cystogenesis. To explore whether the Hh pathway might present a novel target for CKD, we examined the effects of small molecule Hh inhibitors in a cAMP-induced cystogenic kidney explant assay. In the presence of cAMP, cultured *aln* kidneys exhibited a three-fold greater cystogenic potential than wild-type. This was prevented by small molecule Hh inhibitors, Gant61 or Sant2, supporting a role for increased Hh signaling in aln renal cystogenesis. Surprisingly, the small molecules also abrogated cAMP-induced cysts in wild-type kidneys, indicating a beneficial role for Hh inhibitors in cAMP-mediated renal cystogenesis, which is proposed as a general mechanism in CKD. We thus questioned whether the preventive effects of Gant61 and Sant2 might extend to other models of CKD, such as the *jck* mutant, a mouse model for NPHP9. This mutant develops severe CKD by P21 and also showed a three-fold higher cystogenic potential than wild-type using the cAMP cystogenic assay. This increased cystogenesis was dramatically reduced by both Gant61 and Sant2. Our results indicate a protective role for small molecule Hh antagonists in CKD and demonstrate the potential of developing Hh inhibitors as targeted therapies.

173

Mutations in TMEM216 perturb ciliogenesis and cause Joubert, Meckel and related syndromes. *J.H. Lee[1,5], E.M. Valente[2,5], C.V. Logan[3,5], S. Mougou-Zerelli[4,5], J.L. Silhavy[1], C.A. Johnson[3,6], T. Attie-Bitach[4,6], J.G. Gleeson[1,6].* 1) Neurogenetics Laboratory, Institute for Genomic Medicine, Department of Neurosciences and Pediatrics, Howard Hughes Medical Institute, University of California, San Diego, California, USA; 2) Mendel Laboratory, Istituto di Ricovero e Cura a Carattere Scientifico "Casa Sollievo della Sofferenza," San Giovanni Rotondo, Italy; 3) Section of Ophthalmology and Neurosciences, Wellcome Trust Brenner Building, Leeds Institute of Molecular Medicine, St. James's University Hospital, Leeds, UK; 4) Department de Génétique, INSERM U781, Hôpital Necker-Enfants Malades, Université Paris Descartes, Paris, France; 5) Contributed equally to the work; 6) Jointly directed the project.

Joubert syndrome (JBTS), related disorders (JSRDs) and Meckel syndrome (MKS) are ciliopathies. We now report that MKS2 and CORS2 (JBTS2) loci are allelic and caused by mutations in TMEM216, which encodes an uncharacterized tetraspan transmembrane protein. Individuals with CORS2 frequently had nephronophthisis and polydactyly, and two affected individuals conformed to the oro-facio-digital type VI phenotype, whereas skeletal dysplasia was common in fetuses affected by MKS. A single G218T mutation (R73L in the protein) was identified in all cases of Ashkenazi Jewish descent (n = 10). TMEM216 localized to the base of primary cilia, and loss of TMEM216 in mutant fibroblasts or after knockdown caused defective ciliogenesis and centrosomal docking, with concomitant hyperactivation of RhoA and Dishevelled. TMEM216 formed a complex with Meckelin, which is encoded by a gene also mutated in JSRDs and MKS. Disruption of tmem216 expression in zebrafish caused gastrulation defects similar to those in other ciliary morphants. These data implicate a new family of proteins in the ciliopathies and further support allelism between ciliopathy disorders.

174

A new culprit for retinal ciliopathies: cc2d2a is required for retinal photoreceptor outer segment formation/function in a zebrafish model of Joubert syndrome. *R. Bachmann-Gagescu[1,4], IG. Phelps[3], SE. Brockerhoff[2], G. Stearns[2], CB. Moens[4], D. Doherty[3].* 1) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA; 2) Department of Biochemistry, University of Washington, Seattle, WA; 3) Division of Genetics and Developmental Medicine, Department of Pediatrics, University of Washington, Seattle, WA; 4) Fred Hutchinson Cancer Research Center, Seattle, WA.

Ciliopathies are a rapidly expanding group of disorders caused by dysfunction of primary cilia. Joubert syndrome is a ciliopathy characterized by a distinctive hindbrain malformation variably associated with retinal dystrophy, renal cystic disease, liver fibrosis and polydactyly. Mutations in CC2D2A are found in ~10% of patients with Joubert syndrome. Here we describe a zebrafish model for CC2D2A-related Joubert syndrome that displays, in addition to a sinusoidal body shape and pronephric (kidney) cysts, a retinal phenotype with shortened and abnormal photoreceptor outer segments. Vertebrate photoreceptor outer segments are highly specialized primary cilia containing large amounts of photosensitive pigments, or opsins, required for the phototransduction cascade. Given the lack of organelles in the outer segment, all its components are synthesized in the inner segment and require a specialized transport mechanism termed intraflagellar transport (IFT) for proper localization to the outer segment. IFT is crucial for the function and formation of the outer segment as zebrafish and mouse mutants with loss of IFT function display abnormal or absent photoreceptor outer segments as well as mislocalization of opsins. The retinal phenotype of cc2d2a mutant zebrafish includes mislocalization of opsins throughout the cell membrane, inner segment and basal synapse, similar to what is seen in fish with mutations in IFT genes such as ift88. Both rod and cone photoreceptors are affected in cc2d2a mutants and retinal function is abnormal, as demonstrated by a markedly attenuated electroretinogram response. Electron microscopy reveals dysmorphic outer segments with aberrant membrane structures. Connecting cilia are present and appear structurally normal, but ciliary axonemes are shortened in photoreceptors. Primary cilia length is unaffected in other organ systems including the otic vesicle, olfactory pit and pronephric duct, indicating that Cc2d2a is not required for cilia assembly. Partial loss of ift88 function in cc2d2a mutants results in a synergistic pronephric cyst phenotype, indicating that ift88 can act as a genetic modifier for cc2d2a. Based on their similar phenotypes in zebrafish and this genetic interaction, we propose that cc2d2a and ift88 function as part of a shared genetic network during pronephros development, outer segment formation/function and opsin localization.

175

Molecular basis of a syndromic form of ciliopathy combining primary ciliary dyskinesia and Usher syndrome. *M. Legendre[1,2], P. Duquesnoy[2], H. Tenreiro[1], L. Vincensini[2], A. Deschildre[3], A. Delattre[4], D. Escalier[2], E. Escudier[1,2], S. Amselem[1,2].* 1) Service de Génétique et d'Embryologie médicales, APHP, Hôpital Trousseau, Paris, France; 2) Inserm U933, Hôpital Trousseau, Paris, France; 3) Service de Pneumologie pédiatrique, Hôpital Jeanne de Flandres, Lille France; 4) Service d'ORL, Hôpital Roger Salengro, Lille, France.

Cilia are evolutionarily conserved structures that play key roles in diverse cell types. Motile cilia are involved in the most prominent ciliopathy called primary ciliary dyskinesia (PCD, prevalence 1/15000) that combines respiratory symptoms, male infertility, and, in nearly 50% cases, situs inversus. In rare instances, PCD is syndromic; in particular, the PCD phenotype can be associated with retinitis pigmentosa. This kind of association is readily explained by mutations in RPGR, a protein involved in intraflagellar transport (IFT), a physiological process that occurs within the axonemes found both in cilia and in photoreceptors. On the other hand, several genes have been identified in Usher syndromes (prevalence 1/30000) characterized by the association of retinitis pigmentosa and deafness, with no respiratory involvement. Here, we describe the molecular basis of a complex syndrome associating, in two consanguineous siblings, a PCD characterized by the absence of outer dynein arms and an Usher syndrome type 1, thereby suggesting the existence of a molecular defect involving a protein common to the axonemes found in respiratory cilia and in the kinocilium of inner ear cells. Molecular analyses revealed a homozygous deletion of exons 3 and 4 of DNAI2 (p.Ala62GlyfsX115), which most likely results from an illegitimate recombination between two Alu sequences located in introns 2 and 4. As DNAI2 is known to be involved in non-syndromic PCD, these data prompted us to search for molecular defects in a second gene that could account for the Usher syndrome. The analysis of USH1G, located in the vicinity of DNAI2, within a common region of homozygosity, revealed a homozygous nonsense mutation (p.Lys218X) which would lead to a truncated protein, lacking the SAM domain know to interact with harmonin. The molecular defects identified in these two genes (DNAI1 and USH1G), which have not been reported previously, were found in the heterozygous state in the healthy parents of two patients. While the syndromic pattern of several ciliopathies is readily explained by the pleiotropic functions of ciliary proteins, the current observation illustrates the occurrence two rare ciliopathies that result from mutations in two tightly linked genes.

176

Leber Congenital Amaurosis Associated With Abnormalities of Respiratory Cilia. *JM. Rozet[1], I. Perrault[1], JF. Papon[2], A. Coste[2], B. Louis[3], C. Hanein[1], L. Fares Taie[1], S. Gerber[1], E. Escudier[4], J. Kaplan[1].* 1) INSERM U781, Hopital des Enfants Malades, Paris, CDX 15, France; 2) INSERM U955 - Paris 12 University - CHU Henri Mondor, Créteil, France; 3) Physiology and Functional Explorations, INSERM UMR651 - Paris 12 University - CHU Henri Mondor, Créteil, France; 4) Anatomo-Pathology, INSERM U933 - UPMC - CHU Armand Trousseau, Paris, France.

Purpose: CEP290 mutations account for either pleiotropic phenotypes including Senior-Loken and Joubert syndromes or non-syndromic Leber congenital amaurosis. In the retina, CEP290 associates with microtubule-based transport proteins including RPGR. Interestingly, some RP patients with RPGR mutations were reported to be affected with primary ciliary dyskinesia. We made the hypothesis that CEP290 patients with non-syndromic LCA may exhibit infra-clinical symptoms resulting from abnormal functioning of motile respiratory cilia. Methods: 7 patients in 6 families were ascertained. They were either homozygote for loss-of-function alleles or the hypomorphic CEP290 c.2991_1655A>G mutation or compound heterozygote for each mutation type.Patients were examined for ENT symptoms and olfaction. Airway biopsies were obtained from nasal mucosa to allow transmission electron microscopy (TEM) analysis of cilia ultrastructure and High-speed videomicroscopy (HSV) for evaluation of ciliary beat pattern. Results: 6/7patients had ENT history and presented with constant edema and/or congestion of the mucosa. Olfaction acuity was slightly impaired in 3 patients and 1presented with slight unilateral hypoacousia. One patient had no reported ENT history but he presented oligoasthenospermia and bilateral transmission hypoacousia.TEM analyses identified absence of cilia in two patients with an epithelial morphology suggestive a loss of ciliary differentiation. In the other 5 patients, high levels of abnormal cilia were noted (mean 48.2 %). Axonemal structural defects of peripheral microtubules, dynein arms and central microtubules were evidenced alone or in combination. HSV showed significantly decreased numbers of ciliated cells compared to controls and significant percentages of short cilia i.e.<4μm (mean 13.6%). 85 % to 100 % of cells were beating but slight or significant beating abnormalities were evidenced in 4 patients. No genotype correlation could be drawn. Conclusions: These data provide the first argument for respiratory cilia abnormalities in non-syndromic LCA due to CEP290 mutations demonstrating that this non-syndromic LCA should be regarded as a member of the fast increasing family of ciliopathies.

177

Whole-Genome Sequencing and High-Density SNP Genotyping for Type-2 Diabetes. *B.F. Voight[1], J.R. Maguire[1], H.M. Kang[2], T.M. Teslovich[2], K.J. Gaulton[3], R.D. Pearson[3] on behalf of the Genetics of Type-2 Diabetes (GO-T2D) Consortium.* 1) Medical Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.

Although rare Mendelian mutations and common genetic variants have been implicated in susceptibility to type-2 diabetes (T2D), the large majority of heritability has yet to be explained. Additional heritability may be contributed by lower frequency variation, as previously demonstrated in other complex traits (e.g., PCSK9, IFIH1, IL23R, and NOD2 for lipid and autoimmune traits). Optimal strategies for evaluating and testing the role of lower frequency polymorphisms and rare variants broadly remain uncertain.

We are evaluating three strategies designed to systematically interrogate low frequency polymorphism in relation to T2D: imputation from the emerging 1000 Genomes Project, whole-genome low-pass (4x) sequencing, and next-generation 5M SNP chip genotyping in each of 2,650 samples (T2D and euglycemic controls). These analyses will be used to evaluate the performance of each technology, and to test lower frequency variants for association to T2D. Notably, the low-pass sequencing and genotype data will provide a deep and rich imputation resource facilitating association testing in large numbers of samples (T2D cases and controls) previously genotyped on existing whole-genome platforms.

At the time of this abstract submission, we have completed sequencing of 33 type-2 diabetics and 34 associated controls. Preliminary variant calling on 22 samples resulted in the detection of 6.58M high-quality SNPs -- 15% of which were not previously identified in dbSNP b129 or from the 1000 Genomes Pilot Project. In September, we will establish a first-term project freeze of approximately 500 whole-genomes, and will present an initial characterization of these data. We will present an assessment of sequence quality and variant calling and an initial association analysis of rare variation in these samples. We will also use this data impute and analyze variants in additional whole genome data from the WTCCC, FUSION, and DGI studies (totaling 4,100 cases and 5,200 controls for meta-analysis). This analysis will provide an initial evaluation of the proposed strategies for discovery of new genes for complex, common diseases, post the 1000 Genomes Project.

178

Whole-genome sequencing to identify the genetic basis for resistance to HIV-1 infection. *K. Pelak[1], D. Ge[1], J. Fellay[1], K.V. Shianna[1], J.J. Goedert[2], D.B. Goldstein[1].* 1) Center for Human Genome Variation, Duke University School of Medicine, Durham, NC, 27708, USA; 2) Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, 20892, USA.

Human genetic factors have been shown to contribute to clinical variability of viral infections in humans. In the HIV field, host genetic variants have been identified that contribute to susceptibility to infection, viral load at setpoint, and disease progression following infection. However, much of the observed variability in these phenotypes still has not been explained. People who are homozygous for a deletion in the CCR5 gene (CCR5d32), which encodes a receptor that HIV uses to enter the cell, are resistant to HIV infection. About 1% of people of northern and western European descent are homozygous for this deletion. In a population of hemophiliacs who were exposed to contaminated blood products, the proportion of CCR5d32 homozygotes increases to about 25%. However, much of the “HIV-resistance” observed in the remainder of this population has not been explained either by host or viral genetic variants. In this study, we seek to identify additional host genetic variants that may provide resistance to HIV infection. We used the Illumina Genome Analyzer IIx to sequence the genomes of 30 hemophiliacs of European ancestry who were highly exposed to contaminated blood products between 1979 and 1984, but who did not become HIV-positive. We then compared these whole-genome sequences to the genomes of 40 controls that were not exposed to HIV. We expected to see an enrichment of any potentially protective variant in the cases as compared to the controls, and any variant that was at equal or greater frequency in the controls was dropped from further consideration. From this list of potentially protective variants, we then checked the frequency of all coding variants against a population of confirmed HIV-positive people who had undergone whole-exome sequencing. We expected any protective variant to be absent from this population. Using the information provided by these two control cohorts, we chose the most promising variants/genes to follow up by genotyping and Sanger sequencing in a larger population. Within the variants chosen for further follow-up, there were 50 nonsynonymous SNPs and 9 frameshift indels in genes that were previously shown to interact with the HIV virus in a series of RNAi screens. These variants occurred in at least two cases and were absent from the controls. And, there were 26 genes that were observed to be a homozygous knock out in at least 2 of the cases and none of the controls.

179

Exome Sequencing Identifies ANGPTL3 as a Cause of Familial Combined Hypolipidemia. *R. Do[1,2,3], K. Musunuru[1,2,3,4], J.P. Pirruccello[1,2,3], C. Guiducci[3], E. Gonzalez[3], C. Sougnez[3], K.V. Garimella[3], S. Fisher[3], J. Abreu[3], A.J. Barry[3], T. Fennell[3], E. Banks[3], L. Ambrogio[3], K. Cibulskis[3], A. Kernytsky[3], M.A. DePristo[3], M.J. Daly[2,3,4], D. Altshuler[2,3,4,5], G. Schonfeld[6], S.B. Gabriel[3], P. Yue[6], S. Katherisan[1,2,3,4].* 1) Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, United States; 2) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, United States; 3) Broad Institute, Cambridge, MA, United States; 4) Department of Medicine, Harvard Medical School, Boston, MA, United States; 5) Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, United States; 6) Washington University in St. Louis School of Medicine, St. Louis, MO, United States.

Introduction: Familial hypobetalipoproteinemia (FHBL) is an inherited lipid disorder defined by low serum apolipoprotein B (apoB) and low-density lipoprotein cholesterol (LDL-C) (< 5th percentile). FHBL has been linked to mutations in the *APOB* gene, but in other families, *APOB* sequencing failed to identify mutations. We studied a non-apoB-defective FHBL family with 38 members across 3 generations and sought to identify the causal gene with exome sequencing, an emerging unbiased method to identify genes underlying monogenic disorders. **Methods:** Affected members displayed low LDL-C and triglycerides (TG), with 4 individuals having combined hypolipidemia—extremely low LDL-C (29-37 mg/dL), TG (17-24 mg/dL), and HDL-C (16-22 mg/dL). We selected two siblings with combined hypolipidemia for exome sequencing using solution hybrid selection (Agilent SureSelect) and the Illumina GA-II platform. Sanger sequencing was used to genotype the other family members for identified mutations. **Results:** We targeted 28,646,006 bases at 164,688 exons from 15,994 genes. For each individual, we generated ~6 billion bases of sequence, with each targeted base having an average of 200X coverage. Each individual harbored ~270 novel single nucleotide variants, of which ~150 were new missense mutations and ~6 new nonsense mutations. To identify the causal variants, we hypothesized that the combined hypolipidemia was inherited in an autosomal recessive fashion. Therefore, we sought genes harboring novel variants in both alleles in both sequenced individuals. The two sequenced probands with combined hyolipidemia were found to be compound heterozygotes for two distinct nonsense mutations (E129X and S17X), both in the first exon of the *ANGPTL3* gene. Two additional compound heterozygotes and 13 individuals who carried one *ANGPTL3* nonsense mutation were also identified. *ANGPTL3* nonsense mutations decreased LDL-C ($P = 8 \times 10^{-6}$) and TG ($P = 0.004$) under an additive model, whereas HDL-C was decreased under a recessive model ($P = 9 \times 10^{-5}$), suggesting distinct mechanisms of action of *ANGPTL3* on different lipoproteins. **Conclusion:** We have characterized a novel Mendelian lipid disorder— “familial combined hypolipidemia” — caused by nonsense mutations in both copies of *ANGPTL3*. Our finding in humans highlights an unexpected role of *ANGPTL3* in LDL-C metabolism and demonstrates the utility of unbiased exome sequencing for identification of novel genetic causes of inherited disorders.

180

Exome sequencing identifies thousands of novel candidate mutations in coeliac disease. *V. Mistry[1], N. Bockett[1], K.A. Hunt[1], C.A. Mein[2], M. Simpson[3], R. Trembath[3], I. Stanaway[4], D.A. Nickerson[4], V. Plagnol[5, 6], D.A. van Heel[1].* 1) Institute of Cell and Molecular Science, Barts and The London School of Medicine and Dentistry, 4 Newark Street, London E1 2AT, UK; 2) The Genome Centre, Barts and The London School of Medicine and Dentistry, Charterhouse Square, London EC1M 6BQ, UK; 3) Department of Medical and Molecular Genetics, Kings College London, Guys Hospital, Great Maze Pond, London SE1 9RT, UK; 4) Department of Genome Sciences and Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA; 5) UCL Genetics Institute (UGI), Kathleen Lonsdale Building, 2nd Floor Gower Place, London WC1E 6BT, UK; 6) Department of Medical Genetics, Juvenile Diabetes Research Foundation/ Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 0XY, UK.

Coeliac disease is a highly heritable common autoimmune disease involving chronic small intestinal inflammation in response to dietary wheat. The HLA and ~40 newer GWAS loci (many immune genes) account for ~40% of heritability. We hypothesise that rare mutations of larger effect size (OR~2-5) exist, possibly missed by linkage and common variant GWAS.

We exome resequenced (180,000 protein coding CCDS exons, Nimblegen in-solution capture) 44 coeliac disease cases. One Illumina GAIIx 76bp PE lane provided an optimal ~50-60x mean read-depth (on-target non-duplicate reads). Base call quality aware gapped alignment (novoalign) and a custom Bayesian SNP caller was near-identical to BWA/samtools/picard/GATK at high read depth, but more sensitive at lower read depth for HapMap NA12878.

We identified a mean of 15,601 non-reference sequence exonic SNPs per sample (~40% protein altering i.e. missense, nonsense, or splice site). In all 44 case samples we identified a total of: 28,328 protein altering SNPs (of which 8,869 were novel - not seen in dbSNP129, 1000Genomes CEU or 47 control exomes); 318 protein altering small indels (48% 3bp multiples, largest 10bp); and 55 large deletions of ≥2 exons (range 200bp to 2Mb). These data present a challenge to identify disease causing mutations from many thousands of candidate variants. Strategies considered were: A) Near-private biologically plausible case-only mutations: total 335 novel, high conservation score, protein altering SNPs in immune-system genes (44 case samples). B) Utility of multiply affected families: 4,310 protein altering SNPs were called in both affected individuals of a 1st cousin pair from a large multiply affected family, of which 1218 mapped to shared chromosomal segments (computed using Hap300 data) and of these 64 were novel. C) Multiple rare mutations: 474 genes (34 immune system) had ≥3 distinct novel protein altering SNPs found in different individuals.

A combination of these, and other, analysis techniques will be necessary to make sense of exome sequencing data in common polygenic diseases. Coeliac disease genes of interest include SELP, ITGA2B, CD248 and IL32. Our next steps include familial segregation studies; large scale genotyping in thousands of cases and controls; and deeper resequencing.

181

Deep Resequencing Study of 202 Genes in 15,000 Individuals across 12 Diseases to Support Drug Repositioning. *M.R. Nelson[1], M.G. Ehm[1], L. Warren[1], C. Verzilli[2], J. Shen[1], D. Fraser[1], J. Aponte[1], J. Novembre[3], D. Wegmann[3], J. Li[4], S. Zöllner[4], Y. Li[5], P. St. Jean[1], L. Li[1], P. Woollard[6], S. Topp[6], M. Hall[6], K. Nangle[1], G. Abecasis[7], L.R. Cardon[8], J.C. Whittaker[2], S.L. Chissoe[1], V. Mooser[8].* 1) Genetics, GlaxoSmithKline, Res Triangle Park, NC; 2) Genetics, GlaxoSmithKline, Harlow, UK; 3) Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA; 4) Human Genetics, University of Michigan, Ann Arbor, MI; 5) Biostatistics, University of North Carolina, Chapel Hill, NC; 6) Computational Biology, GlaxoSmithKline, Stevenage, UK; 7) Biostatistics, University of Michigan, Ann Arbor, MI; 8) Genetics, GlaxoSmithKline, Upper Merion, PA.

The discovery and development of novel drugs is immensely challenging despite great unmet medical needs. One way to improve productivity is through drug repositioning — identifying new indications for drugs already marketed or in development. Large-scale DNA resequencing provides an opportunity to identify protein function-disrupting variants with a strong influence on disease risk or disease-related traits, and thus to suggest novel drug indications. We approached this by resequencing the exons of 202 genes coding drug targets in approximately 15,000 individuals. The sample design includes a mixture of case-control covering 12 diseases in five therapeutic areas and population samples with 90% of European origin and the remainder of Indian Asian, Chinese or African American ancestry. Single-molecule Illumina sequencing of captured DNA was carried out in pools of 48 tagged DNA samples. Over 98% of the 985kb of target bases were successfully sequenced. Sequencing yielded an average depth over 30x and approximately 85% of target bases covered at a depth of at least 10x. Comparisons to Sanger and external single-molecule sequence data, chip-based genotypes and internal controls showed the resequence-based genotyping accuracy to be high. We found over 10,000 protein-modifying variants of which 95% have allele frequencies less than 0.1%. The impact of variants on protein function was predicted by two methods. Preliminary genotype-phenotype analyses have identified several notable associations, including the statistically significant influence of both common ($p < 10^{-6}$) and rare ($p < 0.001$) coding variants within *CASR* on serum calcium levels and several novel *PLA2G7* variants that we have shown ablate Lp-PLA2 enzyme activity. We will present the overall organization of variation in such a deeply sequenced sample, the patterns of association observed across analyzed traits and key results with the repositioning opportunities to impact the development of drugs that act on these targets.

182

Deep Resequencing Study of the Influence of 202 Drug Target Genes on Metabolic Syndrome. *L. Warren[1], M. Nelson[1], L. Li[1], J. Shen[1], C. Verzilli[3], S. Bacanu[1], K. Song[2], D. Fraser[1], K. King[1], K. Nangle[1], J. Aponte[1], P. St.Jean[1], P. Woollard[4], M. Hall[4], L. Cardon[2], S. Chissoe[1], V. Mooser[2], M. Ehm[1], J. Whittaker[3], D. Waterworth[2].* 1) Statsitical Genetics, GlaxoSmithKline, Res Triangle Park, NC; 2) GlaxoSmithKline, Upper Merion, PA, USA; 3) GlaxoSmithKline, Harlow, UK; 4) GlaxoSmithKline, Stevenage, UK.

The Genetic Epidemiology of Metabolic Syndrome (GEMS) study is a multinational study designed to explore the genetic basis of metabolic syndrome. We conducted an exon-based deep resequence experiment to examine 202 genes related to drug targets including those targeted by compounds currently in development for treatment of dyslipidemia, type II diabetes, and coronary artery disease. An average depth of 30x and coverage of approximately 85% of target bases covered at a depth of at least 10x was achieved. We conducted analyses in 1840 subjects. A total of 12932 variants were observed; however, 681 variants with >20% missing genotypes were excluded. Among them, 4524 (37%) fell within non-coding regions, 2118 (17%) were nonsynonymous and 1852 (15%) synonymous. Most variants were rare, including 56% singletons. Only 18% had minor allele frequencies (MAF) greater than 1%. Common variants (MAF > 1%) were analyzed individually whilst rare nonsynonymous, nonsense and read-through variants were aggregated within each gene. Additional analysis of rare variants included aggregation based on predictions of functionality via either PolyPhen or SIFT. Logistic regression was applied to analysis of binary traits, including dyslipidemia status and coronary artery disease status under an additive model. Continuous traits representing a clustering of metabolic traits such as BMI, HDL-C, triglycerides, LDL-C, adiponectin, inflammatory markers, systolic and diastolic blood pressure were analyzed via linear regression. Preliminary analyses revealed several interesting associations. Reassuringly we replicated several previously know associations, including association between ADIPOQ and adiponectin (smallest P=1.2E-9). Rare nonsynonymous variants in aggregation in several genes also yielded modest association with continuous traits. We will present detailed association analysis results and discuss the challenges and opportunities in working with deep resequence data.

183

Design considerations for Massively Parallel Sequencing Studies of Complex Human Disease. *B.-J. Feng[1], S.V. Tavtigian[2], M. Southey[3], D.E. Goldgar[1].* 1) Dept of Dermatology, University of Utah, Salt Lake City, UT; 2) Huntsman Cancer Institute, University of Utah, Salt Lake City, UT; 3) Dept of Pathology, University of Melbourne, Victoria, Australia.

Massively Parallel Sequencing (MPS) provides order of magnitude improvement in throughput over Sanger sequencing. Sequencing of entire transcriptomes (and genomes) now is feasible at reasonable cost. Recently, Ng et al. (2010) elegantly demonstrated the utility of exomic massively parallel sequencing in the identification of a gene for a rare Mendelian disorder based on analysis of only four unrelated individuals. However, for a complex disease we need to deal with likely substantial degrees of locus, allelic, and phenotypic heterogeneity, as well as complex relationships between genotype and phenotype. This will require careful selection of families to be sequenced and a well-developed strategy for finding the few true susceptibility genes from the many genes that will have rare genetic variants that could plausibly alter protein function. To examine these issues we have performed analytical and simulation-based calculations in order to compare several strategies for MPS sequencing in complex disease. Factors examined include underlying genetic architecture, sample size, individuals sequenced per pedigree (1 vs. 2), and a variety of filters based on bioinformatics, observation in multiple pedigrees, and concordance within pedigrees A two-stage design is assumed where genes selected in the MPS analysis of high-risk families are evaluated in a secondary screening phase of a larger set of probands with more modest family history of disease. Assumptions for these analyses include: 25000 genes in the genome; 400 rare missense variants (rms) and 20 truncating/splice junction (TSJ) non-disease related variants per exome (Ng et al.; Nature 2009); sensitivity of 90% for detecting any given sequence variant and that pathogenic mutations in each true gene are 50% rms & 50% TSJ. Designs are evaluated assuming a two-stage design and that the cost of sequencing the whole exome is 400 times that of sequencing a single candidate gene. Results indicate that: (a) requiring variants to be identified in multiple pedigrees is an effective strategy for reducing false positives; (b) under some genetic models, no design with a single individual sequenced per family finds sufficient loci at feasible cost; and (c) the best strategy depends on the underlying genetic architecture, but designs requiring pedigree concordance and observation of a variant in a given gene in at least one other family are optimal across a wide range of genetic models and sample sizes.

184

Replication Strategies and Rare Variants Discoveries in Genetic Studies of Complex Traits using Next Generation Sequencing Technologies. *S.M. Leal[1,2], D.J. Liu[1,2].* 1) Prof, Molec/Human Gen, Baylor Col Med, Houston, TX; 2) Dept. Statistics, Rice Univ, Houston, TX.

There is solid evidence that complex traits can be due to rare variants. Next-generation sequencing is being employed to generate whole-exome sequence data to investigate the role of rare variants in complex trait etiology. In order to aggregate signal and improve power, multiple rare variants in a gene region are usually jointly analyzed using gene based tests, such as combined multivariate and collapsing (CMC) or the weight sum statistic (WSS). When a significant gene-phenotype association is detected in the exploratory sample (stage 1) these findings need to be replicated in an independent sample (stage 2). Two different replication strategies can be used a.) variant-based replication: nucleotide sites uncovered in the stage 1 are genotyped and analyzed in stage 2 or b.) gene-based replication: the gene region implicated in stage 1 is sequenced in the stage 2 sample and both known and novel variants are analyzed. The efficiency of the two strategies is dependent on the proportions of causative variants discovered in stage 1, as well as the relative error rates of sequencing and genotyping. Using a rigorous population genetic framework, we demonstrate that gene-based replication is consistently more powerful. For smaller-scale studies with a few hundred individuals usually <50% of the causal nucleotide sites are uncovered, but often >80% of the locus population-attributable-risk (PAR) can be explained by these variants. In the scenario when the uncovered variants account for a high proportion of locus PAR, the power advantage of gene based replication is small. For large scale studies with thousands of individuals, a fairly large proportion of causative variants sites can be observed in stage 1. Therefore, customized genotyping can be a temporal solution for replicating large scale genetic studies if an ethnically well-matched stage 2 sample is analyzed. However, sequencing the stage 2 sample is necessary and advantageous if initial study sample is small or the discovery of novel rare variants is also of interest. Using a novel error model for downstream analysis with parameters estimated from next-generation sequencing data, we show that the impact of current error levels for sequencing and genotyping on the power for mapping rare variants is only minimal. Compared with CMC, WSS is less robust to the errors, and suffers greater power loss. However, the advantage of gene-based replication remains in the presence of imperfect data using both tests.

185

Accounting for Imputed Data in Rare Variant Association Studies. *J.L. Asimit[1], A.P. Morris[2], A. Day-Williams[1], R. Mägi[2], E. Zeggini[1].* 1) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom.

Genotype misspecification results in a loss of power in the analysis of rare variants and disease association. Imputation of rare variants is now possible, since the 1000 Genomes (1KG) reference set provides a catalogue of variants with frequencies as low as 0.01. A powerful approach for analyses is to test for an association with accumulation of rare variants within a locus of interest, by collapsing the genotypes across variants within a locus. We extend this method to make use of the probability distributions resulting from imputation at each variant by considering the expected genotype score, rather than the genotype score with the highest probability. In simulations the power of the collapsing method for imputed data is compared with applications to data from (i)GWAS (ii)genotyping variants present in the reference panel (RP) (iii)direct re-sequencing. Some results are provided in a table for RP sizes of 120 (comparable to CEU from 1KG), 500 (comparable to Europeans from 1KG), and 1000. The quantitative trait simulations are based on 100 replicates of 2000 individuals and the RP, where the maximum MAF of a causal SNP is 0.01 and the maximum total MAF of the causal SNPs is 0.05.

RP	GWAS	Imputation	Genotyping	Re-sequencing
120	.05	.19	.86	.96
500	.03	.28	.95	.98
1000	.03	.37	.96	.98

These results suggest that imputation from a RP similar to 1KG offers gains in power over analysis of rare variants on GWAS chips, but there is still a substantial loss in power compared with direct genotyping of rare variants discovered in the RP.

186

Impact of haplotype reference panel on imputation accuracy of rare variants in isolated populations. *C.V. Van Hout, J.A. Douglas.* Department of Human Genetics, University of Michigan, Ann Arbor, MI.

A substantial fraction of the genetic component of complex traits remains unexplained by the results of recent genome-wide association analyses of common SNPs, the so-called missing heritability problem. Thus, the contribution of rare variation to heritability is currently of great interest. Efforts such as the 1000 Genomes Project are underway, in part, to catalog rare variants by deeply sequencing reference population panels (1000genomes.org). For common SNPs (minor allele frequency ≥5%), we recently showed that the Old Order Amish (OOA) of Lancaster County, PA, an isolated population derived from a modest number of founders, and the HapMap CEU participants share similar allele frequencies and linkage disequilibrium profiles (Van Hout et al., 2010). Accordingly, we expect that reference panels like the CEU will adequately characterize common SNPs in the OOA. However, for rare SNPs, we expect the OOA and CEU to differ considerably. Thus, in anticipation of the 1000 Genomes Project data, we assessed the portability of deep sequencing resources by comparing imputation accuracy of rare SNPs between haplotype reference panels representative of CEU participants and the OOA population. Using coalescent theory as implemented in Hudson's ms (Hudson, 2002), we simulated 100 megabases of sequence representative of the OOA and CEU, including 800 OOA and 800 CEU haplotypes (to serve as reference panels) and 400 OOA and 400 CEU pseudo-study participants. We masked ~95% of the study participants' genotypes and used the remaining 5% to impute the masked data from each reference panel using a hidden Markov model as implemented in MACH (Li et al., 2006 & 2009). We define imputation accuracy as the proportion of heterozygous genotypes that are imputed correctly. Consistent with expectation, using the CEU haplotype reference panel, we observed high imputation accuracy (~96%) for SNPs with MAF >5% in the OOA. However, for SNPs with MAF 1-5% in the OOA, imputation accuracy was 68% and 50% using the OOA and CEU reference panels, respectively. By comparison, based on the CEU haplotype reference panel and SNPs with MAF 1-5%, imputation accuracy was 60% for the CEU study participants. Our results suggest that public reference panels similar to those included in the 1000 Genomes Project may poorly catalog rare SNPs in isolated populations like the OOA and that there may be added value in deep genome-wide sequencing in these populations.

187

Reliable eQTL Mapping with F1 Generations of Inbred Mouse by Measuring Allele Specific Differential Expressions. *E. Kang[1], B. Han[1], L. Martin[2], A. Lusis[2], S. Shifman[3], E. Eskin[1,2].* 1) Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, USA; 2) Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA, USA; 3) Department of Genetics, The Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel.

Understanding the effect of causal genetic variation on gene expression in model organism such as mouse is an important problem in genetics. One traditional approach to identify regulatory polymorphisms is to apply statistical hypothesis testing which predicts the association between genetic polymorphisms and gene expression variation. Many previous studies have shown successful results with the genome wide association study approach (GWAS). However the main limitation of the GWAS approach is that due to the linkage disequilibrium (LD) structure among SNPs, discovering actual causal SNP on gene expression variation is difficult. In other words, GWAS analysis can lead to discovery of many correlated SNPs in a LD region. Recent studies show that allele specific differential expression (ASDE) is a widespread phenomenon affecting the gene expression and employing differential allele expression is promising for identifying causal SNPs. In this work, we perform cis-eQTL mapping for mouse between genetic polymorphisms and gene expression levels by measuring allele specific differential expression of F1 generation of inbred mouse. To this end, we examine 6 different F1 generation mice whose parental strains are each pair from 4 different inbred strains whose whole genome sequencing data is available. We measure the relative gene expression of the two copies of each gene from all 6 F1 mice. If we found the differential gene expression between two copies of each gene, there are likely to be heterozygous SNP sites responsible for the differential expression. For each gene, we look at cis-region of genome and use a combinatorial search algorithm that exhaustively searches the candidate genetic variations that are responsible for the differential expression. The advantage of this direct comparison between two copies for one gene is that we can exclude the trans-acting SNP effects on differential gene expression from our analysis. The results from our simulation study show that our approach on average reduces the number of candidate variations by 91.2%. This amount of reduction is expected to increase as we add more F1 mice in our future study. To our knowledge, this is the first differential expression-based mapping study using the whole genome sequencing data. We also propose a design methodology that guides which additional F1 generations from which parental strains to be used to maximize the efficacy of mapping.

188

Genetics of gene expression using fully sequenced human genomes. *S.B. Montgomery, T. Lappalainen, E.T. Dermitzakis.* Genetic Medicine and Development, University of Geneva, Geneva, Switzerland.

Our understanding of common and complex disease is being enhanced by our ability to uncover the effects of genetic variation on cellular state. Specifically, by understanding which variants have impact on the expression of genes, it is likely that we can also find those variants which inform important human conditions. Now, with the availability of complete genomes we are confronted with the challenge of dissecting rare and common as well as small and large variants into a more complete model of association with the aim of pinpointing specific causal variants. To approach this, we analyzed 60 complete genomes from the 1000 Genomes Project with respect to gene expression assayed by RNA-Seq (Montgomery et al., Nature 2010 464(7289):773-7). We have assessed association of 6.5 million common genetic variants (6,554,051 SNPs and 561,844 indels) with the expression of 22,194 GENCODE annotated protein coding genes. We have compared the relative effects of SNPs versus indels on gene expression. Furthermore, to integrate rarer variants, we have also developed a novel approach which exploits allelic association within individuals to uncover genome annotation that is harbouring causal variants. By using this approach to understand the potential for a noncoding variant to be functional, we have been able to hone in on likely causal variants and are closing in on being able to identify rare regulatory variants in single genomes. We hypothesize that this understanding will significantly aid in dissecting the relative impacts of genetic variants in case-control genetics of gene expression studies.

189

Next generation sequencing-based mRNA profiling of total blood in a large human cohort. *P.A.C. 't Hoen[1], J.T. den Dunnen[1], E.J.C. de Geus[2], D.I. Boomsma[2], J.J. Hottenga[2], B.W.J.H. Penninx[3], G.J.B. van Ommen[1].* 1) Center for Human and Clinical , Leiden University Medical Center, Leiden, Netherlands; 2) Department of Biological psychology, Netherlands Twin Registry, VU University, Amsterdam, Netherlands; 3) Department of Psychiatry, Netherlands Study of Depression and Anxiety, VU University Medical Center, Amsterdam, Netherlands.

With rapidly decreasing sequencing cost, sequence-based gene expression profiling becomes an attractive alternative over array-based studies. We report on one of the first sequence-based studies into the inter-individual variability of gene expression levels in total blood (n=104). Despite the high abundance of reticulocyte-derived hemoglobin mRNAs (20-80% of reads), the sequencing depth of 10±2.5 million reads per sample allows for the reliable quantification of mRNAs derived from »12,000 genes with an expression level down to 0.3 copies per cell. The amount of hemoglobin transcripts shows a significant inverse correlation with white blood cell counts at the time of sample collection. The absolute nature of the expression levels obtained with next generation sequencing, the high sensitivity of the technology, and the presence of cell type-specific transcripts allow for the accurate estimation of the relative amounts of white blood cells, including those for low abundant basophils and eosinophils. Since differences in blood cell content are a major confounding factor in blood-based expression profiling studies, it is essential to correct for these differences before analyzing expression differences between subjects. Unlike array-based studies, sequence-based studies enable the quantification of allele-specific expression using variants in the mRNA-derived sequence reads. We found that the majority of genes demonstrate preferred expression of one of the two alleles. Furthermore, we observed remarkable inter-individual differences in the preference of one allele over the other. Another factor contributing to the inter-individual differences in gene expression is the preferred expression of specific splicing isoforms and/or use of shorter or longer 3'-UTRs.

190

Comparison of cis eQTL across different brain regions. *V. Strumba[1], E.M. Schmidt[2], B.J. Keller[3], M. Flickinger[4], J.Z. Li[5], M.H. Dai[2], F. Meng[2], R.C. Thompson[2], E. Sliwerska[2], A.F. Schatzberg[6], J.D. Barchas[7], E.G. Jones[8], W.E. Bunney[9], R.M. Myers[10], S.J. Watson[2], H. Akil[2], M. Boehnke[4], L.J. Scott[4], M. Burmeister[2].* 1) Addiction Research Center, University of Michigan, Ann Arbor, MI; 2) University of Michigan, Ann Arbor, MI, Molecular and Behavioral Neuroscience Institute; 3) Eastern Michigan University, Ypsilanti, MI, Department of Computer Science; 4) University of Michigan, Ann Arbor, MI, Department of Biostatistics and Center for Statistical Genetics; 5) University of Michigan, Ann Arbor, MI, Department of Human Genetics; 6) Stanford University, Stanford, CA, Department. of Psychiatry; 7) Cornell University, Ithaca, NY, Department. of Psychiatry; 8) UC Davis, Davis, CA, Center for Neuroscience; 9) UC Irvine, Irvine, CA, Department of Psychiatry & Human Behavior; 10) HudsonAlpha Institute for Biotechnology, Huntsville, AL.

Mapping of expression quantitative trait loci (eQTL) can help improve our understanding of functionality of disease-causing genetic variants. The majority of human eQTL studies carried out to date have used transformed lymphoblastoid cell lines (LCLs) to measure mRNA expression. Other studies have used disease-relevant tissues such as liver, adipose tissue, or brain cortex. For studies of brain disorders, sampling of total brain cortex can identify eQTLs that have the same direction of effect in multiple brain regions, but may miss eQTLs that are localized to specific brain regions. To better study the brain's structural and functional complexity, we carried out eQTL mapping in several different brain regions: anterior cingulate, amygdala, dorsolateral prefrontal cortex, cerebellum, hippocampus, nuceus accumbens and anterior thalamus. We used the Illumina HumanRef-8 platform to measure expression of 18,187 probes in human postmortem brain samples from 60 individuals with bipolar, major depression and schizophrenia disorders as well as 42 controls. Genotypes for 550K SNPs were obtained using Illumina Human610-Quad chips with an additional 2 million SNPs imputed using HAPMAP data. 2.2 million SNPs were defined as putative cis-acting regulatory elements by pairing each Illumina probe with all SNPs within or near (500 KB on either side) the gene to which the probe mapped. We used an additive regression model with age, gender and disease status as covariates, to test association of genetic variation with gene expression, which resulted in 14.7 million tests. Overall, 6623 SNP-probe associations in 261 genes passed genome-wide multiple testing correction of p-value < $3.7*10^{-9}$ in at least one brain region tested. While many associations are significant across the majority of the brain regions tested and often confirm previous findings, others are brain-region specific. Brain-region specific eQTLs can be due to brain region-specific expression - i.e. when expression is only detectable in some brain regions - but our approach will also identify regulatory elements acting in only one or more but not all brain regions. Our studies will help prioritize SNPs in GWAS studies, and help identify new regulatory circuits in the most complex organ of the body.

191

Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. *A.L. Price[1], A. Helgason[2], G. Thorleifsson[2], S.A. McCarroll[3], A. Kong[2], K. Stefansson[2].* 1) Harvard School of Public Health, Boston, MA; 2) deCODE Genetics, Reykjavik, Iceland; 3) Harvard Medical School, Boston, MA.

Family studies of individual tissues have shown that gene expression traits are genetically heritable. However, it is currently unknown (1) whether heritability can be partitioned into *cis* and *trans* components using local and genome-wide identity-by-descent (IBD), (2) whether heritable components of variance are shared across tissues, and (3) whether heritability extends to distantly related ("unrelated") individuals inheriting IBD segments from distant ancestors. We applied variance-components methods to 722 Icelanders from family cohorts for which long-range phased genome-wide SNP data and gene expression measurements from blood and/or adipose tissue were available, and obtained results for each of ~19,000 genes. Our main results are: (1) We estimate the proportion of gene expression heritability attributable to *cis* regulation as 37% in blood and 24% in adipose. (2) Correlation in gene expression measurements across tissues is entirely attributable to heritable effects and entirely attributable to heritability specifically at the *cis* locus, whereas *trans* regulation is not shared across tissues. This implies that heritability in heterogeneous tissue types is expected to include a larger contribution from *cis* regulation than in more homogeneous tissue types, consistent with our blood vs. adipose results as well as results of previous studies in lymphoblastoid cell lines. (3) Analysis of IBD in unrelated individuals produces similar estimates of the *cis* components of heritability as analyses in related individuals. This suggests that transgenerational epigenetic inheritance does not contribute substantially to the "missing heritability" of gene expression in these tissue types.

192

Gender specificity of regulatory variation sculpts sexually dimorphic phenotypes in humans. *A.S. Dimas[1, 2], S.B. Montgomery[1], B.E. Stranger[3], E.T. Dermitzakis[1].* 1) Department of Genetic Medicine, University of Geneva Medical School, Geneva, Switzerland; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, United Kingdom; 3) Division of Genetics/Department of Medicine, Harvard Medical School/Brigham and Women's Hospital, 77 Avenue Louis Pasteur, Boston, MA, 02115.

Regulation of gene expression has been shown to sculpt sexually dimorphic phenotypes in multiple model organisms. The impact of regulatory variation on human sexual dimorphism, including gender-specific differences in the prevalence and severity disease, has not been explored to date. To address this, we measured gene expression in lymphoblastoid cell lines (LCLs) from individuals from four HapMap populations (CEU: 54 females [F], 55 males [M], CHB 42 F, 38 M, JPT 40 F, 42 M, YRI: 53 F, 55 M) and tested for association with SNP genotypes for each gender separately. Of the 303 and 279 expression-associated genes detected in CEU F and CEU M respectively, 193 (64%) and 169 (61%) were found exclusively in one gender. Similarly, from the 354 and 330 genes detected in YRI F and YRI M, 262 (74%) and 238 (72%) were gender-specific. Equivalent numbers were detected for CHB and JPT (58% F, 67% M in CHB; 64% F, 67% M in JPT). Expression mean and median values for genes with gender-specific eQTLs were similar across sexes, although expression variance was consistently, but not significantly, higher in M. Furthermore, gender-specific genes did not show indication of significance (low p-value enrichment) in the other gender, suggesting prominent gender specificity of regulatory variation. GO terms analyses revealed an over-representation of genes involved in biological processes relevant to sexually dimorphic phenotypes (e.g. spermatocyte formation, oocyte growth, steroid metabolism, muscle development). The direction of allelic effects was consistent across sexes, with effect sizes for gender-specific eQTLs being smaller than those of shared eQTLs. Many genes with significant associations in both sexes were found to possess distinct, independent, gender-specific eQTLs, hinting at the existence of sex-specific regulatory variants. Such cases include SPO11, (spermatocyte formation; disruption in mouse leads to F and M infertility), ODF2L (dense fibre of sperm tails), and PSAP (sphingolipid metabolism, also associated with schizophrenia and prostate cancer). Multiple associations were also detected for genes involved in sex steroid metabolism, supporting the idea that sex hormones play a major role in immune system regulation. Documenting gender-specificity of regulatory variation in LCLs may therefore provide an important link between the observed sex-specific differences in the function of the immune system and sexual dimorphism in the prevalence of disease.

193

eQTLs and eSNPs and their relationship to prevalence of Coronary Artery Disease and cardiovascular risk factors. *T. Zeller[1], S. Szymczak[2], A. Schillert[2], P. Wild[1], M. Rotival[3], F. Cambien[3], L. Tiret[3], L. Lackner[4], T.F. Munzel[1], A. Ziegler[2], S. Blankenberg[1].* 1) Dept Med II, Univ Med Center Mainz, Mainz, Germany; 2) Institute of Medical Biometry and Statistics, Medical University Schleswig-Holstein, University at Lübeck, Lübeck, Germany; 3) INSERM UMRS 937, Pierre and Marie Curie University (UPMC) and Medical School, Paris, France; 4) Institute of Clinical Chemistry and Laboratory Medicine, University Medical Center Mainz, Johannes Gutenberg University Mainz, Germany.

Identifying genes that influence common disease is an important effort undertaken in the field of genetics research. Genome-wide association studies led to the identification of several loci associated with the risk of coronary artery disease (CAD). However, the association signals have yet to be translated into a more complete understanding of the genetic elements mediating CAD susceptibility. To gain more insight into the genetics of CAD we aimed to investigated eQTLs and eSNPs associated with CAD and their relation to prevalence of CAD and intermediate cardiovascular risk factors. In the CADomics study (2,000 cases/3,000 controls) a SNP-eQTL analysis was performed to identify eQTLs and eSNPs associated with CAD. Identified eQTLs and the respective eSNPs were further investigated for relationship with prevalence of CAD and cardiovascular risk factors. 34 CAD-associated SNPs showed association to expression of 16 different transcripts ($p<10^{-8}$). The most interesting genes were *LYZ* (Lysozyme), *PHGDH* (3-phosphoglycerate dehydrogenase), *LIPA* (Lipase A), and *MTAP* (methylthioadenosine phosphorylase). *LYZ* expression revealed the strongest difference between cases and controls with reduced transcript levels in CAD individuals ($P=5.4x10^{-6}$) and showed correlations with blood pressure ($P=8.5x10^{-3}$). The corresponding eSNP was associated with higher HDL-cholesterol levels ($P=2.5x10^{-2}$). *PHGDH* expression was similar in CAD cases and controls, however, strong associations were observed between *PHGDH* expression and multiple cardiovascular risk factors. The corresponding eSNP was associated with lower LDL-cholesterol levels ($P=2.9x10^{-2}$) and lower diastolic blood pressure ($P=3.8x10^{-2}$). *LIPA* expression showed no significant difference between CAD cases and controls but was significantly associated with HDL-cholesterol levels ($P=2.5x10^{-3}$) and endothelial function ($P=4.04x10^{-3}$). *MTAP* expression was significantly lower in CAD cases compared to controls ($P=2.5x10^{-2}$), but no significant association was observed for *MTAP* expression and any cardiovascular risk factor. In contrast, corresponding eSNPs were associated with blood pressure ($P=1.1x10^{-2}$), LDL-cholesterol levels ($P = 4.2x10^{-2}$) and endothelial function ($P=4.4x10^{-2}$). This study demonstrates that a combined SNP-eQTL analysis and further investigation of relationship of eSNPs and eQTLs to cardiovascular risk factors will lead to a more complete biological picture of genetics of common diseases.

194

Transcriptional Regulation of Selenoprotein S and its Potential Role in Cardiovascular Disease Risk. *J.E. Curran[1], M.P. Johnson[1], E.I. Drigalenko[1], M.A. Carless[1], T.D. Dyer[1], H.H.H. Goring[1], M.C. Mahaney[1], A.G. Comuzzie[1], K.R. Walder[2], L.J. Abraham[3], J.B. Jowett[4], E.K. Moses[1], J. Blangero[1].* 1) Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX; 2) Metabolic Research Unit, Deakin University, Geelong, Australia; 3) Biochemistry and Molecular Biology, University of Western Australia, Crawley, Australia; 4) Genomics and Systems Biology, Baker IDI Heart and Diabetes Institute, Melbourne, Australia.

Inflammation plays a major role in and precedes the onset of many common complex diseases including CVD, diabetes and cancer. We have previously identified *SELS* as a significant mediator of the inflammatory response. *SELS* plays a crucial role in protecting the functional integrity of the endoplasmic reticulum (ER) against stress activators. When the ER protection mechanisms fail, the accumulation of stress ultimately leads to activation of the inflammatory response. For the first time, we have directly tested whether the functional G-105A *SELS* promoter variant is associated with altered gene expression *in vivo*. We typed this SNP in 1,100 Mexican Americans from large extended pedigrees and tested for association with *SELS* gene expression (from genome-wide transcriptional profiling of lymphocytes) using the measured genotype approach. As predicted from our prior functional work, the A allele is associated with decreased gene expression (p= 0.016). We have previously observed *in vitro* that the difference between genotypes was increased during challenge with stress agents. Therefore, we created a *SELS* response residual phenotype by controlling for overall ER stress by regression on transcript levels for several ER stress marking genes such as *HSP90B1*, *HERPUD1*, *HSPA5*, and *ATF6*. Conditional upon these ER-related gene transcript levels, *SELS* expression response to ER stress strongly negatively correlated with measures of ER stress-associated damage such as *DDIT3* ($p=9.2\times10^{-9}$) and *NFKB1* ($p=1.1\times10^{-4}$). These statistical results strongly support a major role for *SELS* expression attenuating ER stress damage. When we then tested for the effect of the G-105A promoter variant on this *SELS* response phenotype, the evidence for association dramatically improved ($p=3.0\times10^{-4}$) as predicted from the *in vitro* studies. These results further support a major role for *SELS* in the inflammatory response. We also performed bivariate quantitative genetic analyses between *SELS* expression level and several CVD risk phenotypes. In our data set, *SELS* expression levels are significantly heritable (h2=0.35, $p=1.4\times10^{-13}$). There is strong evidence for genetic correlation between *SELS* and multiple lipoprotein-related, obesity-related, hypertension-related and oxidative stress/inflammation measures. All observed genetic correlations are in the correct predicted direction, reflecting the influence of genetic variants regulating *SELS* expression and their downstream effects on CVD risk phenotypes.

195

High Proportion of Transcripts Associated with Insulin Sensitivity in Fat and Muscle Are Associated with Expression Quantitative Trait Loci (eQTL). *S.C. Elbein[1], E.R. Gamazon[2], S.K. Das[1], N. Rasouli[3], P.A. Kern[4], N.J. Cox[2].* 1) Section on Endocrinology, Dept. of Medicine, Wake Forest U Sch Med, Winston-Salem, NC; 2) Section on Genetic Medicine, Dept. of Medicine, University of Chicago, Chicago, IL; 3) Division of Endocrinology, Dept of Internal Medicine, University of Colorado School of Medicine, Aurora, CO; 4) Division of Endocrinology, Dept of Internal Medicine, University of Kentucky School of Medicine, Lexington, KY.

Impaired insulin action is an early and prevalent feature of type 2 diabetes and metabolic syndrome, and a heritable trait, yet few genetic loci have been identified that alter insulin action. To search for loci associated with insulin sensitivity, we first used Agilent 4x44 expression arrays to identify transcripts that were differentially expressed in adipose and muscle from 62 individuals selected from 184 individuals to be at extremes of insulin sensitivity after correcting for age, gender, and obesity. Using a fold change of 1.5 and q-value of <10%, we identified 43 muscle and 259 adipose transcripts. We genotyped the 62 individuals with the Illumina 1M and tested the association of each SNP with the differentially expressed transcripts to identify both local (within 1 mb of the probe) and distal eQTLs controlling differentially expressed transcripts. In muscle, 14/43 transcripts were associated with local eQTLs ($p<0.001$), including PPARGC1A (3 SNPs) and SOCS2. For distal transcripts we used a more stringent p value of 10-8 and identified 1275 SNPs associated with 33 transcripts, with transcripts associated with 1-136 SNPs. Conversely, of 346 SNPs associated with at least one muscle transcript, 73 were associated with at least 5 differentially expressed transcripts. Even at p<10-10 24/43 transcripts were associated with at least one distant eQTL, and 6 SNPs were associated with 5 or more transcripts. In adipose, 66/259 transcripts were associated with at least one local (<1 mb) eQTL, including strong candidates HK2, PCK1, PTPN3, RDH10, RXRG, TSPAN9, and VEGFA. We restricted distal transcripts to p<10-10 and identified 65/259 transcripts associated with at least 1 eQTL, and 35 transcripts associated with 5 to 105 SNPs. Conversely, of the 207 eQTLs at approximate genome-wide significance (p<10-10), 17 controlled 5-23 transcripts. Among the transcripts with distal eQTLs were ACACA, ELOVL6, FASN, UBQLN1, and several zinc finger proteins. At a less stringent p value (<10-8), 175/259 transcripts were associated with at least one eQTL, and 129 SNP were associated with 5 or more transcripts. We show that both local and distal eQTLs control transcript levels that differ between insulin resistant and insulin sensitive individuals. Many distal eQTLs appear to be master regulators controlling large numbers of transcripts of importance to insulin action. Such master regulators may play an important role in modulating insulin action in adipose and muscle.

196

Rare Sequence Variants at the X-linked *FMR1/AFF2* Contribute to Male Autism Susceptibilty. *K. Mondal, K. Hagen, A.C. Shetty, V. Patel, M.E. Zwick.* Department of Human Genetics, Emory University School of Medicine, Atlanta, GA.

Autism spectrum disorders (ASDs) exhibit a high heritability and affect four times as many males as females. To test the hypothesis that X-linked rare variants contribute to this pattern, we comprehensively sequenced the *FMR1* and *AFF2* genomic region in 127 male ASD probands from the Autism Genetic Resource Exchange (AGRE) collection. Probands were selected from multiplex families with two or more affected males who share the Xq chromosomal arm, which includes *FMR1* and *AFF2*. A total of 607 variants were observed, including 449 single base variants (SBVs) and 158 insertions/deletions (Indels). While the average level of variation (theta = 0.00059) agreed with the chr X expectation, an excess of rare variants was observed (Tajima's D = -1.44). Of 449 SBVs, 2 are exonic replacements in *AFF2*, 6 are exonic silent variants, 18 are in the untranslated regions (UTRs) and 423 are intronic. All of the 158 indels lie in introns. The *AFF2* missense mutations (D714N, R927H) each affect a single proband and occur at highly conserved sites predicted to be highly damaging by Panther and SIFT. A third *AFF2* misssense mutation (R837C) that is also predicted to be deleterious was identified in an ASD proband from the Simons Simplex Collection. A recent study that sequenced the chr X exome of 208 individuals with X-linked intellectual disability (Tarpey et al, Nat Gen, 41:535-43, 2009) identified two additional *AFF2* missense mutations (H612D, P886A). In total, these data suggest that up to 1.5% (5/339) of male developmental delay (ASD or intellectual disability) could result from coding sequence mutations at the *AFF2* locus. Of the 18 3'UTR variants at *FMR1* and *AFF2*, 6 are found at highly conserved sites, do not exist in dbSNP, and could also act as ASD susceptibility loci. To interrogate the functional effects of these variants, we will present the results of ChIP-Seq experiments for the *AFF2* missense mutations and luciferase assays for the UTR variants. Our main finding, that rare DNA sequence in *FMR1/AFF2* genomic region may contribute to ASD susceptibility, can help explain a portion of the ASD male excess and helps elucidate the research paradigm for assessing the common disease-rare variant model for complex human disorders.

197

Germline mutations of the CBL gene define a new genetic syndrome with predisposition to juvenile myelomonocytic leukemia (JMML): "the CBL syndrome". *B. Perez[1,2], J. Lachenaud[1,2], B. Isidor[3], S. Odent[4], N. Philip[5], B. Leheup[6], C. Baumann[1], L. Burglen[7], E. Pasmant[8,9], D. Vidaud[8,9], F. Mechinaud[10], A. Salmon[11], C. Galambrun[12], M. Besnard[13], C. Dupré[14], N. Ben Romdhane[15], N. Djebara[1], B. Cassinat[2], S. Kaltenbach[1], O. Fenneteau[16], A. Baruchel[17], C. Chomienne[2], A. Verloes[1, 18], H. Cave[1,2].* 1) AP-HP, Hopital Robert Debre, Departement de Genetique, , Universite Paris 7-Diderot, Paris, France; 2) INSERM U940, IUH, Hopital Saint-Louis, Paris, France; 3) Service de Genetique, CHRU Nantes, Nantes, France; 4) Service de Genetique, CHRU Rennes, Rennes, France; 5) Service de Genetique, Hopital de la Timone, Marseille, France; 6) Service de Genetique, CHRU Nancy, Nancy, France; 7) AH-HP, Hopital Armand-Trousseau, Service de Genetique clinique, Paris, France; 8) AP-HP, Hopital Beaujon, Service de Biochimie et de Genetique Moleculaire, Clichy, France,; 9) UMR745 INSERM, Universite Paris Descartes, Faculte des Sciences Pharmaceutiques et Biologiques, Paris, France; 10) Service d'Oncologie Hematologie Pediatrique, CHRU Nantes, Nantes, France; 11) Service d'Onco-Hematologie Pediatrique, CHRU Nancy, Nancy, France,; 12) Service d'Hematologie Pediatrique, Hopital de la Timone, Marseille, France; 13) Clinique PAFOAI, Papetee Tahiti, Polynesie française; 14) Service d'Hematologie Pediatrique, CHU Rennes, Rennes, France; 15) Service d'Hematologie, Hopital La Rabta, Tunis, Tunisia; 16) AP-HP, Hopital Robert Debre, Service d'Hematologie Biologique, Paris, France; 17) AP-HP, Hopital Robert Debre, Service d'Hematologie Pediatrique; 18) INSERM U676, Hopital Robert Debre, Paris, France.

Background: CBL missense mutations have recently been associated with juvenile myelomonocytic leukemia (JMML), an aggressive myelodysplastic and myeloproliferative neoplasm of early childhood characterized by excessive macrophage/monocyte proliferation. CBL, an E3 ubiquitin ligase and a multi adaptor protein, controls proliferative signaling networks by downregulating the growth factor receptor signaling cascades in various cell types. Methods and results: CBL mutations were screened in 70 patients with JMML. A homozygous mutation of CBL was found in leukemic cells of 5 patients (7%): p.Y371H (3 cases), p.F418L and 1228-2A>G. In all cases, copy-number neutral loss of heterozygosity of the CBL locus (11q23.3) was demonstrated. CBL mutations were mutually exclusive from other RAS activating mutations found in JMML. An underlying developmental condition was suspected in these 5 patients and a heterozygous germline CBL substitution was evidenced in each of them. The germline mutation represents the first hit. Somatic loss of heterozygosity due to acquired somatic uniparental isodisomy is the second hit, positively selected in JMML cells. The CBL mutation was inherited from the parents in 2/5 cases. Patients displayed a variable combination of mild dysmorphic features, hyperpigmented skin lesions, microcephaly, and cryptorchidism. Learning difficulties and postnatal growth retardation may be part of the phenotype. Four patients received a cord blood allograft and all patients survived their leukemia. Some clinical features were reminiscent of Noonan syndrome (NS) or type 1-neurofibromatosis (NF1), two genetic conditions associated with RAS signaling pathway activation and predisposing to JMML. CBL mutations were screened in patients referred for molecular diagnosis of NF1 (n=47) or NS (n=214). Two patients with CBL mutations were identified. None of them had leukemia but one had hepato-splenomegaly. Conclusion: We report here the existence of an unreported inheritable condition due to germline CBL mutations and associated with a predisposition to JMML. The "CBL syndrome" shares some clinical similarities with Noonan syndrome, another example of syndromic JMML, and CBL could be a new gene to screen in RASopathies. Further studies will permit to refine the clinical description of patients with this syndrome and to better appreciate the risk of leukemia or other malignancies associated with it.

198

DICER1 syndrome - a pleiotropic tumor predisposition syndrome. *I. Slade[1], C. Bacchelli[1], H. Davies[2], A. Murray[1], F. Abbaszadeh[1], S. Hanks[1], R. Barfoot[1], . The FACT Collaboration[1], C. Stiller[3], A.G. Nicholson[4], N. Sebire[5], J.R. Priest[6], K. Pritchard-Jones[7], M.R. Stratton[1,2], J. Douglas[1], N. Rahman[1].* 1) Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey, United Kingdom; 2) The Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 3) Childhood Cancer Research Group, Department of Pediatrics, University of Oxford, Oxford, United Kingdom; 4) Department of Histopathology, Royal Brompton Hospital and the National Heart and Lung Institute, Imperial College, London, United Kingdom; 5) Department of Histopathology and Pediatric Laboratory Medicine, Great Ormond Street Hospital, London, United Kingdom; 6) International PPB Registry, Children's Hospitals and Clinics of Minnesota, MN; 7) Section of Pediatrics, Institute of Cancer Research, Sutton, Surrey, United Kingdom.

Dicer1, an RNase endonuclease type III, is an essential component in the production of microRNAs. Constitutional *DICER1* mutations were recently reported to cause familial pleuropulmonary blastoma (PPB), a rare embryonal malignancy of the lung occurring in childhood. We sequenced *DICER1* in constitutional DNA from 823 unrelated individuals with a variety of tumors and in 781 cancer cell lines, to investigate the contribution and phenotypic spectrum of constitutional and somatic *DICER1* mutations to cancer. We identified constitutional *DICER1* mutations in 19/823 individuals including 11/14 with sporadic pleuropulmonary blastoma, 2/3 with cystic nephroma, 3/6 with ovarian Sertoli-Leydig tumors, 1/243 with Wilms tumor (this individual also had Sertoli-Leydig tumor), 1/1 with intra-ocular medulloepithelioma (this individual also had PPB), 1/86 with medulloblastoma and 1/172 with germ cell tumor. In the 17 families in which inheritance could be investigated, *DICER1* mutations were identified in 25 relatives, 23 of whom did not have a tumor diagnosis, one mother had ovarian Sertoli-Leydig tumor and one half-sibling had cystic nephroma. Six relatives had non-toxic thyroid cysts/goiter. Analysis of eight tumors from *DICER1* mutation-positive individuals demonstrated retention of the wild-type allele in all tumors. We identified *DICER1* truncating mutations in only 4/781 cancer cell lines. All were microsatellite unstable cell lines and hence these are unlikely to be driver mutations. These data demonstrate that constitutional *DICER1* haploinsufficiency predisposes to a broad range of tumors, making a substantial contribution to PPB, cystic nephroma and ovarian Sertoli-Leydig tumors, but only a small contribution to other tumors. Most mutation carriers are unaffected indicating that the risk of tumors is at best modest. We have termed this condition '*DICER1* syndrome'.

199

Genetic variants in DNA repair genes and the risk of cutaneous malignant melanoma in melanoma-prone families with/without CDKN2A mutations. *X. Liang, RM. Pfeiffer, D. Maeder, L. Burdette, M. Yeager, S. Chanock, MA. Tucker, AM. Goldstein, XR. Yang.* Division of Cancer Epidemiology and Genetics, NCI/NIH, Bethesda, MD, USA.

Cutaneous malignant melanoma (CMM) is an etiologically heterogeneous disease with genetic, environmental (sun exposure), host (pigmentation/nevi) factors, and their interactions contributing to risk. CDKN2A is a major susceptibility gene for familial CMM. However, its incomplete penetrance suggests that other factors such as variants in the pigmentation gene, melanocortin-1 receptor (MC1R) may modify its effect. Other common genetic variants in low-penetrance genes in a number of biologically important pathways, such as cell cycle, DNA repair, apoptosis, and detoxification of metabolites, have been reported to confer modest risk for CMM. Genetic variants in DNA repair genes may be particularly important since their altered function in response to DNA damage is directly related to sun exposure. However, studies of genetic variants in DNA repair genes are limited, and even less studies were done in high-risk families. We comprehensively analyzed DNA repair gene polymorphisms and CMM risk in melanoma-prone families with/without CDKN2A mutations where 562 individuals (183 CMM) from 53 families (23 CDKN2A+, 30 CDKN2A-) were genotyped for 3079 tag SNPs in 138 DNA repair genes. Conditional logistic regression, matching on families, was used to estimate trend p-values, odds ratios (ORs) and 95% confidence intervals (CIs) for the association between CMM and each SNP. All analyses were adjusted for age and sex. The two strongest associations were observed for rs12046289 (OR=2.28, 95% CI: 1.46-3.54, p=0.0003) and rs3789613 (OR=1.81, 95% CI: 1.30-2.52, p=0.0004), both in DCLRE1B. SNPs rs7087131 and rs11016879 in MGMT also showed suggestive associations (P=0.0004 and 0.0005, respectively). When we analyzed SNPs with P<0.001 in CDKN2A+ and CDKN2A- families separately, we found that the effects of rs12046289 in DCLRE1B and rs11016879 in MGMT were mainly driven by CDKN2A+ families. In CDKN2A+ families, ORs of rs12046289 in DCLRE1B and rs11016879 in MGMT were 2.87 (95% CI: 1.61-5.11, p=0.0004) and 2.03 (95% CI: 1.36-3.02, p=0.0005). While in CDKN2A- families, ORs of these two SNPs were 1.33 (p=0.29) and 1.47 (p=0.21), respectively. Our finding suggests that polymorphisms in DCLRE1B and MGMT may increase melanoma risk in CDKN2A+ families. Replications of our findings are needed to confirm the role of these two genes in melanoma etiology.

200

Phenotypic features and predictors of mutations among 162 patients with germline PTEN mutations. *R.T. Pilarski[1,2], J.A. Stephens[3], R. Noss[1,2], J.L. Fisher[2], T.W. Prior[4].* 1) Human Cancer Genetics Program and Division of Human Genetics; 2) Comprehensive Cancer Center - James Cancer Hospital and Solove Research Institute; 3) Center for Biostatistics; 4) Molecular Pathology Laboratory, Ohio State Univ, Columbus, OH.

Background: Cowden syndrome (CS) is a multisystem disease of benign hamartomatous lesions and increased risks for thyroid, breast and endometrial cancers. BRRS is an allelic disorder characterized by macrocephaly, intestinal polyps, lipomas and pigmented penile macules. Reportedly 80% of CS and 60% of BRRS patients have detectable germline mutations in the PTEN gene. Prior data on the component clinical features have been based on small numbers of selected patients and/or compilations of case reports prior to development of consensus diagnostic criteria. We sought to determine the clinical features in a large cohort of mutation-positive patients. Methods: Data were reviewed on 1132 patients referred for PTEN gene sequencing by a single laboratory. Clinical features present at the time of testing were ascertained using a checklist completed by the referring clinician. Results: 792 patients (530 female; 67%) had usable clinical data and were included. 162 patients (85 female; 20.4%) were found to have probable deleterious PTEN mutations. Only 34.5% of those meeting CS diagnostic criteria had a mutation, compared to 56% of BRRS patients. Among patients with a mutation, macrocephaly was twice as common as previously estimated (83% vs. ~40%). Mental retardation or developmental delay was reported for 18.3% (26% of males vs. 12% of females). Penile freckling was reported in 50% of males. For female mutation carriers over age 18, 46% had breast cancer and 17% had endometrial cancer. Thyroid cancer was reported at similar rates (13.8 % vs. 13.9%) among females and males over age 18. Logistic regression analyses indicated that female mutation carriers vs. non-carriers were best discriminated by presence of macrocephaly, endometrial cancer, trichilemmomas, papillomatous papules, benign thyroid disease and benign GI lesions. For males, the most discriminating features were macrocephaly, lipomas, papillomatous papules, penile freckling, benign GI lesions and benign thyroid disease. Discussion: This is the largest cohort of PTEN mutation-positive patients to date. The mutation detection rate in patients meeting CS diagnostic criteria was significantly lower than previously found. Differences were also noted from the previously reported frequencies of macrocephaly, endometrial cancer and other clinical features. Clinical features best discriminating between mutation carriers and non-carriers were identified and should prove useful in clinical practice.

201

Clinical investigation of the UK cohort of *FLCN* mutation positive Birt-Hogg-Dubé syndrome patients. *D.H.K. Lim[1,2], E.R. Woodward[1,2], G. Kirby[1], F. Macdonald[2], L. Izatt[3], L. Walker[4], L. Side[4], S.V. Hodgson[5], P.J. Morrison[6], E.R. Maher[1,2].* 1) Department of Medical & Molecular Genetics & Centre for Rare Diseases and Personalised Medicine, University of Birmingham College of Medical and Dental Sciences, Edgbaston, Birmingham B15 2TT, UK; 2) West Midlands Regional Genetics Service, Birmingham Women's Hospital, Edgbaston, Birmingham B15 2TG, UK; 3) Department of Clinical Genetics, Guys Hospital, London, UK; 4) Department of Clinical Genetics, Churchill Hospital, Oxford, UK; 5) Department of Medical Genetics, St. George's University of London, London, UK; 6) Department of Medical Genetics, Belfast City Hospital, Belfast, UK.

Background: Birt-Hogg-Dubé (BHD) syndrome is an autosomal dominantly inherited genodermatosis and familial cancer syndrome characterised by fibrofolliculomas (benign tumour of the hair follicles), pulmonary cysts that can predispose to pneumothorax and renal cancer. BHD is caused by mutations in the *FLCN* gene. Diagnostic testing for *FLCN* mutations by sequencing was first offered in the United Kingdom at the West Midlands Regional Genetics laboratory in 2007 and subsequently screening for deletions and duplication of *FLCN* by MLPA was offered in 2009. Since diagnostic testing has been available, a registry of *FLCN* mutation positive BHD patients in the UK has been established. The registry now contains 110 *FLCN*+ve carriers from 42 families. **Aim of Study:** To define the natural history and age related risks of the various manifestations including tumours/malignancies in a large cohort of BHD patients in the UK. **Methods:** A research study into the clinical genetic investigation of BHD syndrome in the UK involving detailed phenotyping and collecting natural history data of the cohort commenced in January 2009 and is currently ongoing. We also established an online locus-specific database (LSDB) of sequence variants (published and unpublished) in the *FLCN* gene available at www.lovd.nl/flcn. **Results:** We present preliminary data in the UK cohort of BHD patients including natural history, genotype-phenotype correlations and age related risks of malignancies. Analysis of a cohort of 80 UK BHD *FLCN*+ve mutation carriers showed that all cases have been familial with no new mutation cases currently reported. Spontaneous pneumothorax occurred in 25% of mutation carriers. 21 patients developed renal cancer with a median age of diagnosis of 54years (range 25-84).The lifetime risk of renal cancer in this cohort is 40%. 4 patients developed colorectal cancer with a median age of diagnosis of 56years (range 48-64). The lifetime risk of colorectal cancer in this cohort is 12%. Other rare manifestations that occurred include malignant melanoma (4 patients), thyroid follicular adenoma (1 patient), eccrine spiroma (1 patient) and multiple basal cell carcinoma (1 patient). **Conclusions:** A lifetime risk of renal cancer of 40% in our cohort indicates the importance of renal cancer surveillance in BHD patients. *FLCN* mutation testing should be offered to at risk relatives and also to other patients with a suspected clinical diagnosis of BHD.

202

IDH1 Mutation, a Frequent Genetic Alteration in Pediatric and Adult Gliomatosis Cerebri. *C. Godfraind, D. Narasimhaiah.* Pathology, Univ catholique de Louvain, Brussels, Belgium.

Somatic mutations in the isocitrate dehydrogenase 1 gene (IDH1) have recently been associated with glioblastoma of better prognosis and of younger patient age (Science 2008, 321, 1807). IDH1 codes for an enzyme implicated in the Krebs cycle that converts D-isocitrate to alpha-ketoglutarate. When mutated, this gene's activity is diminished and altered, as it rather transforms alpha-ketoglutarate into 2-hydroxyglutarate acid (Nature 2009, 462, 739-744). Interestingly, 2-hydroxyglutarate aciduria is a metabolic disease that has been associated with higher frequency of brain tumors including medulloblastoma and glioblastoma. So far in tested brain tumors, mutations of IDH1 and of IDH2, the second gene coding for a mitochoundrial NADP(+)-dependent isocitrate dehydrogenase, were mostly reported in adult supratentorial primitive neuro-ectodermal tumors and in diffuse adult gliomas of both astrocytic and oligodendroglial lineage. We tested a yet unscreened glioma type named gliomatosis cerebri for mutations of both genes. This was performed by sequencing tumoral DNA extracted for paraffin sections and by using the primers previously reported (JNEN, 2009, 68, 1319-1325). Gliomatosis cerebri is a rare glial tumor characterized by an extensive brain infiltration. Clinical presentation and evolution as well as histological aspects vary from one patient to another. No major genetic alteration has yet been reported. Our series included 15 gliomatosis cerebri, 13 from adult patients and 2 from children. For 10 tumors, IDH1 mutations were observed (67%). They all corresponded to R132H mutation, the most frequent mutation reported for this gene. The two pediatric tumors and eight adult ones were mutated. No IDH2 mutation was detected. Thus, IDH1 mutations appear to be a frequent genetic anomalies in gliomatosis cerebri. This observation ties, by a common and early genetic alteration, the tumorigenesis of two glioma entities previously though to be unrelated, except for their cellular origin.

203

Exome re-sequencing of seven melanoma cell lines to characterize somatic mutations. *S. Nikolaev[1], K. Harshman[2], C. Gehrig[1], M. Guipponi[1], B.J. Stevenson[3,4], A.J. Sharp[1], D. Rimoldi[3], P. Descombes[5], I. Xenarios[4], V.C. Jongeneel[3,4], A. Valsesia[3,4], Ch. Iseli[3,4], S.E. Antonarakis[1].* 1) Dept Gen Med, Development, Univ Geneva, Geneva, Switzerland; 2) Lausanne Genomic Technologies Facility, Center for Integrative Genomics, CH-1015, Lausanne, Switzerland; 3) Ludwig Institute for Cancer Research ,CH-1015, Lausanne, Switzerland; 4) Swiss Institute of Bioinformatics, CH-1015, Lausanne, Switzerland; 5) Platform, NCCR "Frontiers in Genetics", University of Geneva, Geneva, Switzerland.

Melanoma is the most dangerous form of skin cancer that can lead to local and distant metastases. Ultraviolet (UV) light exposure is a well known environmental risk factor. Although a few prevalent mutations have been identified with the sequencing of a single melanoma cell line, the full spectrum of somatic driver mutations remains unknown. To better understand the molecular pathophysiology of melanoma we have used high-throughput sequencing of array-selected exons to detect somatic mutations in protein-coding regions of seven melanoma cell lines (5 unrelated and 2 derived from the same patient). We performed exome selection using CCDS capture arrays (34 Mb, ~20'000 genes) on genomic DNA extracted from cell lines derived from metastatic tumors and from normal tissues of the same patient (EBV-transformed cell lines or Peripheral Blood Lymphocytes). An improved capture protocol resulted in 60-90% of DNA library fragments originating from coding exons. Exomes were re-sequenced using 2-3 lanes of 2x76 bp paired-end reads, providing 40-80 fold coverage per sample. The number of exonic germline variants, relative to the reference human genome sequence, was about 10'000 in each of 6 individuals and the number of somatic mutations in the 7 melanoma cell lines ranged from 156 to 980. We observed that the most abundant class of somatic mutations is C>T, a hallmark of exposure to UV light. In order to detect mutations in putative candidate genes, we compared somatic nonsense or missense mutations among individuals and found that 412 out of the 2468 affected genes occurred in more than one melanoma cell line. In 6 out of 7 melanomas we found somatic non-synonymous mutations in the BRAF gene that is frequently mutated in melanomas. We also detected missense or nonsense somatic mutations in 4 genes (STAB2, ZNF285A, DSC1 and LRP1B) in 5 out of 7 cases. Mutations in LRP1B gene were previously associated with lung adenocarcinoma. In addition, we compared two melanomas derived from a same patient but from distinct metastases with 12 years interval; 150 shared somatic variants were identified (out of 276 and 349 somatic variants in the early and the late cell line respectively). Such results will shed light in the natural history of tumor development.

204

Complete exome resequencing reveals inactivating mutations in a novel metastasis suppressor on chromosome 3 in ocular melanoma. *A. Bowcock[1], E. Roberson[1], M. Onken[2], L. Cao[1], L. Worley[2], S. Duan[1], M. Council[1], C. Helms[1], J. Harbour[2].* 1) Dept Genetics, Washington Univ, St Louis, MO; 2) Dept Ophthamology, Washington Univ, St Louis, MO.

Uveal (ocular) melanoma (UM) is the most common primary cancer of the eye and the second most common form of malignant melanoma. Tumors often result, not only in vision loss, but in death within a few months of diagnosis due to metastasis to the liver. The metastasizing form of UM exhibits a characteristic "class 2" gene expression signature, which is strongly associated with monosomy 3, suggesting the presence of a metastasis suppressor gene on chromosome 3. Although activating oncogenic mutations in GNAQ have been reported in UMs, these mutations do not occur very early in tumorigenesis and are not correlated with metastasis. Despite intensive investigations, however, no metastasis-associated mutations have been identified on chromosome 3 or elsewhere in UM. We performed exome capture and massively parallel sequencing of DNA from two class 2 tumors (MM56 and MM70) and identified a single gene on chromosome 3 with loss of function mutations. One converted a Trp codon (UGG) to a UGA termination codon and the other was an 11bp deletion leading to a frameshift and premature termination of the protein. All coding exons of this gene were then subjected to Sanger sequencing of genomic DNA from an additional five class 1 and seventeen additional class 2 tumors. Matched normal DNA from the same patients was also sequenced. Inactivating mutations were identified in none of the class 1 tumors and in fourteen class 2 tumors. These included 8 premature truncations, five mutations that are predicted to disrupt the active site of this gene, and one read-through mutation that abolished its normal termination codon. Lack of corresponding mutations in normal DNA samples indicated that all changes were somatic in origin. Quantitative RT-PCR confirmed that this gene is expressed in normal uveal melanocytes and UMs, and that expression was significantly higher in melanocytes and class 1 UM compared to class 2 UM ($P = 0.0003$). Plausible cancer causing mutations were not detected in any other gene on chromosome 3 suggesting that loss of function of this gene leads to metastasis. Its identification as a metastasis suppressor for this particularly lethal form of melanoma has important diagnostic and therapeutic ramifications.

205

Genetic Analysis of Five Cancers and Cancer Progression Using Single-Cell Sequencing. *J. Wang, X. Xu, RQ. Li, X. Yang.* Beijing Genomics Institute at Shenzhen, Shenzhen 518083, China.

Tumor tissues contain several different cell subpopulations, which complicate the detection of important genetic changes relevant to cancer development. So far, deep sequencing of tumor tissues or of cancer cell lines has provided an incomplete understanding of the genetic information and mechanisms involved in tumor progression. To provide a more comprehensive picture of the genetic changes that occur in tumors, we have carried out whole-genome sequencing at a single cell level of five common cancers: gastric cancer, liver cancer, renal cancer, bladder cancer, and blood cancer, and assessed the genetic changes that had occurred within these cells. We sequenced 100 single cancer cells and twenty paracancerous normal cells using the Illumina HiSeq 2000 platform, and obtained at least 2-fold coverage for each cancer type. For each cell, we recovered ~85% of the whole genome information. Our cell-to-cell comparative analysis revealed single nucleotide polymorphisms (SNPs), copy number variations (CNVs), and structural variations (SVs) that differed between these cells. These mutations allowed us to cluster these cells into cancer and normal cell types, and then to further divide them into different subpopulations. We then carried out deep sequencing of representative single cell types and identified mutations relevant to tumor progression. These data and the use of single-cell sequencing provide novel information and useful methodology to better understand the genetic mechanisms underlying cancer development and progression.

206

Clinical Application of Microarray Technology to Cancer: Are We Ready for Prime Time? *M.M. Li[1,2,3], X. Hu[1,3], H. Safah[3], A. Iqbal[4], D.J. Wolff[5], J.M Hagenkord[6], C. Lee[7] On behalf of the Cancer Cytogenomics Microarray Consortium.* 1) Hayward Genetics Center; 2) Dept. of Pediatric; 3) Tulane Cancer Center, Tulane University School of Medicine; New Orleans, LA; 4) University of Rochester Medical Center, Rochester, NY; 5) Medical University of South Carolina, Charleston, SC; 6) Creighton Medical Laboratories, Creighton University Department of Pathology, Omaha, NE; 7) Brigham and Women's Hospital, Harvard Medical School, Boston, MA.

Chromosome Microarray (CMA) has been used for several years in clinical laboratories for constitutional applications and in cancer research laboratories as a method to assess copy number alterations (CNAs) and loss of heterozygosity (LOH). However, only a handful of laboratories are offering CMA as a clinical assay for cancer applications. This is largely due to the lack of effective diagnostic algorithms and guidelines for cancer applications, and the absence of a cancer CMA database to aid in post-analytic interpretation. We have established the Cancer Cytogenomic Microarray Consortium (CCMC). The goals of the CCMC include setting up platform-neutral standards for cancer CMA designs, conducting a multi-center, multi-platform clinical validation study to demonstrate the reliability of the cancer CMA for clinical use, and creating a public cancer CMA database. We have designed a combined targeted/whole genome array specific for cancer using the microarray CGH format. Approximately 20,000 high quality oligonucleotide probes were selected to target all exons and exon/intron boundaries of 423 cancer genes from the Sanger cancer gene census, more than 100 known cancer-associated genomic regions, and all subtelomeric regions. Intervals between aforementioned genes or regions were filled relatively evenly with oligonucleotide probes to cover the whole genome. Using this custom designed array, we studied 200 cases of hematological malignancies and 30 normal bone marrow or blood controls. In addition to confirming and clarifying cytogenetic and FISH results, the array identified many cryptic CNAs, including intra-gene deletions and duplications in patients with normal or abnormal karyotype. Many apparently balanced rearrangements harbored cryptic CNAs at or near the breakpoints. These CNAs can be used as novel biological markers to monitor patient responses to treatments and detect minimal residual diseases. The novel CMA design detected all mosaic genomic imbalances of $\geq 15\%$, as determined by FISH, and showed minimal interlaboratory variations. Our experience demonstrates that cancer microarray designs need to target both cancer genes and cancer-associated genomic regions densely as well as cover the whole genome to detect cancer-specific CNAs and identify CNAs in genomic regions that have not been so far associated with cancer. Based on this pilot study, we propose a set of analytic parameters and a preliminary diagnostic algorithm for cancer diagnosis.

207

Mutation in CCS, the copper chaperone for superoxide dismutase (SOD1), implies a novel disorder of human copper metabolism. *P. Huppke[1], C. Brendel[1], C. Korenke[2], I. Marquart[2], O. Elpeleg[3], L.B. Moller[4], J. Christodoulou[5], S. Kaler[6], J. Gartner[1].* 1) Georg August University, Goettingen, Germany; 2) Dept of Neuropediatrics, Children's Hospital, Oldenburg, Germany; 3) Dept of Genetics and Metabolic Disease, Hadassah, Hebrew University Medical Center, Jerusalem, Israel; 4) Kennedy Center, Glostrup, Denmark; 5) Western Sydney Genetics Program, Children's Hospital at Westmead, Australia; 6) Molecular Medicine Program, NICHD, NIH, Bethesda, MD, USA.

Copper is a trace metal with the capacity to readily gain and donate electrons, which renders it both highly desirable as an enzyme cofactor and dangerous as a potential generator of toxic free radicals. Thus, to regulate cellular copper metabolism, an elaborate system of chaperones and transporters has evolved that enable simultaneous utilization of copper and protection from it. Mutations in copper-transporting ATPases produce distinctive disease phenotypes (Menkes and Wilson diseases) but no copper chaperone mutations have been described. Here, we report the first such mutation in a known copper chaperone, delineate the associated phenotype, and apply the findings to further illuminate normal cellular copper metabolism. In a child born to healthy consanguineous parents, we documented neurodevelopmental delay, epilepsy, congenital cataracts, sensorineural hearing loss, liver disease, and low serum copper in association with homozygosity for a mutation (R163W) in the copper chaperone CCS that normally delivers Cu to both copper/zinc superoxide dismutase (SOD1) and XIAP (X-linked inhibitor of apoptosis). The R163W defect is located in domain II of CCS known to facilitate binding to SOD1. Biochemical analyses in patient fibroblasts and transfected HELA cells showed decreased CCS protein, impaired binding of CCS to SOD1, and low SOD activity (46% of wild-type). XIAP independently interacts with COMMD1, a copper export protein defective in the hepatic copper toxicosis of Bedlington terriers. When XIAP is not metallated with copper, it binds to and promotes protosomal degradation of COMMDI, leading to decreased copper export and increased cellular copper concentration. The proband developed liver failure and succumbed at age 3.5 years. We hypothesize that reduced ability to metallate XIAP in this patient led to unbridled degradation of COMMD1 by native unmetallated XIAP, which drastically impaired hepatic copper excretion. In conclusion, we describe a novel autosomal recessive disorder of copper metabolism caused by mutation in CCS, which should be considered in the differential diagnosis of infants with the constellation of low serum copper, neurodevelopmental delay, and liver dysfunction. Further characterization of this condition, including generation of R163W=ACCCS mice, will enhance understanding of the connections between cellular copper homeostasis and apoptosis, and may suggest rational treatment approaches for future affected individuals.

208

Mouse neural stem cells model the Smith-Lemli-Opitz Syndrome brain. *J.-B. Roullet[1], C.A. Wassif[2], S. Impey[1], L.S. Merkens[1], Q. Yang[1], F.D. Porter[2], R.D. Steiner[1].* 1) Oregon Health & Science University, Portland, OR; 2) Eunice Kennedy Shriver, NICHD, Bethesda, MD.

The Smith-Lemli-Opitz syndrome (SLOS) is caused by inactivating mutations of the *DHCR7* gene encoding 7-dehydrocholesterol reductase (DHCR7) the final enzyme in cholesterol synthesis, leading to cholesterol deficiency and precursor (7-dehydrocholesterol, 7DHC) accumulation. SLOS is characterized by mental retardation, and to date there is no cure or treatment. Currently, study of human SLOS brain cells is severely limited by tissue availability. Animal models of the disease also have significant limitations: the *Dhcr7*Δ3-5/Δ3-5 *Dhcr7*$^{-/-}$) KO mouse (Wassif, 2001) die within 1 day after birth, and the phenotype of the hypomorphic *Dhcr7*$^{T93M/}$Δ3-5 mouse (Correa-Cerro, 2006) is biochemically very mild. Therefore, complementary experimental approaches are needed. Because neural stem cells (NSCs) are multipotent, they are potentially useful to elucidate the consequences of DHCR7 deficiency in all brain cell types. In this study we sought to develop a model of the SLOS brain using NSCs isolated from the *Dhcr7*$^{-/-}$ mouse. Whole brain specimens from *Dhcr7*$^{-/-}$, *Dhcr7*$^{+/-}$ and *Dhcr7*$^{+/+}$ E18 fetuses were processed for neurosphere isolation, and adherent NSC cultures were established in poly-DL-ornithine coated dishes with serum-free medium supplemented with growth factors. All 3 cell lines expressed the stem cell markers nestin and Sox2. In contrast to skin fibroblasts isolated from SLOS patients, growth and replication of *Dhcr7*$^{-/-}$ NSCs seemed unaltered. Sterol analysis was performed using GC-MS. Cellular cholesterol concentrations (μg/mg protein) were 27.7, 19.8 and 0.1 for *Dhcr7*$^{+/+}$, *Dhcr7*$^{+/-}$ and *Dhcr7*$^{-/-}$ respectively. 7DHC concentration was 26.2 in *Dhcr7*$^{-/-}$ cells, 0.1 in *Dhcr7*$^{+/-}$ cells and not detected in *Dhcr7*$^{+/+}$ cells. Desmosterol was present in *Dhcr7*$^{+/+}$ and *Dhcr7*$^{+/-}$ cells but not in *Dhcr7*$^{-/-}$ cells. The *Dhcr7*$^{-/-}$ NSC sterol profile is thus similar to that of the whole brain of 1-day old pups (Wassif, 2001). Finally, and in contrast with Wassif 's finding in cortical neurons isolated from *Dhcr7*$^{-/-}$ pups, preliminary experiments with NSCs loaded with fura-2 show preserved glutamate-mediated (Ca^{2+}) signaling in *Dhcr7*$^{-/-}$ NSCs. Thus, DHCR7 deficiency may alter glutamate-sensitive pathways in differentiated neurons or impair differentiation itself. These possibilities are currently under investigation. In conclusion *Dhcr7*$^{-/-}$ NSCs exhibit the typical metabolic phenotype of SLOS. The cells should prove useful in the study of the molecular basis of mental retardation in the disease.

209

Mutations in TMEM165 cause a new Congenital Disorder of Glycosylation type II and affect pH homeostasis. *G. Matthijs[1], F. Foulquier[2], M. Aymere[3], R. Zeevaert[1,4], V. Race[1], R. Bammens[1], L. Keldermans[1], E. Van Schaftingen[5], M. Vikkula[3], J. Jaeken[4].* 1) Laboratory for Molecular Diagnostics, Center for Human Genenetics, University of Leuven, Leuven, Belgium; 2) Université des Sciences et Technologies de Lille, CNRS, Lab de Glycobiologie Structurale et Fonctionnelle, Lille, France; 3) Laboratory of Human Molecular Genetics, de Duve Institute, Université catholique de Louvain, Brussels, Belgium; 4) Center for Metabolic Disease, Department of Paediatrics, University of Leuven, Leuven, Belgium; 5) Laboratory of Physiological Chemistry, de Duve Institute, Université catholique de Louvain, Brussels, Belgium.

Protein N-glycosylation is one of the major biosynthetic functions of the endoplasmic reticulum (ER) and Golgi compartments. The correct oligosaccharide structures observed on secreted and/or membrane glycoproteins depend on the activities of glycosyltransferases and remodeling glycosidases but also on other factors such as a correct localization of Golgi enzymes and/or an adequate Golgi environment. Genetic defects affecting the glycosylation pathway cause a range of diseases known as Congenital Disorders of Glycosylation (CDG). We identified one family with a peculiar phenotype of hypotonia, muscle hypotrophy, osteoporosis, very short stature and joint laxity. The patients were diagnosed as CDG-II, they presented sialylation and galactosylation deficiencies. Using different markers for Golgi sub-compartments, a dilated Golgi morphology associated with a fragmentation of the trans Golgi network (TGN) in fibroblasts from affected individuals was observed. A slight delay in the retrograde translocation of Golgi membranes to the ER was observed in fibroblasts after treatment with brefeldin A. In order to identify the responsible gene, homozygosity mapping was combined with a genome wide transcriptomic analysis to look for genes whose expression would be down regulated. A deep intronic mutation was identified in a novel gene, TMEM165 (TPARL), coding for a protein whose cellular functions are unknown. After screening unrelated CDG-II patients, we identified missense mutations in the same gene in 2 additional cases. To study its subcellular localization, a fusion protein with a fluorescent tag was generated. Co-localization of the tagged protein with relevant markers revealed a late endosomal/lysosomal localization. Because of its peculiar localization, we hypothesize that the function of this new protein is that of a proton transporter. To investigate this possibility, pH measurements in late endocytic structures was done using the Lysotracker and Lysosensor dyes, in patient and control fibroblasts. Compared to control cells, a much stronger staining was observed in patient cells, indicating a more pronounced acidity in the acidic compartments. This result was confirmed in Hela cells by using an siRNA strategy to knock down TMEM165. All together, the discovery of mutations in this unknown gene defined a new type of CDG-II and revealed a completely novel connection between lysosomal pH, glycosylation and intracellular trafficking.

210

Riboflavin responsive OXPHOS complex I deficiency caused by defective ACAD9: new function for an old gene. *M. Gerards[1,2], B.J.C. van den Bosch[1,2], K. Danhauser[3], V. Serre[4], M. van Weeghel[5], R.J.A. Wanders[5], G.A.F. Nicolaes[6], W. Sluiter[7], K. Schoonderwoerd[8], H.R. Scholte[9], H. Prokisch[3], A. Rötig[4], I.F.M. de Coo[10], H.J.M. Smeets[1,2].* 1) Dept Clinic Genomics, Maastricht UMC, Maastricht, Netherlands; 2) Research School GROW, Univ Maastricht, Maastricht, The Netherlands; 3) Institute of Human Genetics, Technische Universität München, Munich, Germany and Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 4) INSERM U781 and Department of Genetics, Hôpital Necker-Enfants Malades, Université René Descartes Paris V, 149 rue de Sèvres, 75015 Paris, France; 5) Department of Clinical Chemistry, Laboratory Genetic Metabolic Diseases, Emma Children's Hospital, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands; 6) Department of Biochemistry, Cardiovascular Research Institute Maastricht, Maastricht University, Maastricht, The Netherlands; 7) Center for Lysosomal and Metabolic Diseases, Erasmus MC University Medical Center, Rotterdam, The Netherlands; 8) Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, The Netherlands; 9) Department of Neuroscience, Erasmus MC University Medical Center, Rotterdam, The Netherlands; 10) Department of Neurology, Erasmus MC University Medical Center, Rotterdam, The Netherlands.

Mitochondrial complex I deficiency is the most common OXPHOS defect. Mutations have been detected in mitochondrial and nuclear genes, but many patients remain unresolved and new genes are likely involved. In a consanguineous family, patients presented since early childhood with easy fatigability, exercise intolerance and lactic acidosis in blood. In muscle, subsarcolemmal mitochondrial proliferation and a severe complex I deficiency were observed. Exercise intolerance and complex I activity improved by supplementation with a high dosage of riboflavin. Homozygosity mapping revealed a candidate region on chromosome three containing six mitochondria-related genes. Four genes were screened for mutations and a homozygous substitution was identified in ACAD9 (c.1594C>T), changing the highly conserved arginine-532 into tryptophan. This mutation was absent in 188 ethnically matched controls. Protein modelling suggested a functional effect due to loss of a stabilizing hydrogen bond in an α-helix and a local flexibility change. To test whether the ACAD9 mutation caused the complex I deficiency, we transduced fibroblasts of patients with wild type and mutant ACAD9. Wild type ACAD9, but not mutant ACAD9, restored complex I activity. An unrelated patient with the same phenotype carried a homozygous ACAD9 c.1405C>T mutation, changing arginine-469 into tryptophan, which was not present in controls. Our data support a new function for ACAD9 in complex I function, making this gene an important new candidate for patients with complex I deficiency, which could be improved by riboflavin treatment.

211

Proteomic changes in very long chain acyl-CoA dehydrogenase (VLCAD) deficient mice reveals a compensatory change in mitochondrial energy metabolism and protein folding. *W. Wang[1,2], J. Palmfeldt[3], N. Gregerson[3], J. Vockley[4].* 1) Department of Human Genetics, GSPH, University of Pittsburgh, Pittsburhg, PA; 2) Center for Inherited Metabolic Disease, Shandong Women's Hospital of Shandong Province Hospital System, Jinan, China; 3) Research Unit for Molecular Medicine, Institute of Clinical Medicine, Aarhus University Hospital and Faculty of Health Sciences, University of Aarhus, Aarhus, Denmark; 4) Department of Pediatrics, School of Medicine, Department of Human Genetics, GSPH, University of Pittsburgh, Pittsburgh, PA.

Background and objective: VLCAD deficiency is an autosomal recessive disease resulting from mutations on ACADVL gene that lead to impaired long-chain fatty acid beta-oxidation. It is clinically heterogeneous and can be lethal if not promptly recognized in the ill child. The expansion of newborn screening with tandem mass spectrometry now allows for presymptomatic identification. The pathophysiologic mechanisms underlying disease symptoms and phenotypic heterogeneity remain unresolved. To better understand the biological changes induced by VLCAD deficiency and explore novel markers to predict prognosis, and guide therapy in patients, we examined quantitative changes in the mitochondrial proteome in mice with knock out of the ACADVL gene. Methods: Mitochondria were isolated from mice liver and purified by Percoll gradient. Trypsin digested samples from wild type and mutant mice were individually labeled with an isobaric tag for relative and absolute quantitation(iTRAQ™)Reagent. All labeled samples for the comparison of protein quantification were combined into one sample mixture and peptides were separated by Isoelectric focusing(IEF),using an immobile pH gradient (IPG) strip. The peptide mixtures extracted from IEF gels were separated and analyzed by nano-LC coupled to LTQ-Orbitrap mass spectrometry through a nano-electrospray source. The generated peak lists of the tandem mass spectra were subjected to search algorithm of Mascot for protein identification and iTRAQ reporter quantification. Results: We found significant alterations in a variety of energy metabolism related proteins in VLCAD deficiency mice. The greatest changes of upregulation were seen in cytochrome b-c1 (complex III), cytochrome c oxidase(complex IV) , H+ transporting mitochondrial ATP synthase(complex V) and the mitochondrial chaperonin HSPD1(Hsp60). Levels of several fatty acid beta-oxidation proteins also changed, with some increasing and others decreasing in a lower magnitude. Conclusion: An increase in electron transport chain related protein levels in mutant mice suggests a compensatory response related to energy deficit induced by ACADVL mutation. Additionally, an increase in mitochondrial chaperonin levels points to a protective stress response. Further study of the diversity of protein changes in mutant animals under stress conditions will help elucidate the ultimate phenotypic variation and pathophysiology of this deficiency.

212

Genetic loss of the serine protease Tmprss6 reduces iron overload in a mouse model of *HFE*-Hemochromatosis by increasing expression of the iron regulatory hormone hepcidin. *K.E. Finberg[1], R.W. Whittlesey[2], N.C. Andrews[2,3].* 1) Dept Pathology, Duke Univ Med Ctr, Durham, NC; 2) Dept Pharmacology & Cancer Biology, Duke Univ Med Ctr, Durham, NC; 3) Dept Pediatrics, Duke Univ Med Ctr, Durham, NC.

HFE-associated Hereditary Hemochromatosis (*HFE*-HH) is a common, autosomal recessive disorder of variable penetrance characterized by excessive storage of iron in multiple organs. Patients with *HFE*-HH display inappropriately low levels of hepcidin, a circulating hormone produced by the liver that acts to inhibit iron absorption in the duodenum. TMPRSS6, a transmembrane serine protease produced by the liver, is a key negative regulator of hepcidin expression; both humans and mice harboring pathogenic *TMPRSS6* mutations display inappropriately elevated hepcidin levels resulting in impaired intestinal iron absorption and systemic iron deficiency. Here we asked if genetic loss of *Tmprss6* could raise hepcidin levels, and thus reduce systemic iron overload, in $Hfe^{-/-}$ mice, a mouse model of *HFE*-HH. To test this, we generated $Hfe^{+/-}Tmprss6^{+/-}$ mice, which we intercrossed to generate offspring of all *Hfe-Tmprss6* genotype combinations for phenotypic characterization. $Hfe^{-/-}$ mice harboring two wild type *Tmprss6* alleles showed systemic iron overload consistent with prior study of $Hfe^{-/-}$ mice. Heterozygous loss of *Tmprss6*, however, markedly reduced the severity of iron overload in $Hfe^{-/-}$ mice, leading to a 50% reduction in liver non-heme iron content. Homozygous loss of *Tmprss6* completely ameliorated iron overload in $Hfe^{-/-}$ mice and led to systemic iron deficiency. $Hfe^{-/-}Tmprss6^{-/-}$ mice were phenotypically identical to $Hfe^{+/+}Tmprss6^{-/-}$ mice, showing microcytic anemia and liver non-heme iron content that was markedly reduced compared to wild type mice. Genetic loss of *Tmprss6* in $Hfe^{-/-}$ mice was associated with increased expression of hepatic mRNA encoding hepcidin as well Id1 and Smad7, two transcriptional targets of a hepatic Bmp/Smad pathway previously implicated in hepcidin regulation. In summary, these results demonstrate that *Tmprss6* is a powerful genetic modifier of the murine *Hfe*-hemochromatosis phenotype that acts by modulating hepcidin transcription by the Bmp/Smad pathway. These results, together with recent findings that common variants in *TMPRSS6* influence laboratory parameters of iron homeostasis in human populations, suggest that natural genetic variation in the human ortholog *TMPRSS6* may contribute to the variable clinical penetrance observed in *HFE*-HH. In addition, these results raise the possibility that pharmacological inhibition of TMPRSS6 activity may prove an effective therapy in *HFE*-HH.

213

Metabolic sink therapy in methylmalonic acidemia using a novel muscle-specific transgenic mouse model. *J.R. Sysol[1], I. Manoli[1], L. Li[2], J. Senac[1], R.J. Chandler[1], V. Hoffmann[3], P. Zerfas[3], J. Schnermann[2], C.P. Venditti[1].* 1) Genetics and Molecular Biology Branch, NHGRI, NIH, Bethesda, MD; 2) Kidney Disease Branch, NIDDK, NIH, Bethesda, MD; 3) Division of Veterinary Resources, ORS, NIH, Bethesda, MD.

Methylmalonic acidemia (MMA) is caused by deficiency of the mitochondrial enzyme methylmalonyl-CoA mutase (Mut) and results in massive elevations of methylmalonic acid in tissues and body fluids. Studies in knockout mice ($Mut^{-/-}$) and transplanted MMA patients have suggested that a large portion of circulating methylmalonic acid derives from extrahepatic organs, mainly the skeletal muscle. To examine the effects of restoring skeletal muscle expression of the Mut enzyme on the $Mut^{-/-}$ phenotype and gain insight into the efficacy of targeting skeletal muscle for metabolic "sink" therapy for MMA, we generated mice that express the *Mut* gene under the control of an insulated, muscle-specific promoter ($Mut^{-/-}$;$Tg^{INS\text{-}MCK\text{-}}$*Mut*). $Mut^{-/-}$;$Tg^{INS\text{-}MCK\text{-}}$*Mut* mice were born in Mendelian proportions, showed greater than 83% survival past day of life 60 (N=40), and achieved 40-50% of their heterozygous littermates weight through the first year of life. Muscle-specific *Mut* RNA expression in $Mut^{-/-}$;$Tg^{INS\text{-}MCK\text{-}}$*Mut* mice was 103±4.6% compared to $Mut^{+/-}$ and was accompanied by abundant immunoreactive enzyme in muscle. To further assess transgene function, we measured the oxidation of 1-^{13}C propionate into $^{13}CO_2$. The $Mut^{-/-}$;$Tg^{INS\text{-}MCK\text{-}}$*Mut* mice metabolized 18.4±3.6% of the label in 25 min, compared to 76.5±4.5% in $Mut^{+/-}$ and 10±2% in $Mut^{-/-}$. Baseline plasma MMA levels (μM) were 1107.9±66 in transgenic mice, compared to <5 in controls. The $Mut^{-/-}$;$Tg^{INS\text{-}MCK\text{-}}$*Mut* animals develop significant liver pathology, characterized by giant eosinophilic vacuoles and megamitochondria formation, which was associated with decreased respiratory chain complex IV activity (18.2±7.4% relative to controls), similar to the $Mut^{-/-}$ mice. More variable changes were noted in the tubular epithelial cells and were associated with a decreased glomerular filtration rate as measured by inulin clearance. Selective muscle expression of the Mut enzyme by transgenesis at levels matching or exceeding the heterozygous controls resulted in near uniform rescue of the neonatal lethal phenotype of the $Mut^{-/-}$ mice, but was unable to prevent liver and kidney damage. This novel murine model demonstrates that enzymatic correction of skeletal muscle can augment metabolism in MMA, but cell-autonomous and/or toxic effects from circulating metabolites may mediate hepatic and renal tubular pathology. It also provides a new platform for testing of liver and/or kidney-directed gene, cell and other therapies for this disease.

214

Handling of methylmalonic acid and its precursor amino acids across the blood brain barrier in patients with methylmalonic acidemia who had received liver transplantation. *M. Yoshino[1], T. Oohira[1], Y. Watanabe[1,2], J. Okada[1], T. Ohya[1], K. Tashiro[2], T. Inokuchi[2], T. Matsuishi[1,2], Y. Shigematsu[3].* 1) Dept Pediatrics & Child Hlth, Kurume Univ Sch Med, Kurume, Japan; 2) Lab for Med Appl Gas-chromatogr Mass spec, Kurume Univ Sch Med, Kurume, Japan; 3) Sch Nurs Fukui Univ, Fukui, Japan.

Background: Methylmalonyl-CoA mutase (MCM) is expressed in neurons as well as some other cells in the central nervous system (CNS). Efflux of short chain dicarboxylic acids, including methylmalonic acid (MMA), out of CNS cells is higher than influx. **Patients report:** Patient 1: This girl with MMA had been fed a dietary regimen that provided 0.5-0.7g/kg/day of natural protein until she underwent liver transplantation (LT) at the age of 7 years and 3 months of age, after which natural protein intake was relaxed to 1.2g/kg/day. Three months before the surgery, she had begun to develop episodes of quick torsional movements of the head. This involuntary movement worsened 3 months after the surgery and she developed an episode of tonic seizure at 8 years and 5 months of age. The concentrations (μmol/L) of MMA in plasma (pMMA) and in cerebrospinal fluid (cMMA) were 47 and 270, respectively, before surgery, and 63 and 419, respectively, after surgery. Patient 2. This boy had been given 0.4-0.7g/kg/day of natural protein until the age of 5 years and 2 months, when he underwent LT. Natural protein intake was then increased to 1.2g/kg/day. At 7 years and 2 months of age, he exhibited weakness of the right extremities and flexion of the right upper extremity. MR imaging recorded at this time revealed high signal intensity on T2-weighted imaging in the bilateral globus pallidus. The concentration of pMMA and cMMA were 32 and 20, respectively, and 100 and 483, respectively, after surgery. **Discussion:** It is known that efflux of methylmalonic MMA out of CNS cells is higher than influx. Nevertheless, the concentration of MMA was 4- to 6-fold higher in the CSF than in blood after the relaxation of natural protein intake in these patients. This observation indicates that the fraction of MMA produced in the MCA-deficient CNS contributes more greatly to the whole MMA pool in CSF than that imported into CNS from blood. Thus, the relaxation of the precursor amino acids may enhance MMA accumulation in the CNS, which cannot be rescued by LT. **Conclusion:** Relaxation of natural protein intake may precipitate neurological symptoms in MCM-deficient patients, even if they undergo LT.

215

Nitric oxide production in subjects with MELAS syndrome and the effect of arginine and citrulline supplementation: interim results. *A. El-Hattab[1], J.W. Hsu[2], W. Craigen[1], L.J. Wong[1], F. Jahoor[2], F. Scaglia[1].* 1) Dept Molec & Human Gen, Baylor Col Med, Houston, TX; 2) Dept of peditarics, Baylor Col Med, Houston, TX.

Background. The mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS) syndrome is one of the most frequent maternally inherited mitochondrial disorders. The pathogenesis of stroke-like episodes remains unclear. It is believed that epithelial dysfunction leads to nitric oxide (NO) deficiency and ischemic events. Both arginine and citrulline act as NO precursors. **Methods.** In this study we are measuring NO production rate using a stable isotope infusion technique and the plasma concentrations of NO metabolites (nitrite and nitrate, NOx), arginine, and citrulline in 15 control subjects in a single admission and in 15 subjects with MELAS before and after arginine and citrulline supplementations. The aims are to determine whether NO production is slower in subjects with MELAS, whether arginine and citrulline supplementation will increase NO production, and whether citrulline will increase NO production more substantially than arginine supplementation due to the fact that arginine requires a transporter to enter the cell and the action of both arginase and nitric oxide synthase on the intracellular arginine. **Results.** Three control subjects and 6 patients have completed the study. Patients have lower arginine, citrulline, and NOx concentrations than control subjects; however, these differences have not reached statistical significance. Additionally, no significant differences between patients and control subjects were found in NO production rate. In patients, citrulline supplementation resulted in more significant increases in plasma arginine and citrulline concentrations than arginine supplementation. Plasma NOx concentration increased after arginine supplementation (17.9±1.6®20.2±1.8 μM, p=0.09); however, the increment was higher after citrulline (18.0±2.4®21.0±2.3 μM, p=0.06). NO production rate increased after arginine supplementation (0.069±0.014®0.095±0.020 μM, p=0.05); however, the increment was more pronounced after citrulline (0.073±0.012®0.378±0.12 μM, p<0.05). **Conclusion.** The interim analysis revealed that in comparison to arginine, citrulline supplementation to subjects with MELAS has led to a more significant increase in NOx, arginine, and citrulline concentrations and NO production rate. Ultimately, the completion of this pilot study could shed light on better therapeutic strategies for the management of stroke-like episodes in subjects with MELAS syndrome.

216

Ammonia (NH3) Control in Children with Urea Cycle Disorders (UCDs); Comparison of Sodium Phenylbutyrate and Glycerol Phenylbutyrate. *U. Lichter-Konecki[1], GA. Diaz[2], JL. Merritt II[3], A. Feigenbaum[4], C. Jomphe[5], JF. Marier[5], M. Beliveau[5], J. Mauney[6], K. Dickinson[6], A. Martinez[7], M. Mokhtarani[7], B. Scharschmidt[7], W. Rhead[8].* 1) Dep. Pediatrics, George Washington University, Div Gen & Metabolism, Children's National Med Ctr, Washington, DC; 2) Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, NY; 3) Seattle Children's Hospital, 4800 Sand Point Way NE, Seattle, WA; 4) The Hospital for Sick Children, 555 University Avenue, Toronto Ontario, Canada; 5) Pharsight Corp., Montreal, 2000 Peel St., Suite 570, Quebec, Canada; 6) Chiltern, 2520 Independence Blvd., Ste. 202, Wilmington NC; 7) Hyperion Therapeutics, Inc., 601 Gateway Blvd., Ste. 200, South San Francisco, CA; 8) Pediatrics, Medical College of Wisconsin, 8701 Watertown Plank Rd., Milwaukee, WI.

Daily NH3 profiles and correlates of drug effect were examined in a phase 2 comparison of sodium phenylbutyrate (NaPBA) and glycerol phenylbutyrate (GPB; glyceryl tri-[4phenylbutyrate] or HPN-100), an investigational drug being developed for UCDs. Study Design. Protocol HPN-100-005 involved open label fixed-sequence switch-over from the prescribed NaPBA dose to a PBA-equimolar GPB dose with controlled diet. After 7 days on NaPBA or GPB, subjects underwent 24-h blood sampling for NH3 and drug metabolite levels as well as measurement of 24-hour urinary phenyacetylglutamine (PAGN). Adverse events (AEs), safety labs and triplicate ECGs were monitored. Subjects completing the switch-over were offered 1-year open label GPB treatment. Results. 11 subjects (9 OTC, 1 ASS, 1 ASL) enrolled and completed the switch-over from NaPBA (mean dose = 12.4g/d or 322 mg/kg/d; range = 198 - 476 mg/kg/d) to GPB (mean dose = 10.8 mL or 0.285 mL/kg/d or 313 mg/kg/d; range = 192 - 449 mg/kg/d). Possibly-related AEs were reported in 1 subject on NaPBA and 4 subjects on GPB. All were mild, except one moderate AE of vomiting on GPB related to an intercurrent illness. No clinically significant laboratory or ECG changes were observed. NH3 values on both drugs were lowest after overnight fast, peaked in the afternoon to early evening and varied widely over 24-h with occasional values > 100 μmol/L without symptoms. NH3 (24 hour AUC) was ~25% lower on GPB vs. NaPBA, as were average daily mean (28.7 vs. 37.7 μmol/L; p ≥0.1) and peak NH3 values (47.8 vs. 55.7 μmol/L; p ≥0.1). The upper 90% (1.002) and 95% (1.061) confidence intervals for the difference between NH3 levels on GPB and NaPBA were less than the predefined non-inferiority margin of 1.25, indicating that non-inferiority was achieved. Plasma phenylacetic acid and PAGN (24 hour AUC) were comparable for GPB vs. NaPBA, as were U-PAGN and the percentage of orally administered PBA excreted as PAGN (68% for GPB vs. 66% for NaPBA). GPB and NaPBA dose correlated best with urinary-PAGN. All 11 subjects entered the ongoing safety extension and none have reported a hyperammonemic crisis during the initial ~60 total patient-months of GPB treatment. Conclusions. These preliminary findings suggest that GPB is at least equivalent to NaPBA in terms of NH3 control and are encouraging with respect to its tolerability and potential utility in pediatric UCD patients. They further suggest that U-PAGN may be a clinically useful biomarker.

217

Neither 'rare' nor 'common' variants can explain much of phenotypic variation. *V. Pihur[1,2], A. Chakravarti[1].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Department of Biostatistics, Johns Hopkins University, Baltimore, MD.

The contribution of common genetic variation to human complex and polygenic traits and diseases has been successfully surveyed by current genome-wide association studies (GWAS). The vast majority of variants discovered to date, however, have small phenotypic effects that explain, on average, only a small proportion of the phenotypic variation or disease risk. Explanations of this "missing" heritability are plenty and several complementary theories could potentially fill the heritability gap. An attractive hypothesis is that rare variants with larger phenotypic effects might explain away the "missing" heritability. But, what constitutes rare or common? To answer the question of how much phenotypic variation is explained by sub-polymorphic (<1%), uncommon polymorphic (>1% but <10%) and common (>10%) variants, we focused our attention to modeling the effect size of an allele as a function of its frequency. Based on published data from a catalog of GWA studies, largely limited to common variants, the average allelic effect size(s) across many quantitative traits is ~0.05 standard deviation units and, for qualitative phenotypes, the average odds ratio is ~1.3. The true means are likely even lower given the 'winners curse' with present-day GWAS sample sizes. Unfortunately, little data exist on less common variants and their effects. We fit the following model to assess the effects of rare alleles versus its frequency (q), $s = \alpha e^{-\beta q} + \gamma$, where $\alpha+\gamma$ is the allelic effect when q=0, γ is the allelic effect when q=1, and β is a scale parameter modulating the frequency effect. The surprising prediction is that neither very rare (<0.1%) nor common (>10%) variants are likely to explain much of the variability. Rather, alleles with frequencies between 0.1-1% make the largest contribution to heritability, followed closely by alleles with frequencies between 1-10%. The second frequency category is within our immediate reach with the progress made by the 1000 Genomes Project, expected to provide a comprehensive catalog of variants with frequency >0.5%. Since the expected number of such variants is very large (>12 million), and imputation is unlikely to cover variants under 2%, a focus on the manageable subset of coding variants (~150,000) can provide adequate statistical power to test for association with current sample sizes. As tempting as it might be to put our faith in rare variants, the next logical step in GWAS studies is to study all polymorphic (>1%) variants.

218

Identifying large-effect rare variants causing synthetic associations by using haplotype association analysis. *A. Coventry[1], L. Bull-Otterson[2], X. Liu[3], A. Clark[1], T. Maxwell[3], J. Hixson[3], T. Rea[4], A. Templeton[5], E. Boerwinkle[3], R. Gibbs[2], C. Sing[4].* 1) Molec Biol & Gen, Cornell Univ, Ithaca, NY; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, 77030; 3) Human Genetics Center, UT Houston Health Science Center, Houston, TX, 77030; 4) Department of Human Genetics, University of Michigan School of Medicne, Ann Arbor, MI 48109; 5) Department of Biology, Washington University, St Louis, MO 63130.

We describe a method for selecting resequencing loci and samples to follow up a marker-based GWAS. We phase samples, then target for resequencing any haplotypes associated with elevated triglyceride levels. This obviates the controversial concern raised by Goldstein and colleagues (*PLoS Biology*, Jan 2010, doi:10.1371/journal.pbio.1000294) that causative variants may be so distant from the GWAS peak they cause that they are missed by the usual *ad hoc* methods for choosing the resequencing locus. That could occur if a causal variant is of sufficiently recent origin that few recombination events have broken up the haplotype on which it arose. By making the haplotype the unit of inference in our association study, we get a clear signal of when that is happening. We also get a clear indication when multiple variants are contributing to a GWAS peak, because different causative variants very likely arose on different haplotypes, leading to multiple haplotypes associated with phenotype. This allows us to also ascertain whether GWAS peaks are often caused by multiple rare variants, as predicted by Goldstein and colleagues.

We are applying this approach to 13,422 human samples from the ARIC cohort, using Affymetrix SNP Array 6.0 genotypes. We are scanning for haplotypes in significant association with elevated triglyceride levels, and are currently resequencing a set of loci in individuals carrying such haplotypes.

Among the nine loci that we have so far found this way, there are four covered by multiple associated haplotypes, suggesting that multiple causative variants do indeed contribute to synthetic associations. For most of these nine loci, the associated haplotypes give a clear indication of involvement by a single gene, but there are two where the associated haplotypes cover multiple genes, suggesting that Goldstein *et al.* are correct in their concern that resequencing loci are typically too short. However, they were perhaps too pessimistic in recommending that to be sure of finding the causative variant, future resequencing efforts should target huge loci. A haplotype-based approach to choosing resequencing loci can substantially narrow down both the target loci and the candidate allele carriers.

219

Coming Full Circle: An Hypothesis-driven GWAS (GWAS-HD) with Application to Meconium Ileus in Cystic Fibrosis. *L.J. Strug[1,2], L. Sun[2], W. Li[1], R. Dorfman[1], C. Taylor[1,2], F. Wright[3], L. Henderson[4], M. Drumm[5], M. Knowles[3], G. Cutting[1,2], J. Rommens[1,2], P. Durie[1,2] on behalf of the North American Cystic Fibrosis Gene Modifier Consortium.* 1) Hosp Sick Children, Toronto, ON, Canada; 2) University of Toronto, Toronto, ON; 3) University of North Carolina at Chapel Hill, Chapel Hill, NC; 4) Johns Hopkins University, Baltimore, MD; 5) Case Western Reserve University, Cleveland, OH.

It is common for genome-wide association studies (GWAS) to result in suggestive, not overwhelming evidence for association. Methods to help prioritize observed results can be useful at this stage. Pathway analysis (GWAS-PA) prioritizes the results, however an interactive pathway is but one hypothesis for disease mechanism and can be restrictive, since contributing genes/proteins can relate to residual or compensating activities in other ways that could span multiple processes. We develop a more general approach, the *Hypothesis-driven GWAS* (GWAS-HD), where *any* hypothesis can be tested within this framework. We applied the GWAS-HD to a modifier gene study of Meconium Ileus (MI) in Cystic Fibrosis (CF). We were interested in testing the hypothesis that constituents of the apical plasma membrane, of which CFTR is a member, may contribute to MI. We used data from the North American CF Gene Modifier Consortium, consisting of 3,655 CF patients, 617 with MI, genotyped at 556,445 SNPs. A GWAS for MI did not contain any genome-wide significant results, based on a p-value criteria of 5x10^-8 or a q-value (FDR adjusted p-value) of 0.05. To apply the GWAS-HD for our hypothesis we: (1) prioritized the genome by generating a list of proteins present in the apical plasma membrane (based on the GO consortium, AmiGO version 1.7); (2) used the stratified FDR (SFDR) to weight our GWAS p-values and determine genome-wide significance for given loci; and (3) used simulation to determine the statistical significance of our hypothesis. Our list consisted of 151 genes spanning 3,734 GWAS SNPs. Using the SFDR, we identified SNPs from four genes within the apical plasma membrane list that reached genome-wide significance (q-value < 0.05), with q-values as low as 6x10^-4. Permuting the case-control status and re-conducting the association analyses of the 3,734 SNPs, allowed us to assess the significance of the whole AmiGO list, while retaining the LD structure of the SNPs within the genes ($p < 0.05$). The GWAS-HD allowed for testing a specific hypothesis within the context of a GWAS, while simultaneously determining significance of individual loci after prioritizing the genome. In our application to MI, the GWAS-HD provided significant evidence that multiple genes present at the apical plasma membrane may contribute to the MI phenotype; and we were able to prioritize four genes for further study despite the absence of genome-wide significance in single-SNP analysis.

220

A new method for detecting epistatic interactions in genome-wide association studies. *M. Yoshida, A. Koike.* Central Res Lab, Hitachi, Ltd, Tokyo, Japan.

Detection of epistatic interactions is one of the significant challenges in genome-wide association studies (GWAS). Because of the sophisticated regulatory mechanisms encoded in the human genome, complex diseases are speculated to be caused by multiple factors and their interactive effects. On the other hand, current GWAS have primarily focused on testing association of only a single SNP at a time. Identifying epistatic interactions in GWAS will therefore likely be a key to further understanding of common complex diseases. In this context, we have developed a new effective method for identifying epistatic interactions in GWAS by extending a learning approach called random forest. The random forest technique is one of the predictive methods that produces a series of classification trees using a large set of predictor variables, and has been proposed for use to discover SNPs which are most predictive of the disease status in large-scale association studies. The technique can detect SNPs which are likely to affect disease susceptibility from among a large number of SNPs taking account of effects of their interactions. It has, however, some limitations for identifying epistatic interactions. First, it may perform poorly for detecting SNPs with little marginal effects. Furthermore, it does not explicitly exhibit information on interaction patterns of susceptibility SNPs. We have extended the random forest framework to overcome the above limitations and established an applicable method for identifying epistatic interactions by means of (i) modifying the construction of the random forest, and (ii) implementing a procedure of extracting interaction patterns from the random forest constructed. The performance of the proposed method has been evaluated by simulation data under a wide spectrum of disease models. The results demonstrate that the new method performs very well in successfully identifying pure epistatic interactions with high precision, and is still more than capable of concurrently identifying multiple interactions under the existence of genetic heterogeneity. The proposed new method gets rid of many of the difficulties that previous approaches have faced, such as high dimensionality problems, multiple testing problems, and genetic heterogeneity, and is promising for practical use in GWAS to search epistatic interactions involved with common complex diseases. Acknowledgements: This study was supported by the integrated database project of MEXT.

221

A novel gene-based test of association for the identification of multiple independent effects. *H. Huang[1,2], A. Alonso[3], J.S. Bader[1,2], D.E. Arking[4].* 1) Dept Biomedical Engineering, Johns Hopkins Univ, Baltimore, MD; 2) High Throughput Biology Center, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN; 4) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.

Genome-wide association studies (GWAS) are now used routinely to identify SNPs associated with complex human phenotypes. These studies have increasingly relied upon larger sample sizes to identify smaller effects, with limited approaches beyond simple univariate tests. It is increasingly clear that many identified loci are likely to have multiple independent genetic effects, and leveraging this observation can lead to a substantial gain in power.

Here we introduce a novel Gene-Wide Significance (GWiS) test that uses Bayesian model selection to identify the number of independent effects within a gene, which are combined to generate a stronger statistical signal. Permutation tests provide p-values that correct for the number of independent tests genome-wide and within each genetic locus.

Applied to simulated data, GWiS has better power than traditional tests to identify genes that have multiple weak associations. We then applied GWiS to a dataset comprising about 2.5 million SNPs in up to 8,000 individuals measured for various ECG parameters. Using a meta-analysis of almost 50,000 individuals as the gold standard, GWiS identified 6 validated associations while traditional univariate methods identified 4 and LASSO, which is a multivariate method using the LASSO algorithm, identified 2 loci. GWiS also provides, for the first time, a systematic assessment of the fraction of disease-associated genes housing multiple independent effects, observed at 35-50% of loci in our study, which in turn strongly suggests that our method will reveal previously unidentified associations when applied to existing data and will improve power for future association studies.

The overall computational cost of GWiS scales linearly with the number of SNPs and the number of individuals in the study. GWiS can be generalized to other study designs, phenotype distributions, and to sequence data. It provides gene-based p-values that are directly compatible with pathway-based meta-analysis. The GWiS source code is available under an open source license from www.baderzone.org.

222

Enhanced disease scoring, imputation and fine-mapping for GWAS in admixed populations: assessment of increased power using African Americans from CARe. *B. Pasaniuc[1,2], S. Pollack[1,2], N. Zaitlen[1,2], G. Lettre[3], A. Tandon[4], L. Kao[5], I. Ruczinski[5], M. Fornage[6], D. Siscovick[7], X. Zhu[8], E. Larkin[8], L. Lange[9], A. Cupples[10], Q. Yang[10], M. Akylbekova[11], S. Musani[12], J. Divers[13], J. Mychaleckyj[14], M. Li[15], G. Papanicolaou[16], S. Myers[17], D. Reich[4], N. Patterson[2], J. Wilson[18], A.L. Price[1,2].* 1) Harvard School of Public Health; 2) Broad Institute; 3) Université de Montréal; 4) Harvard Medical School; 5) Johns Hopkins University; 6) UT Health Science Center; 7) University of Washington; 8) Case Western Reserve University; 9) University of North Carolina; 10) Boston University; 11) Jackson State University; 12) University of Mississippi; 13) Wake Forest University; 14) University of Virginia; 15) University of Pennsylvania; 16) National Heart, Lung and Blood Institute; 17) Oxford University; 18) University of Mississippi Medical Center.

Genome-wide association studies (GWAS), which initially focused on populations of European ancestry, are increasingly being extended to admixed populations such as African Americans and Latinos. While admixed populations offer the promise of capturing additional genetic diversity, standard GWAS analysis techniques fail to take advantage of variation in local ancestry across the genome, which can be inferred with high accuracy. Here we present a complete suite of statistical methods for analyzing GWAS in admixed populations. We illustrate the gain in power achieved by our methods by analyzing 6,209 unrelated African-American samples from the CARe project genotyped on the Affymetrix 6.0 chip, in conjunction with both simulated and real phenotypes. We introduce new case-control scoring statistics that combine SNP and admixture association signals and show that jointly modeling both signals yields an 8% increase in power for finding disease risk loci over the power achieved by standard methods such as the Armitage trend test with correction for genome-wide ancestry, with a much larger increase of over 20% for disease risk variants with large population differences. When the causal variant is untyped and needs to be imputed using data from reference panels, we show that imputation performance is increased when local ancestry is explicitly modeled in the imputation procedure. At imputed SNPs, we observe a 12% increase in power for mapping disease loci when our local ancestry aware imputation framework and the new scoring statistic are jointly employed, as compared to the Armitage trend test with SNPs imputed by MACH using a CEU+YRI reference panel. Notably, the framework is generic and can be incorporated into any imputation method. Finally, we introduce new methods that leverage the genetic variability between the European and African ancestry components as well as the differences between SNP and admixture association information to improve the accuracy of fine-mapping causal variants at associated loci. In previous work we showed that localization success rate of fine-mapping is increased when multiple populations are used in the study compared to a study that involves only one population. Here we report an additional improvement in localization success rate when our new scoring and imputation approaches are employed. Publicly available software implementing our methods is broadly applicable to genome-wide association studies in admixed populations.

223

Genetic Determinants of Telomere Length Localized by Linkage, Association, and Gene Expression Profiling. *J. Kent[1], S. Kumar[1], M. Carless[1], J.E. Curran[1], T.D. Dyer[1], V.P. Diego[1], J. Charlesworth[2], M.P. Johnson[1], A.G. Comuzzie[1], M.C. Mahaney[1], L. Almasy[1], E.K. Moses[1], H.H.H. Goring[1], J. Blangero[1], S. Williams-Blangero[1].* 1) Dept of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX USA; 2) Menzies Research Institute Tasmania, Hobart, TAS Australia.

Telomeres are terminal chromosomal structures consisting of a repetitive DNA sequence and associated proteins. Age- and disease-related reduction in telomere length is associated with genomic instability and cell senescence/death. We measured telomere length (as T/S ratio of telomere repeat copy number to a reference single-copy gene, by multiplexed quantitative PCR) in peripheral blood mononuclear cells (PBMCs) from 1,428 members of extended Mexican American families enrolled in the San Antonio Family Heart Study (SAFHS). As expected, T/S ratio is negatively correlated with age (P=0.00013). Adjusted for sex, age, age^2 and age^n x sex interactions, and normalized, T/S ratio is heritable (h^2=0.25±0.05, P=5.2x10^{-12}). We performed genome-wide association and linkage analysis in 1,187 of the SAFHS cohort. This subset comprised 704 women and 483 men representing a range of ages (15-92ya) and clinical status (median BMI=28.6kg/m^2, range=14-63; 14.8 percent with diabetes). Association tests were conducted in SOLAR allowing for residual non-independence due to kinship. Association at empirical genome-wide significance (P<1.3x10^{-7}) was found on chromosome 6 (rs9500256: peak association at P=1.8x10^{-11}, and rs9476380, bracketing a region that includes multiple tRNA genes), 3 (rs1447826, rs6797523), 21 (rs2236479), 12 (rs1975920), and 1 (rs10922613). The peak QTL by linkage (LOD=2.94) was on chromosome 16p13.13, with rs9939578 in the nucleotide-binding protein 1 gene *NUBP1* showing maximum joint evidence for linkage and association (P=9.4x10^{-6}). PBMC gene expression profiles are available for the genotyped SAFHS participants. The most highly associated SNP for T/S ratio, rs9500256, is also significantly associated (after correction for multiple tests) with 27 RefSeq gene transcripts, including those of the ribonuclease H1 gene *RNASEH1* (P=1.9x10^{-8}), transcription factor *GTF3C4* (P=2.6x10^{-8}), putative tumor suppressor *DOK1* (P=1.8x10^{-8}), and oncogene *JUND* (P=7.7x10^{-7}). These results from a Mexican American sample identify candidate loci not found in previous studies in Caucasians and may suggest novel pathways for telomere maintenance and function.

224

Using Cross-Population Modeling to Improve Inference at Rare Variants. *B.N. Howie[1], J. Marchini[2], M. Stephens[1].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Statistics, University of Oxford, Oxford, UK.

Now that many diseases have been tested for association with common genetic variants, researchers are hoping that rarer variants will explain some of the residual heritability of these phenotypes. While many of the techniques used to measure and analyze rare variants differ from those used on common variants, some fundamental steps remain the same: in many cases it is still necessary to call genotypes from intensity data, phase mutations onto haplotypes, and impute genotypes from dense reference panels into large study cohorts. Sophisticated models of linkage disequilibrium have been successfully applied to these problems before, but they are less effective at rare variants. One promising solution is to combine information across populations: a haplotype that is rare in one population may have drifted to higher frequency in another, implying that data from the latter population could help infer the rare haplotype. This simple idea is hard to implement in modern genetic datasets because of the computational demands of modeling structured populations. We have developed an efficient model that overcomes these limitations by approximating local subtrees of large, structured genealogies; this functionality is implemented in our software package IMPUTE2. Our method requires minimal supervision, is feasible in large, complex datasets, and is more accurate than leading competitors in a variety of applications. To illustrate, we use IMPUTE2 to genotype, phase, and impute rare variants in data from HapMap 3 and the 1,000 Genomes Project. Our results show that IMPUTE2 can robustly combine information across populations, and that this approach can substantially improve inference at rare variants.

225

Leveraging Publicly Available Sequencing Data in the post-GWAS Era to Identify Novel Significant Association Signals. *Y. Li[1,2], D.M. Waterworth[3], L. Li[4], Y. Zhou[2], P. Vollenweider[5], G. Waeber[5], V.E. Mooser[3], G.R. Abecasis[6], M.G. Ehm[4].* 1) Dept Genetics, Univ North Carolina, Chapel Hill, NC, USA; 2) Dept Biostatistics, Univ North Carolina, Chapel Hill, NC, USA; 3) GlaxoSmithKline, Upper Merion, PA, USA; 4) GlaxoSmithKline, Research Triangle Park, NC, USA; 5) Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland; 6) Center for Statistical Genetics, Department of Biostatistics, University of Michigan.

Genome-wide association studies (GWAS) have successfully identified over 2,000 loci associated with disease risk or variations in quantitative traits. With the advent and rapid advances in massively parallel sequencing technologies, sequencing individuals to identify rare variants associated with complex phenotypes will identify disease susceptibility alleles with larger genetic effects. However, the cost of sequencing studies remains prohibitive. In the meantime, large international efforts such as the 1000 Genomes Project are generating sequence data for public usage. We show that one can identify novel signals as well as refine association profiling in known regions by utilizing such public data in the post-GWAS era with no additional costs for experiments. Using 5,479 individuals from the Lausanne, Switzerland, CoLaus population-based study, we have successfully imputed, using MaCH, ~6.2 million SNPs using available Affy 500K GWAS data and as reference ~120 CEU haplotypes generated by the 1000 Genomes Pilot One Project (haplotypes and software both available at http://www.sph.umich.edu/csg/yli/mach/). Using LD among the 6.2 million SNPs, we estimated that the upper bound for the number of independent tests is two million. Evaluation of ~6.2 million variants for a number of cardiovascular and metabolic traits suggested several novel loci for a number traits at stringent genome-wide significance level (p-value < 2.5*10e-8 using a Bonferroni correction for two million independent tests). In addition, we generated association maps with much better resolution by increasing the number of directly examined markers by 2.5 fold over HapMap-based imputation. In particular, we identified several new non-synonymous or splice site junction SNPs in medium to high LD with the top GWAS SNPs. Our study suggests imputation using publicly available sequence data is a sensible post-GWAS next step to help explain the missing heritability, to find the causal variant(s), and to refine GWAS identified regions, before large scale genome-wide medical sequencing becomes widely available.

226

Imputation Accuracy for Rare Variants in Structured Populations. *M. Zawistowski[1,2], S. Zöllner[1,2,3].* 1) Department of Biostatistics, The University of Michigan, Ann Arbor, MI; 2) Center for Statistical Genetics, The University of Michigan, Ann Arbor, MI; 3) Department of Psychiatry, The University of Michigan, Ann Arbor, MI.

Next generation sequencing allows detection of all rare genetic variants in a dataset, making it possible to test their contribution to phenotypic variation. However, for the near future, price may limit the number of samples that can be sequenced. Here, we consider the possibility of increasing sample size of resequencing studies by using existing methods to impute rare variant genotypes. Imputation is routinely used to increase power to detect common risk variants in genome-wide association studies (GWAS) but it is not currently known whether it is sufficiently precise to be applied to rare variants. In particular, rare variants are typically the result of recent mutations, hence recent demographic history within structured populations may determine the extent to which rare variants may be accurately imputed. To assess this, we simulated sequencing datasets using coalescent based models. We partitioned each dataset into a reference sample, assumed to be fully sequenced, and a target sample containing only genotypes for set of SNPs patterned after a typical GWAS experiment. We phased the sequenced sample and used the resulting haplotypes as templates to impute untyped rare variants into the target sample. We used a range of population genetic models to create structured populations from which to draw the reference and target samples, allowing us to consider a variety of imputation strategies. Focusing on variants with maf<5%, we find that the major allele is nearly always called correctly but that accuracy for the minor allele is dependent on various properties of the reference sample. Notably, imputation accuracy depends on how representative the reference sample is of the target sample. Accuracy is optimized when the reference and target samples are drawn from the same underlying population. However, in practice, reference and target samples may be derived from closely related but distinct populations. Therefore, we report accuracy results as a function of population differentiation. These results provide insight into the utility of publicly available resources (ie 1000 Genomes) to serve as reference panels for imputing rare variants. Finally, we show that increasing sample size of resequencing studies by imputation can dramatically improve power for rare variant association tests.

227

Differences in embryo survival depending on meiosis I or meiosis II origin of chromosomal aneuploidy. *A. Kuliev[1], Z. Zlatopolsky[1], I. Kirillova[1], J. Cieslak-Jansen[2].* 1) Reprod Gen Inst, Chicago, IL; 2) Colorado Center for Reproductive Medecine, Denver, Co.

It has been established that the majority of aneuploidies in spontaneously aborted fetuses and live-born children originate from female meiosis I, with the exception of the chromosome 18 trisomy, originating more frequently in meiosis II. We performed the direct testing of the meoisis I and II outcomes in series of 20,946 oocytes, by removal and FISH analysis of PB1 and PB2, using five-color probe specific for chromosomes 13, 16, 18, 21 and 22. 46.8% of oocytes in IVF patients of over 38 were abnormal, varying from 20 to 55% depending on maternal age. Although the observed aneuploidies originated comparably from meiosis I and II, one third of meiosis II errors had also meiosis I errors, with the aneuploidy rescue in 45% of oocytes involving the same chromosome error in meiosis I and II. The majority of meiosis I errors were of chromatid type, with a nonrandom retention of extra chromatids in MII oocytes, which constitutes the major source of aneuploidy, predicting a predominance of trisomy (53%) over monosomy (26%) in the resulting embryos, in agreement with trisomy predominance in spontaneous abortions and live-born children. As much as 40% of the detected aneuploidies were of complex nature, with the involvement of two or more chromosomes, or the same or different chromosomes in both meiotic divisions. There was a nonrandom involvement of different chromosomes in meiosis I and meiosis II errors, showing a higher prevalence of aneuploidies for chromosomes 22 and 21. Of special interest were significant differences of the chromosome specific origin of errors: chromosome 16 and 22 errors originated more frequently in meiosis II (44.4% and 41.5 % meiosis II errors vs. 32.0% and 34.3% meiosis I, respectively), chromosome 18 errors derived more frequently from meiosis I (48.3% in meiosis I vs. 34.6% in meiosis II), and chromosome 13 and 21 derived comparably from both meiosis I and II (40.1% and 41.4 in meiosis I and 36.3% and 36.7% in meiosis II, respectively). The data demonstrates a nonrandom chromosome-specific pattern of error, which is in contrast to the above mentioned observations in spontaneous abortions and live-born children, providing the first evidence for possible embryo survival differences depending on the meiotic origin of aneuploidy.

228

Adverse Effects of Nucleotide Excision Repair and Transcription Gene Abnormalities on Human Fetal and Placental Development. *R. Moslehi[1], A. Kumar[1], C. Signore[2], J.L. Mills[2], A. Dzutsev[3].* 1) Epidemiology & Biostatistics, Ctr Cancer Genomics, SUNY at Albany, Rensselaer, NY; 2) NICHD, NIH, Bethesda, MD; 3) NCI, NIH, Frederick, MD.

Background: Effects of abnormalities in DNA repair and transcription genes in human prenatal life have never been studied. Our recent genetic epidemiologic investigation of gestational outcomes associated with abnormalities in trichothiodystrophy nucleotide excision repair(NER) and transcription genes, namely *XPD*(*ERCC2*), *XPB*(*ERCC3*), *TTD-A*(*GFT2H5*) and *TTDN1*(*C7ORF11*), revealed significantly increased risk of several severe gestational complications including preeclampsia(RR=4.0,P=0.0018), hemolysis, elevated liver enzymes, and low platelets (HELLP) syndrome(RR=35.7,P=0.0002), elevated midtrimester maternal serum human chorionic gonadotropin (hCG) levels(RR=14.3,P<0.0001), small for gestational age(<3rd percentile)(RR=13.9,P<0.0001), preterm delivery(<32 weeks)(RR=12.0,P<0.0001), and decreased fetal movement(RR=3.3,P=0.0018). Abnormal placental development may explain the constellation of observed abnormalities and we hypothesize that DNA repair/transcription genes are involved in normal placental development. **Methods**: To test this hypothesis and decipher biologic mechanisms, we extracted and compared NER/transcription and global gene profiles in normal human placentas and in meta-analysis of placentas from pregnancies with preeclampsia. Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNE) was used to infer transcriptional relationships. **Results**: We found high expression of TTD genes in normal placenta, above the mean of their expression in all organs. *XPD*, *XPB*, and *TTDN1* were consistently expressed from 14 to 40 weeks gestation while expression of *TTD-A* was strongly negatively correlated(r=0.7, P<0.0001) with gestational age. Meta-analysis of preeclampsia datasets implicated imbalance of several pathways including the oxidative stress response pathway. The largest group of downregulated genes belonged to RNA PolymeraseII-mediated pathways and included most of the subunits of transcription factorII(TFIIH) complex such as XPD, XPB and TTD-A. Application of ARACNE revealed *GTF2E1* (component of TFIIA which modulates TFIIH) as the key regulator of gene expression in preeclampsia. **Conclusion**: Our results indicate an important role for TTD NER/transcription gene products during normal human placental development and implicate impaired TFIIA-TFIIH interaction and dysfunction of RNA PolymeraseII-mediated pathways caused by abnormalities in these genes as a relevant mechanism leading to the observed gestational abnormalities such as preeclampsia.

229

"Impact of mtDNA mutations on mtDNA segregation throughout human oogenesis". *S. Monnot[1], N. Gigarel[1], D.C. Samuels[2], P. Burlet[1], L. Hesters[3], N. Frydman[3], R. Frydman[3], V. Kerbrat[3], B. Funalot[1], J. Martinovic[4], A. Benachi[5], J. Feingold[1], A. Munnich[1], J-P. Bonnefont[1], J. Steffann[1].* 1) INSERM unit U781, Necker Hospital, and Paris-Descartes University, Paris, France; 2) Center for Human Genetics Research, Department of Molecular Physiology and Biophysics, Vanderbilt University Medical Center, Nashville, TN, USA; 3) Reproductive Medicine Department, A Beclere Hospital, Clamart, France; 4) Cytogenetics unit, Necker Hospital, Paris, France; 5) Gynaecology-Obstetrics Department, Necker Hospital, Paris, France.

Mitochondrial DNA (mtDNA) mutations cause a wide range of serious diseases with high transmission risk and maternal inheritance. Tissue heterogeneity of the heteroplasmy rate ("mutant load") accounts for the wide phenotypic spectrum observed in carriers. Owing to the absence of therapy, couples at risk to transmit such disorders commonly ask for preimplantation diagnosis (PGD). The lack of data regarding heteroplasmy distribution throughout oogenesis hampers both genetic counseling and implementation of a PGD procedure. Human early embryonic development is indeed fully depending on the mtDNA pool of the mature oocyte, since mtDNA synthesis is thought not to occur in the early embryo prior to implantation. We tracked the segregation of 5 mtDNA mutations, namely m.3243A>G (MTTL1 gene, MELAS syndrome), m.8344T>G (MTTK gene, MERRF syndrome), m.8993T>G and m.9185T>C (MTATP6 gene, NARP/Leigh syndromes), and m.10197G>A (MTND3 gene, Leigh syndrome) during oogenesis. We quantified the corresponding mutant loads in 6 carriers, their 14 oocytes, 53 first polar bodies (PB1), and 65 early embryos, thought to provide an accurate estimate of the mature oocyte mutant load. Our data mainly indicate that 1/ mtDNA segregation is governed by random genetic drift (at least for m.3243A>G) throughout the whole oogenesis; in particular, mtDNA randomly segregates between PB1 and the oocyte, irrespective of the mutation type. 2/ m.3243A>G germ cells are subjected to a "purifying" selection process over a critical heteroplasmy threshold, while m.8993T>G, m.8344T>G, and m.10197G>A cells are not. 3/ the size of the bottleneck operating during oogenesis i) is individual-dependent for m.3243A>G, with a calculable probability of fixing on the mutant or wild-type, and ii) is likely to depend on the type of mtDNA mutation. 4/ While the absence of mutation constantly correlates between PB1 and the corresponding oocyte/embryo, mutant load may vary significantly (+30%) between the 2 clonally-derived daughter cells, irrespective of the mutation type. These data are discrepant with those drawn from heteroplasmic mouse models, supporting the need for studies in humans. 5/ there is no correlation between the PB mutant load and the rate of cleavage, suggesting that oocyte fertilization and early embryonic development are not markedly affected by mtDNA mutations. These data have obvious consequences on genetic counseling and PGD procedures in mtDNA inherited disorders.

230

Trisomic pregnancy and elevated FSH: implications for the oocyte pool hypothesis. *J. Kline[1,2], A. Kinney[1], B. Levin[1], A. Kelly[3], M. Ferin[3], D. Warburton[4].* 1) Mailman School of Public Health, Columbia University, New York, NY 10032; 2) NY State Psychiatric Institute, New York, NY 10032; 3) Dept. Obstetrics and Gynecology, Columbia University, New York, NY 10032; 4) Dept. Genetics and Development, Columbia University, New York, NY 10032.

Some studies, but not all, support the hypothesis that trisomy frequency at conception is related to the size of the oocyte pool, with risk increased for women with fewer oocytes (older ovarian age). We tested the hypothesis by comparing hormonal indicators of ovarian age among women who had trisomic pregnancy losses with those among women with non-trisomic losses or chromosomally normal births. The three primary indicators of advanced ovarian age were low level of anti-Mullerian hormone (AMH), high level of FSH and low level of dimeric inhibin B. The cases, derived from a hospital-based case-control study, comprised 105 women with meiotic errors leading to trisomic, hypertriploid or autosomal monosomic losses. Cases were compared with 36 women with other chromosomally abnormal losses, 57 women with chromosomally normal losses and 279 women with livebirths age-matched to women with losses. We used multiple linear regression models to test the null hypothesis that, at any maternal age, there is no difference between trisomy cases and the comparison groups in ovarian age indicators. We also used conditional logistic regression to compare the proportions of cases and livebirth controls with low AMH (£ 20th percentile in livebirths, £ 0.385 ng/ml), high FSH (≥ 10 mIU/ml) and low inhibin B (< 20th percentile in livebirths, < 24.1 pg/ml). All analyses adjust for age in single years. AMH and inhibin B levels did not differ between women with trisomic losses and any of the three comparison groups. The geometric mean of FSH levels was 22.3% higher (95% CI 11.1%, 34.6%) for trisomy cases compared with livebirth controls; the geometric mean was also higher, though not significantly so, for trisomy cases than for women with other chromosomally abnormal losses (10.7%) or chromosomally normal losses (13.0%). The adjusted odds ratios for trisomy in relation to low AMH and low inhibin B were 1.6 and 1.2, respectively (nonsignificant). High FSH was significantly increased in trisomy cases (12.4%) versus livebirth controls (6.1%); the adjusted odds ratio was 3.4 (95% CI 1.3, 8.7). The association of trisomy with elevated FSH levels is compatible with the oocyte pool hypothesis, while the absence of an association with AMH, expressed by the granulosa cells and thought to reflect the size of the oocyte pool, is not. An alternative explanation for the association with FSH is that elevated levels may have a direct effect and disrupt meiotic processes.

231

Detection of ≥ 1 Mb microdeletions and microduplications in a single cell using a custom oligonucleotide array: Relevance to PGD and non-invasive prenatal testing. *W. Bi[1], A. Breman[1], C.A. Shaw[1], P. Stankiewicz[1], G. Fruhman[1], X.-Y. Lu[1], Z. Ou[1], J.M. Sederstrom[1], A. Patel[1], S.W. Cheung[1], L.G. Jackson[2], J.R. Lupski[1], I. Van den Veyver[1], A.L. Beaudet[1].* 1) Baylor College of Medicine, Houston, TX, USA; 2) Drexel University College of Medicine, Philadelphia, PA, USA.

Objectives: High resolution detection of genomic copy number abnormalities in a single cell is very relevant to preimplantation genetic diagnosis (PGD) and potentially to noninvasive prenatal diagnosis. The sizes of most common genomic disorders range between 1 - 5 Mb, which is beyond the proven resolution of single-cell array comparative genomic hybridization (array CGH). Our objective is to develop a reliable array CGH platform to detect genomic imbalances of > 1 Mb in a single cell. **Methods:** We experimentally optimized the conditions of whole genome amplification (WGA) and oligonucleotide-based array CGH on single cells from multiple lymphoblastoid cell lines with known copy number abnormalities. To improve resolution, we designed custom arrays covering clinically relevant regions with empirical selection of oligonucleotide probes based on our experience with >28,000 postnatal cases and on probe performance with single-cell WGA DNA hybridization. **Results:** Using single-cell WGA and custom arrays with different probe densities in the targeted regions, we found that detection of copy number changes in a single cell is influenced by the number of probes clustered in the interrogated region and that detection was improved by removing data from poorly performing probes. Using our custom array with a high density coverage in the regions for all common microdeletion/duplication syndromes, we reproducibly detected trisomy 21, a 5 Mb Prader-Willi syndrome deletion, a 3.7 Mb Smith-Magenis syndrome deletion and reciprocal Potocki-Lupski syndrome duplication, a 3 Mb DiGeorge syndrome deletion, a 1.5 Mb Williams syndrome deletion, and a 1.5 Mb CMT1A duplication, as well as X- and Y-chromosome copy number differences in gender-mismatched hybridizations. Replicate analyses yielded consistent results. **Conclusion:** Aneuploidy and genomic imbalances as small as 1.5 Mb in a single cell are detectable by array CGH using arrays with high-density coverage in the targeted regions. This approach has the potential to be applied in PGD and noninvasive prenatal diagnosis to detect aneuploidy and almost all common microdeletion/duplication syndromes. Attempts to analyze PGD samples and fetal cells from the maternal circulation are in progress.

232

Validation of single nucleotide polymorphism (SNP) microarray preimplantation genetic diagnosis on single cell(s) from embryos. *W.G. Kearns[1,2], A. Benner[1], C. Chipko[1], E. Widra[3], R. Leach[1].* 1) Shady Grove Center for Preimplantation Genetics, Rockville, MD; 2) Gynecology/Obstetrics, The Johns Hopkins University School of Medicine, Baltimore, MD; 3) Shady Grove Fertility Reproductive Science Center, Rockville, MD.

Objective: To amplify DNA from a single blastomere and to perform complex SNP microarray genetic analyses. Validation studies will be discussed. Design: Prospective study Materials and Methods: Multiple displacement amplification was performed on 962 single cells from 270 cytogenetically abnormal embryos as determined by 10-probe FISH and 34 known cell lines. Invariant DNA genomic loci were used to ensure the entire genome was amplified and TaqMan PCR to ensure heterozygous allele amplification. The Illumina Human HapMap or Cyto-12 microarrays were employed including genotype data for 300-370K SNPs. Data was analyzed with deCODE Genetics Disease Miner Professional and Illumina GenomeStudio and KaryoStudio software. Results: Analyses of blastomeres and cell lines showed in many cases, a genomic coverage > 98%, a heterozygous allele detection rate > 90% and a microarray detection rate and genotype call rate > 90%. A 23-chromosome molecular karyotype was obtained from over 99% of all blastomeres and all 34 cell lines. Structural chromosome imbalances were identified from all cytogenetically abnormal cell lines. These aberrations included deletions and duplications. Genotype information was also obtained for over 85% of all SNPs analyzed. Upon reconstruction of all cells from individual embryos, 5% of embryos showed aneuploidy for the same 3 chromosomes, 64% of embryos were mosaic diploid/aneuploid, 23% were mosaic aneuploid and 8% were complex mosaics. These genome wide scans can also identifiy some subtle DNA aberrations associated with disorders such as Beckwith-Wiedemann syndrome, some types of Prader Willi / Angelman syndrome, DiGeorge syndrome 1/ Velocardiofacial syndrome and some single gene disorders. Uniparental disomy and Copy Number Variations can also be identified. Conclusions: We successfully validated and obtained complex genetic information from 962 single embryonic cells using a modified WGA protocol and SNP microarray analyses. These analyses may be performed on polar bodies, blastomeres or trophectoderm cells.

233

Single nucleotide polymorphism (SNP) microarray preimplantation genetic screening (PGS): A comparison between abnormal day-3 blastomeres and corresponding inner cell mass and trophectoderm cells. *K.D. Nguyen[1], R. Ross[2], A. Benner[1], Y. Zhao[3], P. Brezina[3], W.G. Kearns[1,3].* 1) Genetics, Shady Grove Center for Preimplantation Genetics, Rockville, MD; 2) LaJolla IVF, LaJolla, Ca; 3) Gynecology/Obstetrics, The Johns Hopkins University School of Medicine, Baltimore MD.

OBJECTIVE: To compare SNP microarrays for PGS on day-3 blastomeres, inner cell mass (ICM) cells and trophectoderm (TE) cells from corresponding blastocysts DESIGN: Observational study MATERIALS AND METHODS: One cell from each of 11 day-3 embryos were biopsied by laser and sent to a California genetics reference lab for 23-chromosome SNP PGS analysis. The mean maternal age was 34. All embryos were diagnosed as abnormal and differentiated in culture to the blastocyst stage. Once the embryos reached the blastocyst stage, the ICM and TE cells were separated by laser and placed into 5 ul of DNA stabilizing buffer. Approximately 5-10 cells were placed in each tube. The cells were then shipped to our genetics laboratory in Rockville, MD. Upon receipt of cells, we amplified the DNA using a modified whole genome amplification (WGA) protocol. We employed the Illumina Cyto-12 microarray to determine chromosome aberrations and to obtain genotype data for ~300K SNPs. Data was analyzed with deCODE genetics Disease Miner Professional and Illumina GenomeStudio software. RESULTS: Our SNP microarray analysis differed from all 11 day-3 blastomere SNP PGS microarrays performed at the referring genetics laboratory. Our analysis showed that six of 11 embryos previously determined to be abnormal by day-3 biopsy and 23-chromosome SNP microarray analysis had a normal karyotype in both the corresponding ICM and TE cells. Three of the remaining 5 embryonic ICM and TE cells showed discordance between trophectoderm and ICM cells involving triploidy, trisomy, deletions, and duplications. The remaining 2 embryos showed a concordant cytogenetic abnormality between the ICM and TE cells. CONCLUSION: Our preliminary data suggests that day-3 PGS abnormal blastomeres with complex aneuploidy may self-correct during differentiation to the blastocyst stage of development. Alternative explanations could include a diagnostic error from the first genetics laboratory or an error by our genetics laboratory.

234

Comparative Analysis of Chromosomal Aberration Detection Methods in Invasive Prenatal Diagnosis. *L. Armengol[1], J. Nevado[2,3], C. Serra-Juhé[3,4], A. Plaja[5], C. Mediano[5], F. García-Santiago[2,3], M. García[1], O. Villa[3,4], E. Mansilla[2,3], C. Preciado[3,4], M.A. Mori[2,3], P.D. Lapunzina[2,3], L.A. Pérez-Jurado[3,4,5,6].* 1) QGenomics Laboratory, Barcelona, Spain; 2) Instituto de Genética Médica y Molecular, Hosp Univ La Paz, Madrid, Spain; 3) Centro de Investigación Biomédica en Red en Enfermedades Raras (CIBERER), Spain; 4) Unitat de Genètica, Universitat Pompeu Fabra, Barcelona, Spain; 5) Programa de Medicina Molecular i Genètica, Hosp Univ Vall d'Hebron, Barcelona, Spain; 6) Dept of Genome Sciences, Univ of Washington, Seattle, WA.

Objective: Comparative study of currently available methodologies for invasive prenatal detection of chromosomal abnormalities. **Methodology**: Multicentric collection of a yearly series of pregnant women with indication for prenatal invasive sampling (amniocentesis or corionic villous biopsy). Simultaneous utilization of three screening methodologies when enough (uncultured) material could be obtained: 1) Karyotype + Quantitative Fluorescent Polymerase Chain Reaction (QF-PCR) (current national medical standard), 2) Multiplex Ligation-dependent Probe Amplification (MLPA) with two panels and 3) array-based Comparative Genomic Hybridization (aCGH) with targeted BAC microarray. **Results**: A total of 900 pregnant women provided informed consent to enter the study with very high acceptability (94%). Technical performance was very good for karyotype, QF-PCR and aCGH (~0.5 failure rate), but relatively poor for MLPA (10% failure rate). Turnaround time was ~8 days for aCGH or MLPA and 14-23 for karyotype, with similar combined costs for all three methods. A total of 54 samples (6%) were found to carry pathogenic or clinically relevant aberrations (4,2% and 1,8%, respectively). aCGH alone yielded the best detection rate, ~25% higher than any of the other methodologies, missing a single chromosomal alteration (triploidy 69, XXX). Detection yield was higher among the group with abnormal ultrasound findings (22%). The identification of variants of uncertain clinical significance by aCGH (1%) doubled that of karyotype and MLPA, but the alterations could be classified as likely benign after testing parental samples given that they all were inherited. **Conclusion**: Higher detection rate (~2% of samples undetected by other methods) with similar combined cost and lower turnaround time favour aCGH as the best method for invasive prenatal detection of chromosomal abnormalities in at risk pregnancies. The high detection rate (close to 1/200) also in the *a priori* low risk group (maternal anxiety or advance maternal age with normal triple screening), suggests that evaluation of the indications for prenatal invasive testing with the new methodologies should be considered.

235

Standard-of-Care Screening Beyond Cystic Fibrosis: Targeting the Most Frequent Gene and Chromosome Mutations. *R.V. Lebo[1], W.W. Grody[2].* 1) Dept Pathology, Akron Children's Hosp, Akron, OH; 2) Department of Pathology, School of Medicine, University of California Los Angeles, California.

Ranking genetic disease mutations by disease frequency will facilitate developing optimally useful carrier screening tests that identify the largest proportion of affected fetuses at the fewest number of disease gene sites tested. For example, testing 100,000 Northern European pregnant patients for 23 cystic fibrosis mutations would identify 35 mothers carrying affected fetuses among the 141 at-risk couples carrying two CFTR gene mutations, whereas testing for 77 additional cystic fibrosis mutations would detect only 1 additional affected fetus. In contrast, testing instead for the common SMN1gene deletion would identify 10 additional affected fetuses, and including 6 recurrent DMD mutation sites would identify 10 more affected fetuses for a total of 55 affected fetuses among 100,000 pregnant women tested. Simultaneous aneuploidy analysis of 35 selected chromosome regions would detect 97% of all visible abnormal chromosome abnormalities while avoiding inconclusive polymorphic sites. Furthermore, population mutation screening lists can be targeted according to patient age. For instance, fetal testing might also include common diseases like Rett syndrome with high de novomutation rates, while older patients can be tested for diseases that are not expressed until after birth or only as the patient ages. Testing and reporting only listed mutation sites would optimize test accuracy by simultaneously minimizing the frequency of false positive and negative test results that increase along with the total number of tested sites. Publicly available lists of diseases and mutations ordered by frequency as we propose here would provide professionals, laboratories, oversight committees, third-party payers and government officials an objective means of evaluating potential genetic screening targets based upon rational to optimize clinical utility and health care costs.

236

Common genetic variants are significant risk factors for early menopause: results from the Breakthrough Generations Study. *M. Weedon[1], C.E. Bennett[1], J.R.B. Perry[1], P.A. Jacobs[2], D.H. Morris[3], N. Orr[4], M.J. Schoemaker[3], M. Jones[3], A. Ashworth[4], A.J. Swerdlow[3], A. Murray[1], ReproGen Consortium.* 1) Peninsula Medical School, University of Exeter, St Lukes, Exeter, EX1 2LU, UK; 2) Wessex Regional Genetics, Salisbury District Hospital, Salisbury, SP2 8BJ, UK; 3) Section of Epidemiology, The Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK; 4) Breakthrough Research Centre, The Institute of Cancer Research, 237 Fulham Road, London, SW3 6JB, UK.

Women become infertile approximately 10 years before menopause and as more women delay childbirth into their thirties, the number of women who experience infertility is likely to increase. Tests that predict the timing of menopause would allow women to make informed reproductive decisions. Current predictors are only effective just prior to menopause, there are no long range indicators. Age at menopause and early menopause are highly heritable, suggesting a genetic aetiology. Recent genome wide scans have identified four loci associated with variation in age of normal menopause (40-60 years). We aimed to determine whether theses loci are also risk factors for early menopause. We tested the four menopause-associated genetic variants in a cohort of approximately 2000 women with menopause £ 45 years from the Breakthrough Generations Study. All four variants significantly increased the odds of having early menopause. Comparing the 4.5% of individuals with the lowest number of risk alleles (2 or 3) to the 3.0% with the highest number (8 risk alleles), the odds ratio was 4.1 (95% CI, 2.4 - 7.1, p = 4.0x10-7). In combination the four variants discriminated early menopause cases with an ROC area under the curve of 0.6. Four common genetic variants have a significant impact on the odds of having early menopause and may be useful for predicting reproductive lifespan. The advantage of a genetic predictive test is that it can be carried out at any age and does not rely on peri-menopausal changes to have occurred.

237

SRD5A3 is required for converting polyprenol to dolichol and is mutated in a new congenital disorder of glycosylation. *V. Cantagrel[1], B.G. Ng[2], Z. Guan[3], D. Lefeber[4], E. Morava[5], J.L. Silhavy[1], S.L. Bielas[1], D. Babovic-Vuksanovic[6], L. Al-Gazali[7], R.A. Wever[5], C.R.H. Raetz[3], H.H. Freeze[2], J.G. Gleeson[1].* 1) Neurogenetics Laboratory, HHMI, Department of Neurosciences, University of California, San Diego, La Jolla, CA; 2) Genetic Disease Program, Sanford Children's Health Research Center, Sanford-Burnham Medical Research Institute, La Jolla, CA; 3) Department of Biochemistry, Duke University Medical Center, Durham, NC; 4) Department of Laboratory Medicine, Institute for Genetic and Metabolic Disease, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 5) Department of Paediatrics, Institute for Genetic and Metabolic Disease, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 6) Departments of Medical Genetics, Pediatric Neurology, Laboratory Genetics, Pediatric Endocrinology, and Dermatology. Mayo Clinic, Rochester, MN; 7) Departments of Pediatrics and Pathology, United Arab Emirates University, Faculty of Medicine and Health Sciences, Al Ain, United Arab Emirates.

Dolichol phosphate is the lipid carrier of the oligosaccharides used for N-linked glycosylation in eukaryotic cells. The last step of de novo synthesis of dolichol is known to be the reduction of the alpha-isoprene unit of polyprenol by an unidentified polyprenol reductase. We identified a large consanguineous family with a multisystem disorder of cerebellar hypoplasia, ocular coloboma, developmental delay and skin ichthyosis. Linkage analysis and systematic genes sequencing revealed a truncation mutation in the Steroid 5 alpha- reductase gene (SRD5A3). Based on sequence homology, SRD5A3 has been classified as a member of the steroid reductase enzyme family, but its function is unknown. Serum transferrin test revealed a congenital disorder of glycosylation type I (CDG-I) in this family. Lipid-linked oligosaccharide (LLO) characterization showed a severe decrease in total amount of LLO in patients' fibroblasts. Screening of CDG-Ix patients allowed us to identify additional independent homozygous mutations. Analysis of a yeast strain mutated for a homologous gene showed a severe defect in N-glycosylation which was specifically rescued by the human SRD5A3 gene compared to other members of the same family. The characterization of a mouse mutant with a gene-trap allele for the Srd5a3 gene revealed that the homozygous mutant is embryonic lethal and shows an activation of the unfolded protein response pathway. Investigation of polyprenoids present in yeast and mouse mutants using liquid chromatography-mass spectrometry (LC-MS) revealed strong accumulation of polyprenols in mutant versus wt samples. Finally, by over-expressing the human enzyme in the yeast mutant as well as using in vitro conditions we were able to detect efficient reduction of polyprenols. Further investigation of the developmental defects in the mouse mutant is underway. This work resulted in the identification of a new glycosylation disorder and the characterization of a gene necessary for polyprenol reduction and N-glycosylation in yeast, mouse and human.

238

Identity-by-Descent Filtering of Exome Sequence Data identifies PIGV mutations in Hyperphosphatasia-Mental Retardation syndrome (HPMR). *P.N. Robinson[1,2,3], P.M. Krawitz[1,2,3], M.R. Schweiger[2,3], C. Rödelsperger[1,2,3], C. Marcelis[4], U. Kölsch[5], C. Meisel[5], F. Stephani[4], T. Kinoshita[6], Y. Murakami[6], J. Hecht[1,3], S. Bauer[2], J. Grünhagen[2], M. Isau[3], A. Fischer[3], A. Dahl[3], M. Kerick[3], B. Jonske de Condor[2], H.G. Brunner[4], P. Meinecke[7], E. Passarge[8], M. Thompson[9], D.E. Cole[9], D. Horn[2], T. Roscioli[4,10], S. Mundlos[1,2,3].* 1) Institute for Medical Genetics, Charité-Universitätsmedizin Berlin, Berlin, Germany; 2) Institut für Medizinische Genetik, Charité Universitätsmedizin Berlin, Berlin, Germany; 3) Max-Planck -Institute für Molekulare Genetik, Berlin, Germany; 4) Department of Human Genetics, University Medical Centre St. Radboud, Nijmegen, The Netherlands; 5) Institut für Medizinische Immunologie, Charité Universitätsmedizin Berlin, 10117 Berlin, Germany; 6) Department of Immunoregulation, Research Institute for Microbial Diseases, Osaka University, Osaka, Japan; 7) Medizinische Genetik, Altonaer Kinderkrankenhaus, Hamburg, Germany; 8) Institut für Humangenetik, Universitätsklinikum Essen, Essen, Germany; 9) Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto ON, Canada; 10) Department of Molecular and Clinical Genetics Sydney University, Australia.

Recessive mutations are relatively common in the human genome but their identification remains challenging, especially in cases in which methods such as linkage analysis are not applicable. With the development of next-generation sequencing technologies these diseases can now be re-addressed. Initial efforts at using exome sequencing for disease gene discovery analyzed small numbers of unrelated individuals and used intersection filters to search for genes with mutations in all affected persons. Here, DNA from three sibs of non-consanguineous parents with HPMR was analyzed by exome sequencing using ABI SOLiD following enrichment of exonic sequences. Called variants were filtered to exclude variants not found in all affected persons as well as common variants, which however still left 14 candidate genes. To further reduce the search space, we developed a statistical model that allows identical by descent regions to be inferred from exome sequences of only the affected children. In autosomal recessive disorders, the affected children inherit identical maternal and paternal haplotypes in a region surrounding the disease gene, meaning that both haplotypes originate from the same maternal and paternal haplotype, but not necessarily from an identical ancestor (IBD=2). HMM-IBD=2 analysis reduced the candidate list from 14 to 2 genes, SLC9A1 and PIGV. Further homozygous and compound heterozygous PIGV mutations affecting highly conserved residues of PIGV were detected in patients from three other unrelated families. PIGV encodes the second mannosyltransferase in the GPI anchor biosynthesis pathway. We showed a significant reduction of GPI-anchor expression in patient lymphocytes together with a marked reduction in expression of the GPI-anchored protein CD16. Transfection of a PIGV clone in a PIGV-deficient cell line was able to restore surface expression of GPI-anchored marker proteins, but a PIGV clone with one of the mutations identified in HPMR patients could not. In summary, we have identified PIGV mutations in the autosomal rececssive syndromic mental retardation syndrome HPMR using whole-exome capture and SOLiD sequencing in combination with an HMM algorithm to identify IBD=2 regions. Our algorithm can be used in combination with other bioinformatic filters to streamline gene discovery in exome sequencing projects.

239

Targeted high throughput sequencing shows that mutations in WDR62, encoding a centrosome-associated protein, cause microcephaly and abnormal cortical architecture. *T.W. Yu[1,2,3,4], G.H. Mochida[1,2,4], D. Tischfield[1,2], S. Sgaier[1,2], L. Flores-Sarnat[5], C. Sergi[6], M. Topcu[7], M. McDonald[8], B. Barry[1,2], J. Felie[1,2], C. Sunu[1,2], W.B. Dobyns[9], R. Folkerth[10], A.J. Barkovich[11], C.A. Walsh[1,2,3].* 1) Division of Genetics and HHMI, Children's Hospital Boston, Boston, MA; 2) Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA; 3) Broad Institute of MIT and Harvard, Cambridge, MA; 4) Division of Child Neurology, Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA; 5) Department of Clinical Neurosciences, Division of Paediatric Neurology, Alberta Children's Hospital, University of Calgary Faculty of Medicine, Calgary, AB T3B 6A8, Canada; 6) Department of Laboratory Medicine & Pathology, University of Alberta Hospital, 5B4.09 Walter Mackenzie Health Sciences Centre, Edmonton, AB T6G 2B7, Canada; 7) Department of Pediatrics Section of Pediatric Neurology, Hacettepe University, Medical Faculty, Ihsan Dogramaci Children's Hospital, Sihhiye 06100, Ankara, Turkey; 8) Division of Genetics, Duke University, Durham, NC 27710; 9) Division of Genetics, University of Washington at Seattle, WA 98195; 10) Division of Neuropathology, Brigham and Women's Hospital, Boston, MA 02115; 11) Department of Radiology, University of California San Francisco, San Francisco, CA 94143.

Genes associated with human microcephaly, a condition characterized by a small brain, have identified critical regulators of proliferation, cell fate, and DNA repair. Here we describe a new autosomal recessive syndrome of severe congenital microcephaly, with simplification of the brain's gyral (folding) pattern. Genome-wide linkage screens in two families identified a 7.5 Mb locus on chromosome 19q13.12 containing 148 annotated genes. Comprehensive sequence analysis of genes in the interval from one affected individual in each family using targeted high throughput methods yielded over 4000 DNA variants and implicated a single gene, *WDR62*, as harboring rare, potentially deleterious changes in both families. The pathogenicity of these *WDR62* mutations was confirmed by identification of additional *WDR62* mutations in four additional families. MRI and postmortem brain analysis of affected individuals support an important role for *WDR62* in proliferation of neuronal precursors and potential roles in cell migration and axon outgrowth. WDR62 is a WD40 repeat-containing protein expressed in neuronal precursors in the developing brain and localizes to mitotic spindles. *WDR62* represents a new and surprising mediator of neurogenesis and patterning in the human cerebral cortex.

240

A Novel Mouse Model Recapitulates Defects Seen in 1p36 Deletion Syndrome. *B. Kim[1], H. Zaveri[1], Z. Yu[1], O. Shchelochkov[2], S. Kang[1], M. Justice[1], B. Lee[1], D. Scott[1].* 1) Dept. of Molecualr and Human Genetics, Baylor College Med, Houston, TX; 2) Dept. of Pediatrics, University of Iowa, Iowa, IA.

Monosomy 1p36 is the most common terminal deletion syndrome with an incidence of 1 in 5,000 birth and results in cardiovascular defects, cardiomyopathy, mental retardation, CNS abnormalities, postnatal growth deficiency, and vision problems. Although most features of monosomy 1p36 have been attributed to genes located within 6.2 Mb of the telomere, Kang et al. has reported a cohort of patients with similar symptoms caused by interstitial deletions of 1p36 located proximal to this region. The smallest deletion in this cohort involved a 2.9 Mb region containing *RERE* (*Atrophin 2*), which encodes a nuclear receptor co-repressor. Mice that are homozygous for the *Rere* null mutation openmind (*om*) die at E9.5 with open neural tubes and cardiac defects. However, using a newly identified hypomorphic allele of *Rere* known as *eye3*, we have generated an allelic series of mice with partial *Rere* function. A portion of *Rere* $^{om/eye3}$ mice live into adulthood and show many of the characteristics seen in 1p36 deletion patients. Fibrosis, which is a histological feature of cardiomyopathy, was found in the hearts of *Rere* $^{om/eye3}$ mice at adult stage. The cerebellums of *Rere* $^{om/eye3}$ mice show a limited fissure formation, delay of granule cell migration, and attenuation of Purkinje cell dendrite branching during the postnatal development. *Rere* $^{om/eye3}$ mice also fared poorly in tests requiring normal balance and coordination. A decrease of ganglion cell number was observed in the retina starting at E17.5 and an increase in apoptotic cells are also seen in the ganglion cell layer at E17.5. Since cardiomyopathy, CNS abnormalities, postnatal growth deficiency and vision problems are found in *Rere* $^{om/eye3}$ mice, it is likely that haploinsufficiency of *RERE* contributes to the development of these features in children with 1p36 deletions involving this gene. We are presently using our *Rere* mouse models to identify the molecular interactions between RERE and specific nuclear receptors that are responsible for these defects.

241

Dominant negative mutations in α-II spectrin cause early onset West syndrome with severe cerebral hypomyelination, spastic quadriplegia, and developmental delay. *H. Saitsu[1], J. Tohyama[2], T. Kumada[3], K. Egawa[3], K. Hamada[4], I. Okada[1], T. Mizuguchi[1,17], H. Osaka[5], R. Miyata[6], T. Furukawa[3], K. Haginoya[7], H. Hoshino[8], T. Goto[9], Y. Hachiya[10], T. Yamagata[11], S. Saitoh[12], T. Nagai[13], K. Nishiyama[1], A. Nishimura[1], N. Miyake[1], M. Komada[14], K. Hayashi[15], S. Hirai[15], K. Ogata[4], M. Kato[16], A. Fukuda[3], N. Matsumoto[1].* 1) Dept Human Gen, Grad Sch Med, Yokohama City Univ, Yokohama, Japan; 2) Dept Pediatrics, Epilepsy Center, Nishi-Niigata Chuo National Hospital, Niigata, Japan; 3) Dept Physiology, Hamamatsu Univ Sch Med, Handayama, Japan; 4) Dept Biochemistry, Grad Sch Med, Yokohama City Univ, Yokohama, Japan; 5) Division of Neurology, Clin Res Ins, Kanagawa Children's Medical Center, Yokohama, Japan; 6) Dept Pediatrics, Tokyo Kita Shakai Hoken Hospital, Tokyo, Japan; 7) Dept Pediatrics, Tohoku Univ Sch Med, Sendai, Japan; 8) Division of Neurology, National Center for Child Health and Development, Tokyo, Japan; 9) Dept Neurology, Tokyo Metropolitan Children's Medical Center, Fuchu, Japan; 10) Dept Neuropediatrics, Tokyo Metropolitan Neurological Hospital, Fuchu, Japan; 11) Dept Pediatrics, Jichi Medical Univ, Tochigi, Japan; 12) Dept Pediatrics, Hokkaido Univ Grad Sch Med, Sapporo, Japan; 13) Dept Pediatrics, Dokkyo Med Univ, Koshigaya Hospital, Saitama, Japan; 14) Dept Biol Sciences, Faculty of Bioscience and Biotech, Tokyo Institute of Technology, Yokohama, Japan; 15) Dept Mol Biol, Grad Sch Med, Yokohama City Univ, Yokohama, Japan; 16) Dept Pediatrics, Yamagata Univ Sch Med, Yamagata, Japan; 17) Present address: Lab of Biochem and Mol Biol, National Cancer Institute, NIH, Bethesda, USA.

A *de novo* 9q33.3-q34.11 microdeletion involving *STXBP1* has been found in one of four individuals (group A) with early onset West syndrome, severe hypomyelination, poor visual attention, and developmental delay. While haploinsufficiency of *STXBP1* was involved in early infantile epileptic encephalopathy in a previous different cohort study (group B), no mutations of *STXBP1* were found in two of the remaining three subjects of group A (one was unavailable). We assumed that another gene within the deletion might contribute to the phenotype of group A. *SPTAN1* encoding α-II spectrin, which is essential for proper myelination in zebrafish, turned out to be deleted. In two subjects, an in-frame 3-bp deletion and a 6-bp duplication in *SPTAN1* were found at the initial nucleation site of the α/β spectrin heterodimer. *SPTAN1* was further screened in six unrelated individuals with WS and hypomyelination, but no mutations were found. Recombinant mutant (mut) and wild-type (wt) α-II spectrin could assemble heterodimers with β-II spectrin, but α-II (mut)/β-II spectrin heterodimers were thermolabile compared with the α-II (wt)/β-II heterodimers. Transient expression in mouse cortical neurons revealed aggregation of α-II (mut)/β-II and α-II (mut)/β-III spectrin heterodimers, which was also observed in lymphoblastoid cells from two subjects with in-frame mutations. Clustering of ankyrinG and voltage-gated sodium channels at axon initial segment (AIS) was disturbed in relation to the aggregates, together with an elevated action potential threshold. These findings suggest that pathological aggregation of α/β spectrin heterodimers and abnormal AIS integrity due to *SPTAN1* mutations were involved in pathogenesis of infantile epilepsy.

242

Next-Generation Sequencing of a 4.1 Mb Linkage Interval Reveals FLVCR2 Deletions and Mutations in Lethal Cerebral Vasculopathy (Fowler Syndrome). *S. Thomas[1,2], F. Encha-Razavi[1,2,3], L. Devisme[4], H. Etchevers[1,2], B. Bessière[5], G. Goudefroye[3], N. Elkhartoufi[3], A. Ichkou[3], M. Bonnière[3,6], P. Marcorelle[7], P. Parent[8], S. Manouvrier[9], M. Holder[9], A. Laquerrière[10], L. Loeuillet[11], J. Roume[12], J. Martinovic[3], S. Mougou-Zerelli[1,2,13], M. Gonzales[14], P. Wookey[15], V. Meyer[16], C. Boyle Feysot[17], P. Nitschke[18], N. Leticee[19], A. Munnich[1,2,4], S. Lyonnet[1,2,4], G. Guapay[16], B. Folliguet[20], M. Vekemans[1,2,4], T. Attié-Bitach[1,2,4].* 1) INSERM U781, Paris, France; 2) Université René Descartes, Paris 5, France; 3) Département de Génétique, Hôpital Necker-Enfants Malades, AP-HP, Paris, France; 4) Pôle de Pathologie, Centre de Biologie Pathologie, Lille, France; 5) Laboratoire d'Anatomo-Foeto-Pathologie, Institut de Puériculture et de Périnatalogie, Paris, France; 6) Laboratoire Nord Pathologie, Lille, France; 7) Laboratoire d'anatomopathologie, CHRU Hôpital Morvan, Brest, France; 8) Département de pédiatrie et génétique médicale, CHRU Hôpital Morvan, Brest, France; 9) Service de Génétique Clinique, CHRU de Lille, Hôpital Jeanne de Flandre, Lille Cedex; 10) Laboratoire d'Anatomie Pathologique, Hôpital de Rouen, Rouen, France; 11) Service d'Anatomie et de Cytologie Pathologiques,CHI Poissy, Saint Germain en Laye, France; 12) Génétique Médicale, CHI Poissy, Saint Germain en Laye, France; 13) Service de Cytogénétique, Genetique moleculaire et Biologie de la reproduction, Hôpital Farhat Hached, Sousse, Tunisie; 14) Service de Génétique et d'Embryologie Médicales, Hôpital Armand Trousseau, AP-HP, et Université Pierre et Marie Curie, Paris 6, France; 15) Department of Medicine, University of Melbourne, Australia; 16) Centre National de Séquençage-Genoscope, Evry, France; 17) Plateforme de génomique, Fondation IMAGINE; 18) Service de Bioinformatique, Université Paris - Descartes; 19) Service de Gynécologie Obstétrique, Hôpital Necker Enfants Malades, Paris, France; 20) Laboratoire de Biologie de la Reproduction et du Développement, Maternité de Nancy, France.

Rare lethal disease gene identification remains a challenging issue amenable to next generation sequencing. Despite poor quality DNA derived from seven fetuses deceased in utero from cerebral proliferative glomeruloid vasculopathy (PGV or Fowler syndrome), sequence capture and high-throughput sequencing of a 4.1MB critical region mapped by homozygosity mapping unambiguously identified *FLVCR2* as the causative gene. A specific analysis tool has been developped to annotate detected variants with genomic location, percentage of variants, known SNPs in dbSNP130 or HapMap, and amino acid consequences according to the various transcripts. 3,457 variants on chromosome 14 were found to not correspond to known SNPs, and were absent from the normal control individual. After initial exclusion of non-exonic and synonymous variants, 42 variants in 29 candidate genes remained. In 20 of these genes, a single variation was found in one individual, two and three variations were found in six and two genes, respectively, and *FLVCR2* was the only gene with variations identified in four out of seven individuals. Careful examination of the FLVCR2 locus in one proband revealed a homozygous deletion of exons 2 to 10, as the absence of both nucleotide variations and reads over a 46.8 MB genomic region. Direct sequencing identified two additional mutations such that mutant *FLVCR2* alleles were identified in the 7 families studied (5 homozygotes and 2 compound heterozygotes). FLVCR2 is a member of the major facilitator superfamily (MFS) of transporter proteins that transport small molecules in response to ion gradients. Based on the cell types in which it is expressed, and MFS transport of chelated complexes of divalent metal ions, the FLVCR2 transporter had been postulated to be a gatekeeper for the controlled entry of calcium into target cell types. Immunostaining of normal and diseased tissues will bring further insight into the function of FLVCR2 within the cephalic vasculature. Furtheremore, a yeast two-hybrid assay is currently being performed to find FLVCR2 partners. This is one of the first disease-causing genes to be identified by comprehensive high throughput sequencing of an entire linkage interval, a methodology that will be useful for many other genetically homogeneous rare disorders. Finally, identification of the gene for PGV will permit accurate genetic counselling and prenatal diagnosis, in particular for the late-onset forms of the disease without arthrogryposis.

243

Mutations disrupting selenocysteine formation cause a progressive neurological phenotype. *O. Agamy[1], B. Ben Zeev[2], D. Lev[3], B. Markus[1], D. Fajersztein[1], D. Su[4], G. Narkis[1], R. Ofir[1], C. Hoffmann[2], E. Leshinsky-Silver[3], S. Sivan[1], D. Söll[4], T. Lerman-Sagie[5], O.S. Birk[1,6].* 1) The Morris Kahn Laboratory of Human Genetics, National Institute for Biotechnology in the Negev (NIBN), Ben Gurion University, Beer-Sheva, Israel; 2) Pediatric Neurology Unit, Safra Pediatric Hospital and Department of Diagnostic Imaging, Sheba Medical Center, Ramat-Gan, Israel; 3) Institute of Medical Genetics, Wolfson Medical Center, Holon, Israel; 4) Departments of Molecular Biophysics and Biochemistry, and 5Chemistry, Yale University, New Haven, CT 06520-8114; 5) Pediatric Neurology Unit, Wolfson Medical Center, Sackler School of Medicine, Tel- Aviv University, Holon, Israel; 6) Genetics Institute, Soroka University Medical Center, Beer-sheva, Israel.

The essential micronutrient selenium is found in proteins as selenocysteine (Sec), the only genetically encoded amino acid in humans whose biosynthesis occurs on its cognate transfer RNA (tRNA). In the final step of selenocysteine formation the essential enzyme O-phosphoseryl-tRNA:selenocysteinyl-tRNA synthase (SepSecS) catalyzes the conversion of O-phosphoseryl-tRNA (Sep-tRNA) to Sec-tRNA. We now demonstrate through linkage analysis studies that two different SepSecS mutations cause an autosomal recessive progressive neurological syndrome. This is the first demonstration of a human disease associated with the enzymatic pathway to the 21st amino acid, selenocysteine, which is present in at least 25 human proteins.

244

Disruption at the PTCHD1 locus on Xp22.11 results in autism spectrum disorder and intellectual disability. *A. Noor[1], A. Whibley[2], C.R. Marshall[3], AGP. Autism Genome Project Consort.[4], G.A. Rouleau[5], C.C. Hui[6], F.L. Raymond[2], S.W. Scherer[3,6], J.B. Vincent[1,7].* 1) Neurogenetics Section, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 2) Cambridge Institute of Medical Research, University of Cambridge, Cambridge, UK; 3) Program in Genetics and Genome Biology and The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) See List of authors and affiliations; 5) Center of Excellence in Neuromics, Centre Hospitalier de l'Université de Montréal, and Department of Medicine, University of Montreal, Montreal, Quebec, Canada; 6) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 7) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada.

Autism is a common neurodevelopmental disorder with a complex mode of inheritance. It is one of the most highly heritable of the complex disorders, however, the underlying genetic factors remain largely unknown. Here, we report mutations in the X-chromosomal *PTCHD1* (patched-related) gene, in seven families with autism spectrum disorder (ASD) and in three families with intellectual disability (ID). A 167 Kb microdeletion spanning exon 1 was found in two ASD brothers and a 90 Kb microdeletion spanning the entire gene was found in three males with ID in a second family. In 900 ASD and 208 ID male probands we identified seven different missense changes in eight probands, all male and inherited from unaffected mothers, and not found in controls. Two of the ASD individuals with missense changes also carried a de novo deletion at another ASD-susceptibility locus (*DPYD* and *DPP6*), suggesting complex genetic contributions. In 7 additional males with ASD, we identified deletions in the 5' flanking region of PTCHD1 disrupting a complex non-coding RNA and potential regulatory elements; equivalent changes were not found in 4,829 male control individuals ($p=1.2 \times 10^{-5}$). Systematic screening at *PTCHD1* and 5'-flanking regions, suggests involvement of this locus in ~1% of ASD and ID individuals. Furthermore, we also demonstrated that the PTCHD1 protein represses Gli-dependent transcription, and is likely to be a third receptor for Hedgehog signaling, and implicating a role for this signaling pathway in ASD and ID.

245

UBE3A regulates monoamine synthesis by increasing GTPCH1 activity via transcriptional co-activation. *L. Reiter[1], F. Ferdousy[2], W. Bodeen[1], K. Summers[1], O. Doherty[2], O. Wright[2], J. O'Donnell[2].* 1) Dept Neurology, UTHSC, Memphis, TN; 2) Dept Biology, Univ Alabama, Tuscaloosa, AL.

The underlying defect in Angelman syndrome (AS) and autism spectrum disorder (ASD) may be in part due to basic defects in synaptic plasticity and function. In some individuals serotonin reuptake inhibitors, which increase the free pools of serotonin at the synapse, can ameliorate some symptoms. Loss of *UBE3A* expression causes AS, while duplications of chromosome 15q11.2-q13 that include the *UBE3A* gene cause ASD when maternally inherited, implicating the maternally expressed *UBE3A* gene in the ASD phenotype. In a *Drosophila* screen for proteins regulated by UBE3A, we identified a key regulator of monoamine synthesis, the gene *Punch*, or GTP cyclohydrolase I/GCH1. Here we show that *Dube3a*, the fly *UBE3A* ortholog, regulates Punch/GCH1 in the brain. Over-expression of *Dube3a* elevates tetrahydrobiopterin (THB), the rate limiting cofactor in monoamine synthesis, and dopamine levels, while loss of Dube3a has the opposite effect. These fluctuations were associated with hyperactivity in flies. In addition, both wild type *Dube3a* and a ubiquitination-defective *Dube3a*-C/A form were found at high levels in nuclear fractions and appear to be poly-ubiquitinated in vivo by endogenous Dube3a. We propose that the transcriptional co-activation function of Dube3a can regulate GCH1 activity in the brain. These results provide a connection between dopamine/serotonin and Dube3a expression that may explain why some individuals with ASD or AS respond better to selective serotonin reuptake inhibitors than others.

246

Isoform-specific transgenic complementation of MeCP2 deficiency. *J. Soto[1], B. Kerr[1], M. Saez[1], K. Walz[2], J. Young[1,2].* 1) Centro de Estudios Cientificos, Valdivia, Chile; 2) John P Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, 33136, FL, USA.

Rett syndrome (RTT) is a complex disorder that affects communication, movement and behavior. RTT is caused by mutations in MECP2, an X chromosome gene encoding for MeCP2, a protein that regulates gene expression. MECP2 generates two alternative splice variants encoding two protein isoforms that differ only in the N-terminus. Although the two variants show regional and age-related differences in transcript abundance, no functional differences have been identified for these splice variants so far. However, it has been suggested that the RTT phenotype may occur in the presence of a functional MeCP2-e2 protein, since MeCP2-e1 specific mutations have been identified in patients diagnosed with RTT. This suggests that the two isoforms might be functionally distinct and that endogenous MeCP2-e2 is not able to compensate for the lack of MeCP2-e1. On the other hand, it was shown that expression of MeCP2-e2 was sufficient to prevent the development of disease signs in mice lacking endogenous Mecp2 expression. Here, we show that transgenic mice expressing either the MeCP2-e1 or MeCP2-e2 splice variant on an otherwise Mecp2 null background are phenotypically indistinguishable from wild-type mice. Thus, our data show for the first time that the sole expression of MeCP2-e1 was able to compensate for the lack of MeCP2 in mice. Life span, as well as physical and neurobehavioral phenotypes of the Mecp2-/y mice were normalized by expressing transgenic MeCP2-e1 in the brain. Notably, the approximate expression of MeCP2-e1 in the transgenic line that showed a significant normalization of the Rett-like phenotypes presented by Mecp2-/y mice was only 40-60% of endogenous, suggesting that equaling endogenous level of expression is not strictly necessary for a significant phenotypic rescue. Our results indicate that the two MeCP2 splice variants can substitute for each other and fulfill the basic functions of MeCP2 in the brain. Our demonstration that expression of MeCP2-e1 by itself was sufficient to significantly extend the life span of Mecp2-/y mice, to prevent the development of common manifestations of neurological dysfunction in mice such as clasping, and to normalize anxiety-related and motor phenotypes is relevant to the design of efficient strategies aiming to restore MeCP2 activity, including gene therapy and protein administration approaches.

247

Genome-wide linkage study of mammographic density, a heritable and strong risk factor for breast cancer. *C.M.T. Greenwood[1,2], A.D. Paterson[1,2], L. Linton[3], I.L. Andrulis[4,5], C. Apicella[6], A. Dimitromanolakis[1], M.C. Southey[6], E. Satariano[7], V. Kriukov[3], L.J. Martin[3], A. Salleh[3], E. Samiltchuk[1], E.M. John[7,8], J.L. Hopper[6], N.F. Boyd[3], J.M. Rommens[1,5].* 1) Program in Genetics & Genome Biology, The Hospital for Sick Children, 101 College Street, East Tower, Toronto, ON CANADA M5G 1L7; 2) Dalla Lana School of Public Health, University of Toronto, Toronto, ON; 3) Campbell Family Insitute for Breast Cancer Research, Ontario Cancer Institute, Toronto, ON; 4) Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON; 5) Department of Molecular Genetics, University of Toronto, Toronto, ON; 6) The University of Melbourne, Melbourne, Australia; 7) Cancer Prevention Institute of California, Fremont, CA; 8) Stanford Cancer Center, Stanford, CA.

Mammographic density (MD), determined by mammography, is known to be highly heritable (h^2~0.6) and a strong risk factor for breast cancer. To identify loci influencing variation in MD we performed a genome-wide linkage study. Epidemiological and genotype data were assembled for 1415 families from several sources, including the Australia, California and Ontario Breast Cancer Family Registries, twin and sisters studies in Australia and Ontario, and families being studied for MD and breast cancer risk. Families known to carry BRCA1/2 mutations were excluded. Most families consisted of one pair of sisters; 75 families contained more than 4 women with all relevant data. Information on factors known to influence MD was collected by questionnaire. Mammograms were obtained to digitize the cranio-caudal view, and MD was measured by a single reader using a computer-assisted thresholding method. MD measurement reliability was over 90%. Pre-diagnostic mammograms of the contra-lateral breast were used for women who had a previous diagnosis of breast cancer. Mammograms from pairs of sisters were mostly taken within five years of age of each other. Genotyping was performed by the Center for Inherited Disease Research using the Illumina Infinium II Human Linkage-12 panel. Principal component analysis showed that 92.6% of the families were Caucasian. Using a variance components method, quantitative trait linkage analysis was performed by adjusting the MD measurements for measured covariates. Our primary trait was formed by fitting a linear model to the square root of percent MD, adjusting for age at mammogram, number of live births, menopausal status, weight, height, weight squared and previous or current use of hormonal replacement therapy. The maximum LOD score from the genome-wide scan was 2.69 on chromosome 7p (at 46.5Mb, NCBI Build 36) with a 1 LOD interval spanning 8.6cM. A suggestive score (LOD=2.44) was also found on chromosome 17. We did not find supporting evidence for a locus on chromosome 5p, as previously reported by Vachon *et al.* in 2007. Simulations using 1678 sister pairs showed that this sample had 72% power to detect LOD scores of 3 or greater for a locus accounting for 25% of phenotypic variance. Further investigations to strengthen evidence for linkage and detect genetic association are underway to identify genes that determine MD variation with the long term goal of understanding breast biology and susceptibility to breast cancer.

248

Genotype at Low Penetrance Breast Cancer Risk Loci - What are the Implications for Individual Risk Estimation? *J. Berg[1,2], A. Martin[1], S. White[1], J. Dunlop[2], A. Onen[1], J. Gale[1], P. Quinlan[3], L. Baker[3], A. Thompson[3], P. Pharoah[4].* 1) Human Genetics, Centre for Oncology and Molecular Medicine, Division of Medical Sciences, University of Dundee, Dundee, UK; 2) East of Scotland Clinical Genetics Service, Ninewells Hospital and Medical School, Dundee, UK; 3) Surgery and Oncology, Centre for Oncology and Molecular Medicine, Division of Medical Sciences, University of Dundee, Dundee, UK; 4) Departments of Oncology and Public Health and Primary Care, University of Cambridge, Cambridge, UK.

A significant proportion of women are at increased genetic risk of breast cancer. Mutations in high penetrance genes such as BRCA1 and BRCA2 are only responsible in the minority of cases, with low penetrance polymorphisms in other genes expected to account for the majority of the remaining genetic risk. An increasing number of such low penetrance polymorphisms are being identified, but each polymorphism only contributes a small amount to overall risk. Currently, in clinical practice, women at increased risk of breast cancer are identified by their family history, and the role of genetic testing for multifactorial risk remains uncertain. We have taken the population frequency and genotype relative risk information for the 13 most established breast cancer risk loci. Genotyping at these 13 loci could provide significant risk information for an individual. The top 1% of women in the genotype risk distribution would have a risk of breast cancer of 2.4 times the general population. At this level of risk, they would qualify for breast cancer screening from age 40 according to evidence based guidelines issued by the UK National Institute of Clinical Excellence (NICE). In addition, the top 5% of the population are at 1.7 times risk of breast cancer and would have the same risk at age 40 as a 50 year old at population risk who would qualify for breast screening according to UK and US National Screening Guidelines. To investigate whether low penetrance genotype has greater potential if combined with other risk factors, we used a simple multiplicative model to combine family history risk of cancer derived using BOADICEA with genotype information. This suggests that genotype would result in a significant reclassification of individual risk. For example, 20% of women who had a sister affected with breast cancer at 55 would qualify for additional screening under NICE criteria if genotype were taken into account. Taking a similar approach with 160 complex family histories from the Tayside family history breast clinic, we have shown that genotyping could result in reclassification and change of management for 19.4% of women being assessed in this clinic, with 12.6% of women moving into a higher risk category, and 6.8% of women moving into a lower category. These data suggest that genotyping for low penetrance breast cancer risk loci is clinically relevant, and that it will be more powerful if it can be combined with other established risk factors.

249

A PALB2 mutation associated with high risk of breast cancer. *M.C. Southey[1], Z.L. Teo[1], J.G. Dowty[2], F.A. Odefrey[1], D.J. Park[1], M. Tischkowitz[3,4], N. Sabbaghian[3], G.B. Byrnes[5], C. Apicella[2], I.M. Winship[6,7], L. Baglietto[8], G.E. Giles[8], D.E. Goldgar[9], W.D. Foulkes[3,4], J.L. Hopper[2], Breast Cancer Family Registry, kConFab.* 1) Department of Pathology, University of Melbourne, Victoria, Australia; 2) Center for MEGA Epidemiology, The University of Melbourne, Victoria, Australia; 3) Program in Cancer Genetics, Departments of Oncology and Human Genetics, McGill University, Montreal, Canada; 4) Segal Cancer Center, Lady Davis Institute, Jewish General Hospital, Montreal, Canada; 5) International Agency for Research on Cancer, Lyon, France; 6) Department of Medicine, The University of Melbourne, Victoria, Australia; 7) Royal Melbourne Hospital, Victoria, Australia; 8) Cancer Epidemiology Centre, The Cancer Council Victoria, Australia; 9) University of Utah School of Medicine, Salt Lake City, USA.

As a group, women who carry germline mutations in PALB2 are at increased risk of breast cancer. However, little is known about the size of this risk, or whether risk differs by mutation or family history, due to the paucity of studies of cases unselected for family history. We screened a population-based, age-at-onset-stratified sample of Australian women with invasive breast cancer for PALB2 mutations. We found five independent case probands with the protein truncating mutation PALB2 c.3113G>A (W1038X); two of 695 diagnosed before age 40 years and three of 708 diagnosed when aged 40-59 years. The two early-onset carrier case probands both had very strong family histories of breast cancer. Further testing found that the mutation segregated with disease in these families. No carriers were found in 764 population-based unaffected controls. The age-specific risk of breast cancer was estimated from the cancer histories of first- and second-degree relatives of the five mutation carrying probands using a modified segregation analysis that included a polygenic modifier, and conditioned on the carrier case proband. The hazard ratio was estimated to be 30.1 (95% CI 7.5 - 120; $p < 0.0001$) and the corresponding cumulative risk estimates were 49% (95% CI 15 - 93) to age 50 and 91% (95% CI 44 - 100) to age 70. We also screened 779 Australasian multiple-case breast cancer families ascertained through cancer genetics services and found a further eight families carrying this mutation. The PALB2 c.3113G>A mutation, while rare, appears to be associated with substantial risks of breast cancer that are of clinical relevance.

250

Rare, Evolutionarily Unlikely Missense Substitutions in CHEK2 Contribute to Breast Cancer Susceptibility: Results from the Breast Cancer Family Registry. *F. Lesueur[1], F. Le Calvez-Kelm[1], F. Damiola[1], F. Guénard[2], A. Thomas[3], G.B. Byrnes[1], F. Durocher[2], J.L. Hopper[4], M.C. Southey[5], I.L. Andrulis[6], E.M. John[7], S.V. Tavtigian[8].* 1) International Agency for Research on Cancer, Lyon, France; 2) Dept of Molecular Medicine, CHUQ Research Center, CHUL, Laval University, Quebec, Canada; 3) Dept of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT, USA; 4) Centre MEGA Epidemiology, University of Melbourne, Carlton, Victoria, Australia; 5) Dept of Pathology, University of Melbourne, Carlton, Victoria, Australia; 6) Cancer Care Ontario, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Dept of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 7) Cancer Prevention Institute of California, Fremont, CA, USA, and Stanford Cancer Center, Stanford, CA, USA; 8) Dept of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT, USA.

CHEK2 is a breast cancer susceptibility gene. While both protein truncating variants and some missense substitutions are known to be pathogenic, no large-scale study has used full open reading frame mutation screening to assess the contribution of rare missense substitutions in CHEK2 to risk of breast cancer. We have applied a previously used method for an in silico assessment of risk for unclassified variants to data from CHEK2 mutation screening data of 1,320 cases and 1,109 controls from the Breast Cancer Family Registry and 95 familial cases and 95 controls from the INHERIT BRCAs study. We stratified rare missense substitutions into a series of grades a priori ordered from least to most likely to be evolutionarily deleterious, using the program Align-GVGD, and then used a logistic regression test for trend to compare the frequency distributions of the graded missense substitutions in cases versus controls. We found evidence of increased risk associated with rare, evolutionarily unlikely CHEK2 missense substitutions. Three additional findings were (1) the risk estimate for the most severe grade of CHEK2 missense substitutions (denoted C65) was approximately equivalent to that of CHEK2 protein truncating variants; (2) the population attributable fraction for rare missense substitutions, if causal, was about the same as for protein truncating variants; and (3) post-hoc power calculations implied that scaling case-control mutation screening up to examine entire biochemical pathways would require mutation screening roughly 2,000 cases and a similar number of controls to achieve acceptable statistical power. We conclude that CHEK2 harbors many rare, pathogenic sequence variants, a substantial proportion of which are missense substitutions. Moreover, this study replicates the success, first demonstrated from ATM data, of our method for analysis of case-control mutation screening data.

251

Epistatic interaction between *BRCA1, BRCA2* and intermediate-penetrance breast cancer predisposition genes. *C. Turnbull[1], S. Seal[1], A. Renwick[1], D. Hughes[1], D. Pernet[1], A. Elliott[1], M. Warren-Perry[1], D.G. Evans[2], D. Eccles[3], M.R. Stratton[4], N. Rahman[1], Breast Cancer Susceptibility Collaboration (UK).* 1) Section of Cancer Genetics, Insitute of Cancer Research, Sutton, United Kingdom; 2) Regional Genetic Service. St Mary's Hospital, Manchester, United Kingdom; 3) Wessex Clinical Genetics Service, Princess Ann Hospital, Southampton, United Kingdom; 4) Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom.

Mutations in *BRCA1* and *BRCA2* confer a high risk of breast cancer (relative risk (RR)>10), have formed the basis of our understanding of genetic predisposition to breast cancer and are used widely for clinical genetic testing. Recent genome-wide association studies have identified common 'low penetrance alleles' which confer risk of breast cancer not only in the general population (RR 1.06-1.26) but also in the presence of a mutation in *BRCA1* and/or *BRCA2*. We have identified four further breast cancer predisposition genes, *CHEK2, ATM, BRIP1* and *PALB2* which encode proteins that physically interact with *BRCA1* and/or *BRCA2* in DNA repair pathways. Like *BRCA1* and *BRCA2*, multiple loss-of-function mutations arise in these genes but the breast cancer risks conferred are more modest (RR 2-4). We sought to explore the statistical interplay of mutations in these 'intermediate penetrance genes' with those in *BRCA1* and *BRCA2*. We screened genomic DNA through *BRCA1* and *BRCA2* and genotyped the *CHEK2* 1100delC mutation in 5708 index female cases of familial breast cancer. We additionally screened 1448 of these cases through the coding exons and intron-exons boundaries of *ATM* in 71 fragments. We observed a significant deficiency of *CHEK2* 1100delC ($P=5x10^{-4}$) and mutations in ATM ($P=6x10^{-4}$) in *BRCA*-positive individuals compared to *BRCA*-negative individuals, consistent with a model of negative epistasis between mutations in *CHEK2* and *ATM* and mutations in *BRCA1* and *BRCA2*. This would imply that neither a mutation in *ATM* nor the *CHEK2* 1100delC mutation has significant additional impact in the presence of a mutation in *BRCA1* or *BRCA2* and thus does not modify risk in *BRCA* mutation carriers. This contrasts with the multiplicative combination of risk observed between *BRCA1/BRCA2* mutations and the common low penetrance alleles. This result is biologically informative, suggesting that whilst the common alleles may function in pathways essentially independent of *BRCA1* or *BRCA2*, abrogation of the function of ATM or CHK2 may have little additional impact on a common pathway already radically subverted by a mutation in *BRCA1* or *BRCA2*. This observation of negative epistasis is also clinically important, demonstrating that the statistical interplay of breast cancer predisposition factors is heterogenous, that genetic risk factors can not be assumed independent and that caution must be exercised in risk estimation utilising new genetic risk factors in combination with *BRCA1* and *BRCA2*.

252

Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *T. Walsh[1], M.K. Lee[1], S. Casadei[1], A.M. Thornton[1], A.S. Nord[2], S.M. Stray[1], C.C. Pennil[3], J.B. Mandell[1], E.M. Swisher[3], M-C. King[1,2].* 1) Medicine (Medical Genetics), University of Washington, Seattle, WA; 2) Genome Sciences, University of Washington, Seattle, WA; 3) Obstetrics and Gynecology, University of Washington, Seattle, WA.

Inherited loss-of-function mutations in the tumor suppressor genes *BRCA1*, *BRCA2*, and multiple other genes, predispose to high risks of breast and/or ovarian cancer. All cancer-associated inherited mutations are individually rare but collectively account for a significant fraction of familial breast cancer. Genetic testing for *BRCA1* and *BRCA2* mutations has become an integral part of clinical practice, but testing is generally limited to these two genes and to women with severe family histories of breast or ovarian cancer. To determine whether massively parallel, "next generation" sequencing would enable accurate, thorough, and cost-effective identification of inherited mutations for breast and ovarian cancer, we developed a genomic assay to capture, sequence, and detect all mutations in 22 breast and ovarian cancer genes. Germline genomic DNA from 20 subjects with known inherited mutations, ranging in size from 1 to 19 basepairs for nucleotide substitutions and insertion and deletion mutations, and between 1 and 100 kilobasepairs for genomic deletions and duplications, was hybridized to custom cRNA oligonucleotides, then sequenced using an genome analyzer. Analysis was carried out blind to the mutation in each sample. Average coverage was >1200 reads per basepair. After filtering sequences for quality and number of reads, all single nucleotide substitutions, small insertion and deletion mutations, and large genomic duplications and deletions were detected. There were zero false positive calls of nonsense mutations, frameshift mutations, or genomic rearrangements for any gene in any of the test samples. We are currently evaluating the feasibility of using multiplexed, bar-coded samples to increase throughput and reduce cost of this approach. This approach enables widespread genetic testing and personalized risk assessment for breast and ovarian cancer.

253

Validation and implementation of the 454-NGS platform for the molecular analysis of the breast cancer genes BRCA1/2. *G. Michils, S. Hollants, J. Van Houdt, L. Vliegen, V. Brys, J. Vermeesch, H. Cuppens, G. Matthijs.* Centre for Human Genetics, University of Leuven, Leuven, Belgium.

We aim at developing diagnostic applications on the Roche 454-GS FLX Titanium sequencer with a focus on clinical use and accreditation. Hence, we dealt with quality aspects such as accuracy, sensitivity (coverage, variant frequency), efficiency (quick import and analysis, limited number of false positives) and reporting (mutation identification, layout). Besides this, special attention was given to the detection of mutations in or near homopolymer stretches. Overall, the method has been validated by testing DNA samples of breast cancer patients that have been previously sequenced with the Sanger method. First, amplicons spanning the coding region of the BRCA1 gene were amplified using different DNA samples with known variations, to generate an (artificial) control sample that contained in total 40 variations, including 16 indels. These data files were analyzed with four commercial software tools: CLC Genomics Workbench (CLC bio), Sequence Pilot (SeqNext module, JSI medical systems), GS Amplicon Variant Analyzer (AVA, Roche) and NextGENe (Softgenetics). Secondly, the whole coding regions of the BRCA1/2 genes were interrogated in patient samples, after amplification using a multiplex assay (Multiplicon). This was performed to generate figures for the sensitivity and reproducibility of the method. Two software programs, SeqNext and NextGENe, could detect all heterozygous variations. The other tools missed individual variants. The variant frequencies (% mutant versus wild type reads) for heterozygous variations varied around 46 ± 8 (range: 19-61). The multiplex assay is robust for DNA from different sources and all of the 169 amplicons designed in the assay were represented. Over 93% reached a minimum coverage of 30x and only 0.08% failed. There was also intra- and inter coverage variability among the samples and the amplicons, showing that the pooling is crucial. This might also be explained by stochastic amplification events during the emulsion PCR or by pipetting errors. We believe that the software tools still have to undergo further developments and can only be implemented with circumspection. Clearly, all parameters have to be monitored to avoid variation (i.e. automatized PCR preparation). This is necessary to reach the maximal capacity of the sequencer and warrant the validity of the results.

254

In vivo genetic abnormalities in human breast cancer stem/progenitor cells. *A.L. Hernandez, R.F. Pommier, E. Han, K. Massimino, P. Muller, B.S. Diggs, J. Murphy, J. Hansen, A. Naik, J. Vetto, S.J. Pommier.* Surgery, Oregon Health & Science University, Portland, OR.

Objective: While it is believed that the clinical course of breast cancer is directed by breast cancer stem cells, their origin is unknown. We hypothesize that when critical genetic mutations occur in normal breast stem/progenitor cells, these cells transform into cancer stem cells. To test this phenomenon in vivo, normal and malignant stem/progenitor cells were collected from fresh surgical specimens. The goal of this study was to determine if oncogene mutations and stem-cell related gene expression are found in breast cancer stem/progenitor cells and not found in normal breast stem/progenitor cells. Methods: Seven Stage 2 intraductal breast cancers and 8 normal breast samples from reduction mammoplasty were examined. Stem/progenitor cells were collected from fresh surgical specimens. Cells were labeled with fluorochrome-conjugated monoclonal antibodies against human CD45 and CD31(FITC), and CD24(PE), CD49f(PE-Cy 5). Isotype controls included FITC, PE and PE-Cy 5. The various cell subpopulations were separated based on surface antibody labeling, and collected by discriminatory gating. CD31+, CD45+ endothelial cells and leukocytes were removed. Cells were sorted into 4 different populations: CD49f+or- CD24+or- . Cells were tested for 384 mutations in 26 oncogenes, and gene expression in 96 stem cell genes. Genetic findings were compared between normal and cancer stem cells using Students t-tests. Results: Few significant gene expression differences were observed between cancer and normal stem/progenitor cells. Significant differences were observed between the four stem/progenitor populations. CD49f- CD24+ malignant and normal cells over-expressed DNMT3B. CD49f+24- over-expressed POU5F1 and CD49+CD24+ SOX2. No mutations were found in normal stem/progenitor cells. However, a single mutation in AKT1, HRAS or PI3KCA genes was present in 6 of 7 tumors, exclusively in CD49f+CD24+ and CD49f- CD24+ cells. Specific exon mutations correlated with clinical severity. The node negative tumor had no stem/progenitor cell mutations. Conclusions: Gene expression similarities between normal and cancer stem cells suggest a common origin. Oncogene mutations in stem/progenitor cells will produce malignant behavior and retain stem cell potential. Five of the 7 mutations were activating PIK3CA mutations that can simultaneously exert extensive damage on multiple cellular functions. These genetic alterations may be sufficient to initiate and sustain tumor development.

255

LOW TMPRSS6 (MATRIPTASE-2) GENE EXPRESSION ASSOCIATES WITH TRIPLE NEGATIVE BREAST CANCER. *A. Mannermaa[1,2,3], T. Nykopp[1,3], J. Kauppinen[1,3], M. Eskelinen[4], V. Kataja[5,6], V.-M. Kosma[1,2,3].* 1) Dept Pathology & Forensic Med, Univ Eastern Finland, Kuopio, Finland; 2) Department of Pathology, Kuopio University Hospital, Kuopio, Finland; 3) Biocenter Kuopio, Kuopio, Finland; 4) Department of Surgery, Kuopio University Hospital, Kuopio, Finland; 5) Department of Oncology, Kuopio University Hospital, Kuopio, Finland; 6) Department of Oncology, Vaasa Central Hospital, Vaasa, Finland.

TMPRSS6 encodes for matriptase-2, a member of the type II transmembrane serine protease family. It has been implicated in breast cancer risk and tumor progression. We have previously shown an association with a SNP rs733655 in matriptase-2 gene (TMPRSS6) with breast cancer. Here we report a TMPRSS6 gene expression study in an Eastern Finnish population sample of 41 breast cancer cases. Primary breast tumor tissue samples were obtained during the initial cancer surgery, cooled in liquid isopentane and liquid nitrogen and stored in -70°C. RNA was extracted by using Trizol and cDNA was synthesized from 2.5 µg of total RNA using High-Capacity cDNA Archive kit (Applied Biosystems). Real-time gene expression analysis was performed using TaqMan® Gene Expression Assays (Applied Biosystems) according to manufacturer's instructions. HPRT1 (hypoxanthine phosphoribosyltransferase 1) was used as an endogenous control. Estrogen receptor (ER), progesterone receptor (PR), cytokeratine (CK5/6) and epidermal growth factor receptor (EGFR) expression were evaluated by immunohistochemical staining from paraffin embedded tissue samples. Statistical analyses were carried out using SPSS 17 for Windows. Differences between groups were analyzed by either non-parametric Kruskal-Wallis test or by non-parametric Mann - Whitney U-test. The relative mean TMPRSS6 gene expression was significantly different in grade I, II and III tumors (p=0.004) with the highest mean expression in grade I and the lowest in grade III tumors. There were no differences in expression between the genotypes of rs733655. Low expression of TMPRss6 was associated with negative expression of ER (p=0.0001) and PR (p=0.011) and with positive HER2 expression (p=0.021. Triple-negative tumors, classified as ER-, PR- and HER2- had significantly lower matriptase-2 gene expression (p=0.002) than non triple-negative tumors. No clinically useful target molecule has been identified for triple-negative breast cancer. Our results provide further evidence for TMPRSS6 as a risk factor in breast cancer and urges for additional studies on the role of matriptase gene family in breast cancer.

256

Protein Tyrosine Phosphatase Receptor - type Kappa (PTPRK) gene induces growth arrest and senescence in Breast and Ovarian tumor cells. *R. Athwal, G. Kaur, V. Zhawar.* Fels Inst Cancer Res, Temple Univ Sch Med, Philadelphia, PA.

We have determined that Ectopic expression of PTPRK gene restores normal growth pattern and senescence in breast and ovarian tumor cells. Three chromosomal regions 6q16-21, 6q22.3-23.1 and 6q25-27, have been described to carry tumor suppressor genes for breast and ovarian cancers. Microcell transfer of a normal chromosome 6 or 6q inhibits cell growth and restores senescence in a variety tumor cell lines (Sandhu et al. 1994, PNAS 91: 5498-5502; Sandhu et al. 1996, Oncogene 12: 247-252). Applying a deletion strategy, following the transfer of chromosome 6 or 6q into immortal breast tumor and SV40 immortalized cells, we mapped a candidate cell senescence locus, within 1cM genetic interval, at 6q22.3. Human genome database search, for cDNAs mapped within this locus, identified PTPRK (Protein-Tyrosine Phosphatase Receptor Type K) as candidate cell senescence/tumor suppressor gene. Following initial identification, we recovered a 5.6 Kb full length PTPRK cDNA by RT-PCR, and characterized for genomic structure and regulatory regions. These studies revealed the expression of two different transcript variants for PTPRK gene in tumor v/s normal cells. Open reading frame for PTPRK gene was cloned in a retroviral expression vector, in frame with EGFP tag, where EGFP is added after the signal peptide. The ectopic expression of cloned PTPRK gene, in breast and ovarian tumor cells, led to complete growth arrest and senescence. Gene transfer clones that multiplied indefinitely contained a truncated form of introduced PTPRK gene, confirming that intact gene indeed induces senescence in tumor cells. These results suggest that PTPRK is a cell senescence/tumor suppressor gene that may play a role in the etiology of breast and ovarian cancer. PTPRK is a member of a large family of genes for Protein Tyrosine Phosphates (PTPs). The interplay of Protein Tyrosine Kinases (PTKs) and PTPs in phosphorylation and dephosphorylation of proteins, involved in the cascade of signaling circuitry, govern physiological functions involved in cell growth and differentiation. We hypothesize that PTPRK gene is either mutated or abnormally expressed in tumor cells, and Loss or abnormal expression of PTPRK gene contributes to the neoplastic cell growth and tumor development. Future studies are designed to delineate the signaling pathways responsive to the expression of PTPRK gene.

257

Diversity and Recombination Rate Variation in Western Chimpanzees. *A. Auton[1], A. Fledel-Alon[3], R. Hernandez[3,4], P. Humburg[1], Z. Iqbal[1], E. Leffler[3], G. Lunter[1], J. Maller[1], S.C. Melton[3], S. Pfeifer[5], L. Segurel[3], O. Venn[1], R. Bontrop[2], R. Bowden[1], S. Myers[5], P. Donnelly[1], M. Przeworski[3], G. McVean[1,5].* 1) WTCHG, Oxford, United Kingdom; 2) Biomedical Primate Research Centre, Netherlands; 3) Dept. of Human Genetics & Dept. of Ecology and Evolution, University of Chicago, IL, USA; 4) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, USA; 5) Department of Statistics, University of Oxford, UK.

Despite the evolutionary significance of chimpanzees, relatively little is known about both diversity levels and recombination rates in our closest evolutionary relative. For this reason, we have sequenced 10 Western chimpanzees (Pan troglodytes verus) to 8-10X coverage using paired end sequencing with 50-bp reads, and have identified over 3 million novel SNPs. This data provides insight into genome-wide patterns of chimpanzee diversity, and allows us to construct a fine-scale genetic map that reveals a recombination landscape dominated by hotspots, virtually none of which are shared between chimpanzees and humans. This lack of sharing is indicative of the rapid evolution of recombination hotspots. By comparing hotspots and coldspots, we investigate sequence features and motifs associated with chimpanzee hotspots. In contrast to the fine scale recombination rates, we observe striking correlation in human and chimpanzee recombination rates at broad scales. A clear exception to this pattern is the region around the chromosome 2 fusion site - a region in which the telomeric regions of chromosomes 2a and 2b in chimpanzee correspond to a section of the long arm of human chromosome 2. In this region, we observe that human and chimpanzee recombination rates strongly diverge. As such, this is an ideal region to investigate the effect that recombination has on patterns of diversity, the rate of change of the recombination landscape, and the driving mechanisms.

258

De novo assembly and evolutionary analyses of liver-expressed genes in 16 mammal species. *P. Melsted[1], J. Marioni[1], G.H. Perry[1], Y. Wang[1], K. Michelini[2], M. Stephens[1,3], J.K. Pritchard[1,2], Y. Gilad[1].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Howard Hughes Medical Institute, Chicago, IL; 3) Department of Statistics, University of Chicago, Chicago, IL.

Changes in gene regulation have been proposed as critical in the evolution of phenotypic diversity of primates. However, the lack of high-quality reference genomes for most species and the limited number of independently derived transcripts for non-human primates has made it difficult to study gene regulation across multiple primates. To overcome these problems, we used massively-parallel sequencing to interrogate mRNA samples extracted from the livers of 16 species (12 primates including human, and 4 non-primate out groups; 4 samples per species). Of the 12 primate species, 8 do not have currently available reference genome sequences (vervet, galago, slow loris, and 5 lemur species), which means that we had to assemble the transcriptomes de novo for these species. This study design results in nucleotide sequence, quantitative expression, and gene structure data from thousands of genes, providing insight into gene regulation and sequence evolution across a broad spectrum of primate species. Our analysis has revealed sets of genes whose expression level patterns are consistent with the action of natural selection along individual or ancestral primate lineages. We also investigated the relationship between alternative splicing within species and complete exon gains and losses between species. Interestingly, we also identified a large number of genes, highly conserved at the sequence level, that are expressed in the livers of some taxa, but that are unexpressed in others. Further, we have identified genes whose function and expression pattern we hypothesize may underlie specific adaptive processes. For example, SDR16C5 and AKR1B10, which are involved in retinol metabolism, are found to be highly expressed in marmosets relative to other species. Marmosets have a number of striking craniofacial adaptations, which allow them to gouge holes in trees, through which they can feed extensively on exuded latexes, saps, and gums. Retinol is one of the major derivatives of isoprene, the monomer of latex. It is believed that the initial stages of latex digestion are facilitated by bacteria in the large intestine, in which case, large quantities of retinol may be absorbed through the intestinal wall and filtered by the liver. Therefore these increases in expression may reflect adaptations in the marmoset lineage to their latex-rich diets.

259

Extensive selection for translation efficiency on silent mutations in recent human history. *Y. Waldman[1], T. Tuller[2,3], A. Keinan[4], E. Ruppin[1,5].* 1) Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel; 2) Department of Molecular Genetics; 3) Faculty of Mathematics and Computer science, Weizmann Institute of Science, Rehovot 76100, Israel; 4) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 5) School of Medicine, Tel Aviv University, Tel Aviv, Israel.

Synonymous substitutions are considered to be 'silent' as they do not affect protein sequence. Thus, they are often taken as a measure for neutral evolution rate. However, different silent codons have different translation efficiency (TE), raising the question to what extent such substitutions are really `silent'. In the context of human evolution, this question has been previously examined limitedly on a relatively small scale (Comeron, PNAS 2006). Here we perform the first genome-wide study on TE selection in recent human history, surveying more than 13,000 silent SNPs in 1198 unrelated individuals from 11 populations in HapMap3. We contrast SNPs that are predicted to have a high effect on TE with SNPs of low effect. As these groups differ only in their TE effect, any difference between the two points to TE selection and is insensitive to ascertainment bias. Taking F_{ST} as a measure for genetic differentiation between populations our main results are: (1) We find an enrichment of very low F_{ST} for SNPs with high TE ($P = 1.47\times10^{-11}$), which points to a marked negative TE selection. Selection is stronger in genes with higher expression levels and in genes whose products interact with a relatively large number of proteins. Similarly, selection is stronger in complex members, where changes in protein levels may imbalance complex formation. Moreover, selection is stronger in functional regions, presumably due to the selective pressure against translation errors. (2) We find an enrichment of very high F_{ST} for SNPs with high TE ($P = 7.99\times10^{-7}$) implying for positive TE selection. Selection is stronger in genes whose products interact with a small number of proteins. Interestingly, essential genes are enriched for negative selection on TE and, in parallel, underrepresented for positive selection. (3) Notably, for both positive and negative selection, the likelihood of a silent SNP to be targeted by selection is correlated with its quantitative TE effect. (4) TE selection, both positive and negative, is stronger near gene start, in accordance with recent results on universal TE profiles in various organisms (Tuller et al., Cell 2010). These results were also replicated when considering allele frequency-based tests of selection (such as Tajima's D) instead of F_{ST}. Taken together, they demonstrate the significant role of TE in recent human history and hence underscore the importance of considering silent SNPs in future association studies of complex human diseases.

260

Recent purifying selection in cis regulatory variants in human populations. *T. Lappalainen[1], S.B. Montgomery[1], A.S. Dimas[1,2], A.C. Nica[1,3], E. Migliavacca[1], B.S. Stranger[4], E.T. Dermitzakis[1].* 1) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 3) Wellcome Trust Sanger Institute, Cambridge, UK; 4) Harvard Medical School/Brigham and Women's Hospital, Boston, MA.

Changes in gene regulation constitute an important mechanism of evolution and are likely to contribute to phenotypic diversity in human populations. In this study, we investigated the patterns of natural selection behind of regulatory variation in human populations. We focused on expression quantitative trait loci (eQTLs) discovered in a European (CEU) and an African population (YRI) of the 1000 genomes resequencing data and gene expression array data from LCLs. These data provide an access to an unparalleled number of genetic variants without the ascertainment bias in SNP selection to arrays. Additionally, we covered a wide spectrum of regulatory variation in different populations and tissues by using SNP and expression array based eQTL data from LCLs of eight HapMap3 populations, and from fibroblasts, T-cells and B-cells of a single population. Overall, several lines of evidence suggest that a larger proportion of regulatory variation in Europe has negative fitness effects than in Africa, which is consistent with the hypothesis of natural selection being less efficient in Europe due to past population bottlenecks. Eurasian but not African eQTLs show a significant deficiency of high allele frequency differences in both HapMap3 and 1000g data, suggesting that purifying selection often keeps eQTLs from differentiating by random genetic drift. The fold change of eQTLs is negatively correlated with derived allele frequency in both CEU and YRI, which can be a sign of selection against regulatory variants with major effects. However, a relative enrichment of eQTLs of low derived allele frequencies in the CEU, especially of those with recessive effects, indicates smaller intensity of purifying selection in Europe compared to Africa. This is supported by European eQTLs having significantly higher fold changes, and residing in evolutionarily more conserved sites. The patterns now observed for regulatory variants are recapitulating those of nonsynonymous SNPs both in the 1000 genomes data and in earlier studies. We are continuing to investigate the importance of the enrichment of deleterious variation outside Africa for susceptibility to both rare and common disease. In conclusion, the results indicate that while there are several well-supported cases of regulatory variation underlying evolutionary adaptations in human populations, also purifying selection plays a major role in shaping the genetic diversity that affects gene expression.

261

Adaptive evolution of human glycophorin loci in malaria endemic African populations. *W.-Y. Ko, K. Kaercher, S.A. Tishkoff.* Departments of Genetics and Biology, School of Medicine and School of Arts and Sciences, University of Pennsylvania, Philadelphia PA.

Over 90% of one million deaths per year due to malaria occur in sub-Saharan Africa, a region with high levels of genetic variation and population substructure. However, little is known about nucleotide variation at genetic loci that are relevant to malaria-susceptibility across geographically and genetically diverse ethnic groups in Africa. Invasion of erythrocytes by Plasmodium falciparum parasites is central to the pathology of malaria. Glycophorin A and B are two of the major receptors expressed on the surface of red blood cells that interact with parasite ligands. Amino acid variants at these loci determine MN and Ss blood types. Here we analyze nucleotide diversity at three highly homologous glycophorins, glycophorin A (GYPA), B (GYPB), and E (GYPE) from 15 geographically and ethnically diverse African populations that experience different levels of malaria exposure. We observed extraordinarily high levels of genetic variation and gene conversion at these genes and different types of natural selection (balancing and positive selection) acting upon different parts of the extracellular domain of GYPA. Populations with higher levels of malaria exposure show the strongest signals of balancing selection. Furthermore, we have identified a novel haplotype that carries three nonsynonymous SNPs at GYPB exon 2 in five populations that appears to be a target of adaptive evolution. We also demonstrate the important role of gene conversion as a genetic mechanism for creating novel haplotypes upon which natural selection can act.

262

Sequencing of fifty human exomes reveals adaptation to high altitude. *E. Huerta-Sanchez[1,3], X. Yin[2,3], Y. Liang[2,3], X. Jin[3,4], Z.X. Ping Cuo[2,5], J.E. Pool[1,6], X. Xu[3], H. Jiang[3], N. Vinckenbosch[1], T.S. Korneliussen[7], Y. Li[3], Y. Zhang[3], X. Zhang[3], R. Li[3,7], S. Li[3], H. Yang[3], R. Nielsen[1,7], J. Wang[3,7], J. Wang[3], BGI-TIBET Consortium.* 1) Stats/Integrative Biology, UC Berkeley, Berkeley, CA; 2) The Graduate University of Chinese Academy of Sciences, Beijing 100062, China; 3) BGI-Shenzhen, Shenzhen 518083, China; 4) Innovative program for undergraduate students, School of Bioscience and Biotechnology, South China University of Technology, Guangzhou, 510641, China; 5) The People's Hospital of the Tibet Autonomous Region, Lhasa, 850000, China; 6) Department of Evolution and Ecology, UC Davis, Davis, CA 95616, USA; 7) Department of Biology, University of Copenhagen, Copenhagen, Denmark.

The extreme altitudes of the Tibetan Plateau represent one of the most challenging environments settled by humankind, and residents of this region possess heritable adaptations to their environment. We present the sequencing of 50 exomes of ethnic Tibetans from the People's Republic of China. These coding sequences encompass approximately 92% of human genes with an average coverage of 18x per individual. We identify genes showing allele frequency changes specific to the high altitude sample, which represent strong candidates for altitude adaptation. Genes implicated include the oxygen transporters β-globin and G γ-globin 2, but the strongest signal is observed at EPAS1, a transcription factor involved in oxygen regulation. One SNP at EPAS1 shows a 78% frequency difference between Tibetan and Han samples, representing the fastest allele frequency change observed at any human gene, and an association with erythrocyte levels, consistent with the role of EPAS1 in adaptation to hypoxia.

263

Whole genome low-pass sequencing combined with GWAS data detects variants associated with cholesterol and hemoglobin levels in individuals from the island of Kosrae, Micronesia. *I. Pe'er[1], A. Gusev[1], M.J. Shah[2], E.E. Kenny[1,3], A. Ramachandran[1], J.K. Lowe[5], J. Salit[3], C.C. Lee[2], E.C. Levandowsky[2], T.N. Weaver[2], Q.C. Doan[2], H.E. Peckham[2], S.F. Mclaughlin[2], M.R. Lyons[2], V.N. Sheth[2], M. Stoffel[4], F.M. De La Vega[2], J.M. Friedman[3], J.L. Breslow[3].* 1) Dept Computer Sci, Columbia Univ, New York, NY; 2) Life Technologies Inc., Beverly, MA; 3) Rockefeller University New York, NY; 4) ETHZ Zurich, Switzerland; 5) University of California Los Angeles, CA.

High throughput sequencing opens a window to detect new, rare variants affecting complex traits. However, genomewide sequencing is typically resource constrained to scales below the thousands of individuals needed to for sampling a rare variant in sufficiently many copies for powered association testing. We reason that in a bottleneck population, sequencing a few individuals directly ascertains variants from the population bottleneck that may be rare elsewhere, and can be imputed to a larger sample relying on shared haplotypes detected in SNP array data. We present data and analysis on seven sequenced individuals from the bottleneck population of Pacific island of Kosrae, Federated States of Micronesia, where multi-trait GWAS had been conducted. We report identification of long regions with haplotypes co-inherited between pairs of individuals and methodology to leverage such shared genetic content for imputation. Our estimates show that sequencing as few as 40 personal genomes allows for imputation in up to 60% of the 3,000-person cohort at the average locus. We ascertained a pilot data-set of whole-genome sequences from four Kosraean individuals, with average 4X coverage. This dataset identified 4,567,947 unique single nucleotide variants in total, with 1,075,708 previously un-annotated. These Kosraen variants are unusually enriched for alleles that are rare in other populations. We find that regions of shared haplotypes between pairs of samples based on SNP data are supported by local depletion in opposite-allele homozygotes in sequencing data. We specifically interrogated two regions implicated in Kosraens by haplotype-based association with Hemoglobin A1c ($P=1.6\times10^{-23}$, 10.1% frequency) and Total Cholesterol ($P=2.94\times10^{-10}$, 4.4%). Variants in these regions were validated in three additional low-pass Kosraen personal genomes, for a total of three HBA1c carriers and four TC carriers. In both regions, we report novel, putatively functional variants present exclusively in the carriers. In particular, we highlight a large deletion associated with HBA1C. This effort presents a first study of association using whole genome sequencing.

264

An epistatic model for estimating the strength of balancing selection in genes of the human immune system. *E.O. Buzbas, N.A. Rosenberg.* Human Genetics, University of Michigan, Ann Arbor, MI.

High levels of polymorphism suggest that gene families of the human immune system, such as human leukocyte antigen (HLA) and killer-cell immunoglobulin-like receptor (KIR) genes, experience balancing selection. Further, allele frequencies at HLA and KIR loci are shaped by between-locus epistatic interactions, which result from cooperation between HLA and KIR regulated components of the immune system to produce an efficient response against pathogens. The presence of such interactions complicates inference on the strength of selection and requires models and methods specifically designed for taking into account the effect of epistasis. Motivated by the HLA-KIR system, we have developed a multi-locus Wright-Fisher model of balancing selection, allowing epistatic interactions between loci. Given the allele frequencies at the loci of interest, we develop computational and (bayesian) statistical methods to choose a plausible model of epistasis and to make inference on the strength of balancing selection. An algorithm to generate multi-locus allele frequencies under balancing selection is also presented as a tool to understand the eff ects of balancing selection and epistasis in shaping the allele frequencies. We demonstrate the methods by analyzing previously reported allele frequency data at HLA-A, HLA-B, and KIR loci. Our work fills a gap in theoretical population genetics by clarifying the role of balancing selection on multi-locus genotypes in producing high levels of polymorphism under epistasis.

265

Genome-wide association analysis of the Norfolk Island isolate implicates novel variants in migraine susceptibility. *L.R. Griffiths[1], H.C. Cox[1], R.A. Lea[1], C. Bellis[2], M. Carless[2], T.D. Dyer[2], J. Charlesworth[2], E. Matovinovic[1], S. MacGregor[3], J. Blangero[2].* 1) Genomics Research Centre, Griffith Health Institute, Griffith University, Southport, QLD, Australia; 2) Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio 78227 TX; 3) Statistical Genetics Laboratory, Queensland Institute of Medical Research, Herston, QLD, Australia.

Migraine is a debilitating neurovascular disorder with a complex genetic aetiology. The disease affects ~12% of the Caucasian population and 3 times more females are diagnosed. To identify genes involved in migraine susceptibility we performed a GWAS of the isolated population of Norfolk Island, located 1500 kilometres west of Australia. This unique population derives from a small number of British (Bounty mutineers) and Polynesian female founders forming ~5700 member pedigree spanning 11 generations and exhibiting substantial inbreeding and admixture. These population genetic features may facilitate gene mapping strategies. 377 founder-related adults were phenotyped for migraine using the diagnostic criteria of the International Headache Society and 285 of these individuals were using the Illumina® 610-Quad Beadchip. Heritability and linkage-based association analysis of the migraine phenotype was performed using the SOLAR program. GWAS results were adjusted for sex, age, admixture and inbreeding and SNPs prioritised based on both statistical and biological significance. A total of 544,590 SNPs across chromosomes 1 to 22 were available for analysis with the genome-wide statistical significance threshold adjusted by the number of effective tests estimated empirically (α=1×10-7) and biological significance assessed using annotation data analysed by the WGAViewer program and a gene-wide statistical significance threshold (α=1×10-4). We identified 96 migraine affected individuals in the Norfolk pedigree yielding a point prevalence estimate of 25.5%. Pedigree analysis indicated that the migraine phenotype had strong heritability (h2=0.53, P=0.016). GWAS analysis and SNP prioritisation incorporating biological annotation implicated 13 SNPs in 9 genes as being associated with migraine risk at the gene-wide level. Interestingly, 4 of the SNPs tag haplotypes in a single block and one of the four SNPs confers a Thr-Ala amino acid change providing a plausible candidate variant for involvement in disease causation. In conclusion this study has identified a high prevalence and heritability of migraine in the Norfolk Island genetic isolate and GWAS results have implicated a biologically plausible variant in disease susceptibility. Further studies are required to determine the specific functional role of this variant in migraine and whether the disease associations observed in our genetic isolate studies extend to the general out-bred population.

266

Mitochondrial Haplogroup X is Associated with Successful Aging in the Amish. *M.D. Courtenay[1], J.R. Gilbert[1], L. Jiang[2], A.C. Cummings[2], P.J. Gallins[1], L. Caywood[1], L. Reinhart-Mercer[1], D. Fuzzell[2], C. Knebusch[2], R. Laux[2], J.L. McCauley[1], C.E. Jackson[3], M.A. Pericak-Vance[1], J.L. Haines[2], W.K. Scott[1].* 1) Hussman Institute for Human Genomics, University of Miami, USA; 2) Center for Human Genetics Research, Vanderbilt University, USA; 3) Scott & White, Temple, TX.

Avoiding disease, maintaining physical and cognitive function, and continued social engagement in late life describe successful aging (SA). Mitochondrial lineages described by patterns of common genetic variants ("haplogroups") have been associated with increased longevity in different populations. We investigated the influence of mitochondrial haplogroups on SA in a sample of Amish over age 80. Cognitively intact volunteers aged ≥80 (n= 263) were enrolled in a population-based door-to-door survey of Amish communities in Indiana and Ohio. Individuals scoring in the top third of the sample for lower extremity function, needing little assistance with self-care tasks, having no depression symptoms, and expressing high life satisfaction were considered SA (n=73). The remainder (n=190) were retained as controls. These individuals descend from 46 matrilines in a single 13 generation pedigree constructed from the Anabaptist Genealogy Database. Mitochondrial haplogroups were assigned using the 10 mitochondrial single nucleotide polymorphisms (mtSNPs) defining the nine most common European haplogroups. An additional 17 polymorphic mtSNPs from a previous genome wide association study were also investigated. Associations between haplogroups, mtSNPs, and SA were determined by logistic regression models accounting for sex, age, BMI, and matriline via Generalized Estimating Equations. A significant positive association was found with haplogroup X (freq=0.065, OR=8.50, 95%CI 2.56-28.23, p=0.0005), while a negative association was found with haplogroup J (freq=0.046, OR=0.38, 95%CI 0.25-0.58, p=<0.0001). Individual mtSNP results were consistent with these haplogroup results. All positively associated alleles were found together on haplogroup X: 1719A (OR=8.72, 95%CI 2.63-28.84, p=0.0004), while all negatively associated alleles fell in haplogroup J: rs2854122 (OR= 0.094, 95%CI 0.017-0.52, p= 0.0066), rs3135030 (OR=0.11, 95%CI 0.033-0.36, p=0.0003), 10398G (OR=0.41, 95%CI 0.26-0.64, p=<0.0001). These data represent a novel association of haplogroup X with SA and conflict with previous associations of haplogroup J with longevity in other populations. More detailed analysis is needed to identify specific functional sequence variants on each haplogroup that explain these associations.

267

An evaluation of meta-analysis approaches in the presence of allelic heterogeneity. *L. Zgaga[1], J. Asimit[2], I. Rudan[1], A. Day-Williams[2], E. Zeggini[2].* 1) Public Health Sciences, University of Edinburgh, Edinburgh, Scotland; 2) Wellcome Trust Sanger Institute, The Morgan Building, Wellcome Trust Genome Campus, Hinxton, Cambridge.

Meta-analysis of genome-wide association studies (GWAS) is a powerful approach to identifying associations between complex traits and variants with modest effects. Most GWAS meta-analyses have thus far focused on populations of European descent and common variation. The study of low frequency and rare variants and studies of genetically diverse populations can introduce varying degrees of allelic heterogeneity and greatly dilute meta-analysis power. The objective of this study is to evaluate the power of different meta-analysis approaches in the presence of allelic heterogeneity. We simulated genotypic (110 SNPs) and phenotypic data (2000 cases, 2000 controls) for 3 populations with a total of 5 causal SNPs (1 in POP1, 2 in POP2 and 2 in POP3, all unique) on the basis of 1000 genomes data. We designed the simulations to reflect two paradigms of disease association: (1) low frequency/rare (freq. 0.5-4%) causal alleles with moderate OR (2-3) (i.e. multiple rare variants hypothesis), and (2) common (freq. 15-40%) causal alleles with small OR (1.1-1.2) (in trans-ethnic mapping). For each population we performed 3 association analyses: single-SNP analysis, a test of accumulation of multiple rare variants at the same locus, and a locus-wide allele-based matching association test, for which knowledge of the risk allele is not required. We then meta-analyzed the 3 populations using two p-value-based strategies: accounting for and ignoring direction of effect. Analyses comprised 1000 replicates. For common-frequency causal SNPs, the proportion of meta-analysis iterations that detected 0, 1, 2 and 3-5 causal variants was 17.7%, 25.6%, 27.0% and 29.7% respectively (taking direction of effect into account). When ignoring direction of effect, the proportion of those that detected 0, 1, 2 and 3-5 causal variants was 1.0%, 5.6%, 16.5% and 76.9%. For rare causal SNPs these proportions were 11.9%, 29.8%, 34.4% and 23.9% accounting for direction of effect, and 0.8%, 6.9%, 21.0% and 71.3% when ignoring direction. The performance of P-value-based meta-analysis on SNP-specific association results in the presence of allelic heterogeneity is suboptimal, both when taking direction of the effect into account and when ignoring it. Therefore, there is a need to develop better approaches for meta-analyzing data in the presence of allelic heterogeneity. We propose that locus-specific approaches perform better than SNP-specific approaches; our exhaustive evaluation is in progress.

268

Increased power of mixed-models facilitates association mapping of 10 loci for metabolic traits in an isolated population. *E.E. Kenny[1,2], M. Kim[2], A. Gusev[2], J. Salit[1], J.K. Lowe[3,4], J. Gustav Smith[5], S. Kovvali[3], H.M. Kang[6], C. Newton-Cheh[5], M.J. Daly[3,4], M. Stoffel[7], D.M. Altshuler[3,4,5], J.M. Friedman[1], E. Eskin[8,9], J.L. Breslow[1], I. Pe'er[2].* 1) Rockefeller University, New York, NY 10065, United States of America; 2) Computer Science dept, Columbia University, New York, NY 10027, United States of America; 3) Program in Medical and Population Genetics, The Broad Institute of Harvard & MIT, 7 Cambridge Center, Cambridge, MA 02142, United States of America; 4) Department of Molecular Biology, Massachusetts General Hospital, 185 Cambridge St, Boston, MA 02114, United States of America; 5) Center for Human Genetics Research, Massachusetts General Hospital, 185 Cambridge St, Boston, MA 02114, United States of America; 6) Bioinformatics Graduate Program, University of Michigan, 109 S. Observatory St , Ann Arbor, MI 48109-2029, United States of America; 7) Institute of Molecular Systems Biology, Swiss Federal Institute of Technology (ETH), Wolfgang-Pauli-Str. 16, 8093 Zurich, Switzerland; 8) Department of Computer Science, University of California Los Angeles, Los Angeles, California, United States of America; 9) Department of Human Genetics, University of California Los Angeles, Los Angeles, California, United States of America.

The potential benefits of using population isolates in genetic mapping, such as reduced genetic, phenotypic and environmental heterogeneity, are offset by the challenges posed by the large amounts of direct and cryptic relatedness in these populations confounding basic assumptions of independence. We have evaluated four representative specialized methods for association testing in the presence of relatedness which can be classified as either; (i) within-family (ii) within- and between-family and (iii) mixed-models methods, using simulated traits for 2,906 subjects with known genome-wide genotype data from an extremely isolated population, the Island of Kosrae, Federated States of Micronesia. We report that mixed models optimally extract association information from such samples, demonstrating 88% power to rank the true SNP as among the top 10 genome-wide with 56% achieving genome-wide significance, a >80% improvement over the other methods. We then used the mixed model method to re-analyze data for 17 published phenotypes relating to metabolic traits and electrocardiographic measures, along with another 8 previously unreported. As proof of the pudding, we replicate nine genome-wide significant associations with known loci of plasma cholesterol, high density lipoprotein, low density lipoprotein, triglycerides, thyroid stimulating hormone, homocysteine, C-reactive protein and uric acid, with only one detected in the previous analysis of the same traits. Further, we leveraged shared identity-by-descent genetic segments in the region of the uric acid locus to fine-map the signal, refining a known locus by a factor of four, and supporting independence between the two previously reported signals in the region. We also report a novel association for height (rs17629022, $p<2.1x10^{-8}$). A broader examination of variants that are associated to traits in large, predominantly European-based studies revealed that >50% do not replicate on Kosrae due to bottleneck effects rather than lack of power. On the other hand, associations to known variants that are observed on the island show stronger effects ($p<0.027$), potentially due to increased genetic and environmental homogeneity, demonstrating that isolate populations can have greater power to non-isolate populations for detecting common variants.

269

Conditioning on Known Associations Improves the Power of Association Studies. *N. Zaitlen[1,2,3], B. Pasaniuc[1,2,3], H.M. Kang[4], A.L. Price[1,2,3].* 1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) Department of Biostatistics, Harvard School of Public Health, Boston, MA; 3) Broad Institute of Harvard and MIT, Cambridge, MA; 4) Department of Biostatistics, University of Michigan, Ann Arbor, MI.

Standard GWAS analyses compute a marginal P-value for each SNP independently, and conditioning on associated SNPs is typically used as tool for fine mapping or to determine if additional signal exists within an associated locus. We show via analytical proof, extensive simulation, and application to real data that this current standard GWAS analysis method is not optimally powered for detecting associated loci and introduce a novel statistic that carefully utilizes strongly associated SNPs to elicit up to multiple orders of magnitude improvement in P-values while preserving the type I error rate. We compare our approach to conditioning on SNPs of large effect, which can also elicit a substantial change in test statistic. For example, in the original WTCCC T1D scan SNP rs2542151 on chr 18 has a modest reported P-value of $2x10^{-6}$; conditioning on the most significant SNP on chr 6 improves this tenfold to a P-value of $2x10^{-7}$, meeting the study's significance threshold. Surprisingly, conditioning on SNPs of strong effect does not always improve power and in certain situations can result in severe loss of power. In a realistic simulation of a continuous phenotype with one SNP of large effect and one SNP of weak effect, conditioning improved power from 67% to 95%. However, in a similar case control simulation for a disease with low prevalence, conditioning lowered the power from 79% to 61%. This substantial difference in the effect of conditioning is a function of study design and disease prevalence, affecting both continuous and dichotomous phenotypes. Our statistic properly accounts for these issues and we offer analytical proof that it outperforms both the standard marginal approach as well as the simple conditioning approach, and examine its relative benefits over existing methods via simulation. When multiple causal variants exist in the same gene, or when tag SNPs are differently correlated, their signals may interfere when tested independently preventing their discovery in a GWAS. In the WTCCC RA study for example, two linked SNPs rs9268403 and rs2844559 have marginal χ^2 values 281.1 and 2.5 when tested independently. When modeled together their χ^2 values become 412.9 and 43.3 respectively, a several order of magnitude change in P-value for both SNPs. We derived the distribution of test statistics over multiple linked SNPs and show how conditioning can recover such lost signals possibly accounting for some of the missing heritability in current GWAS.

270

Genome wide association of familial late onset Alzheimer's disease in the NIA LOAD and NCRAD samples replicates BIN1 and nominates CUGBP2 in interaction with APOE. *E. Wijsman[1,2], N. Pankratz[3], Y. Choi[2], J. Rothstein[1], K. Faber[3], R. Cheng[4], J. Lee[4], T. Bird[5,6], D. Bennett[7], R. Diaz-Arrastia[8], A. Goate[9], M. Farlow[10], B. Ghetti[11], R. Sweet[12], T. Foroud[3], R. Mayeux[4], The NIA-LOAD/NCRAD Family Study Group.* 1) Div Med Gen/Dept Med, Univ Washington Sch Med, Seattle, WA; 2) Dept. of Biostatistics, University of Washington, Seattle, WA; 3) Dept Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN; 4) The Gertrude H. Sergievsky Center, The Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University College of Physicians and Surgeons, New York, NY; 5) Geriatric Research Education and Clinical Center, Veterans Affairs Puget Sound Health Care System, Seattle Division, Seattle, WA; 6) Dept. of Neurology, University of Washington, Seattle, WA; 7) Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL; 8) Dept. Neurology, University of Texas Southwestern Medical Center, Dallas, TX; 9) Dept. Psychiatry, Washington University School of Medicine, St. Louis, MO; 10) Dept. Neurology, Indiana University School of Medicine, Indianapolis, IN; 11) Dept. Pathology, Division of Neuropathology, Indiana University School of Medicine, Indianapolis, IN; 12) Dept. of Psychiatry, University of Pittsburgh, Pittsburgh, PA.

Late-onset Alzheimer's disease (LOAD) is the most common form of dementia in the elderly, with strong evidence for a genetic basis. However, other than the well-established contribution of APOE to disease risk, identification of relevant risk loci has been challenging, despite genome scans in multiple samples. We carried out a joint genome-wide association study (GWAS) of multiplex LOAD families from The National Institute of Aging-Late Onset Alzheimer's Disease Family Study and the National Cell Repository for Alzheimer's Disease, with analyses limited here to the European-American subjects (3,839 affected and unaffected individuals from 992 families, and 1,002 additional unrelated neurologically evaluated normal subjects). Genotyping was carried out by CIDR using the 610 IlluminaQuad panel. Statistical analysis used kinship coefficients estimated from the data to correct for the presence of both known and cryptically related individuals in the sample, with this cohort representing the largest family-based GWAS of LOAD to date. Analysis was carried out on the full sample; on strata defined by APOE genotypes; and on strata defined by north European, south European, and Ashkenazi ancestry, as inferred from the data. SNPs near APOE gave highly significant results, but no other genome-wide significant evidence for association was obtained in analysis of the full sample. However, analyses that stratified on APOE genotypes identified SNPs on chromosome 10p14 with genome-wide significant evidence for association within APOE $\varepsilon 4$ homozygotes, with replication in an independent cohort. This APOE-stratified analysis eliminated all residual association near APOE. Evidence of association for recently-reported LOAD risk loci revealed support for BIN1 across identified European-American sub-populations, with evidence for association increasing after APOE adjustment. Nominal evidence for CLU was also observed, but was reduced after APOE adjustment. Our results indicate that genetic structure, coupled with ascertainment biases resulting from the strong APOE association, affect genome-wide results of LOAD. We show that loci with large effects and strong association with disease, such as APOE, can lead to confounding that affects genome-wide results, requiring appropriate accommodation in analysis to avoid both false positive and false negative evidence of association. Similar adjustments may also be needed for many other large multi-site studies with strong risk loci.

271

The missing heritability of a model common trait - adult height - is partially detectable by a long "polygenic tail" of common variant signals with very small effect. *H. Lango Allen[1], K. Estrada[2], G. Lettre[3], S.I. Berndt[4], M.N. Weedon[1], F. Rivadeneira[2], G.R. Abecasis[5], M. Boehnke[6], C. Gieger[7], D. Gudbjartsson[8], N.L. Heard-Costa[9], A.U. Jackson[6], A.V. Smith[10], N. Soranzo[11], C. Willer[5], A. Kumar[12], A. Mahajan[13], W. Rayner[12], N. Robertson[12], A.D. Morris[14], C.N.A. Palmer[14], A.G. Uitterlinden[2], C.M. Lindgren[12], M.I. McCarthy[12], T.M. Frayling[1], J.N. Hirschhorn[15], The Genetic Investigation of ANthropometric Traits (GIANT) Consortium.* 1) Peninsula Medical School, University of Exeter, Exeter, United Kingdom; 2) Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands; 3) Montreal Heart Institute (Research Center), Université de Montréal, Montréal, Québec, Canada; 4) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA; 5) Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, 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) Icelandic Heart Association, Kopavogur, Iceland; 11) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 12) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 13) Institute of Genomics and Integrative Biology, CSIR, Delhi, India; 14) Biomedical Research Institute, University of Dundee, Ninewells Hospital & Medical School, Dundee, UK; 15) Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Boston, Massachusetts, USA.

Height is a classic, highly heritable polygenic trait. Since it is an easily and accurately measured phenotype, it is available for large number of samples, and can be used as a model for other common traits. We previously reported results from a genome-wide association (GWA) analysis of height using 133,600 European individuals from the Genetic Investigation of ANthropometric Traits (GIANT) Consortium. We identified 118 independent common variants (defined as 1Mb windows either side of the lead SNP) at $P<5x10^{-8}$. These signals together contributed <10% of phenotypic variation in height, or <12.5% of the heritable component. One of the most important questions emerging from GWA studies has been the location of "missing heritability". We hypothesized that common variant signals that individually do not reach conventional levels of GWA significance would increase the proportion of heritability explained. We took forward SNPs representing 89 independent signals at $5x10^{-8}<P<5x10^{-6}$ into an in-silico replication set of 50,000 samples of European ancestry. Of the 89 SNPs, 62 reached overall genome wide significance and 88 were directionally consistent with the initial analysis. We also assessed 227 independent signals at $5x10^{-8}<P<5x10^{-4}$ in a separate sample of 7000 individuals from the UK T2D Genetics Consortium, genotyped on the metabochip. The following number of SNPs showed directional consistency in these P-value ranges: $5x10^{-8}<P<5x10^{-7}$: 25 out of 28 SNPs (89.3%, sign test $P=2.7x10^{-5}$); $5x10^{-7}<P<5x10^{-6}$: 47/54 SNPs (87.0%, $P=2.3x10^{-8}$); $5x10^{-6}<P<5x10^{-5}$: 71/83 SNPs (85.5%, $P=2.4x10^{-11}$); $5x10^{-5}<P<5x10^{-4}$: 48/61 SNPs (78.7%, $P=7.7x10^{-6}$). Consistent with these results, a deep set of independent variants (in the range of $0.05>P>5x10^{-8}$) accounted for up to 16.8% of phenotypic variation in height, or ~20% of the heritable component. A second potential source of common variation that may increase the proportion of heritability explained are common variants that fall within an already identified locus, but are overlooked because they are in partial linkage disequilibrium with the confirmed variant. Conditional analysis showed that at 19 loci there is at least one additional independent signal (post-conditioning P-values were between $1x10^{-14}$ and $2x10^{-7}$). In summary, we provide strong evidence that there are many more additional variants associated with common polygenic traits among loci that miss the strict genome-wide significance threshold.

272

Large-scale association studies of adult height using a 50K SNP array. *B. Keating[1], Y. Guo[1], M. Lanktree[2], C. Simpson[3], J. Glessner[1], H. Ongen[5], M. Kumari[6], K. Wang[1], H. Qui[1], P. Talmud[6], J. Peden[5], A. Reiner[4], S. Grant[1], R. Hegele[2], H. Hakonarson[1].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Robarts Research Institute, The University of Western Ontario, 4288A-100 Perth Drive, London, Ontario, Canada, N6A 5K8; 3) National Human Genome Research Institute, National Institutes of Health 333 Cassell Drive, Suite 1200 Baltimore, MD 21224; 4) Department of Epidemiology University of Washington Seattle, WA 98195; 5) Wellcome Trust Center for Human Genetics, Roosevelt Drive, University of Oxford, OX37EQ, UK; 6) Centre for Cardiovascular Genetics Dept of Medicine, University College London 5 University St, London WC1E 6JF.

Adult height is often regarded as a model polygenic trait. Although complex processes such as age of puberty, perinatal environment and nutritional intake affect stature, up to 85% of the variation in adult height in populations can be attributed to heritability. While a number of loci have been shown to associate with height through large scale meta-analyses of genome wide association studies, collectively they only explain a small portion of the genetic variance. We designed and utilized a cardiovascular (CV) gene-centric array comprising of ~50K SNPs across ~2000 loci. As this 50K SNP array comprises ~15K non-HapMap SNPs, and with most loci having greater coverage than conventional genome wide association (GWA) tools, it affords a strong opportunity to assess deeper screening of genetic variation in these CV related loci to account for additional independent signals within known and novel height related loci. We performed a detailed analysis and comparison of common and rare variants associated with adult height in over 67,200 individuals from 32 studies using discovery and replication phases unveiling 26 novel genetic signals which underpin adult height ($p < 5 \times 10\text{-}8$) and confirm 12 previously described signals. For 'array wide' significance of $p < 2 \times 10\text{-}6$ we described 45 replicated signals. We describe a number of signals from non-HapMap SNPs which are poorly tagged, and 'imputable', by the conventional GWAS tools and are thus strong candidates to explain portions of missing variance that would not be arrived at through conventional GWA meta-analysis. After conditional analysis we found a large portion of independent signals. Using admixture mapping refines the causal signals in 14 of these loci. We are currently analyzing height phenotypes from an additional 48,650 samples genpotyped on this array.

273

Thirty two novel loci for age at menarche identified by a meta-analysis of genome-wide association studies. *A. Murray[1], C.E. Elks[2], J.R.B. Perry[1], P. Sulem[3], L.J. Bierut[4], D.I. Boomsma[5], H.A. Boyd[6], D.I. Chasman[20,21], L. Crisponi[7], E. Demerath[8], C.M. van Duijn[9], M.J. Econs[10], T.B. Harris[11], D.J. Hunter[12,13,14,15], R.J.F. Loos[2], A. Metspalu[16,17,18], G.W. Montgomery[19], P.M. Ridker[20,21], T.D. Spector[22], E.A. Streeten[23], K. Stefansson[3,24], U. Thorsteinsdottir[3,24], A.G. Uitterlinden[25,26], E. Widen[27], J.M. Murabito[28,29], K.K. Ong[2,30], ReproGen Consortium.* 1) Peninsula Med School, University of Exeter, Exeter, UK EX1 2LU; 2) Medical Research Council (MRC) Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK; 3) deCODE Genetics, Reykjavik, Iceland; 4) Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA; 5) Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands; 6) Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; 7) Istituto di Neurogenetica e Neurofarmacologia, Consiglio Nazionale delle Ricerche, Cagliari, Italy; 8) Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA; 9) Genetic-Epidemiology Unit, Department of Epidemiology and Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands; 10) Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indiana, USA; 11) Laboratory of Epidemiology, Demography, and Biometry, Intramural Research Program, National Institute on Aging, Bethesda, Maryland, USA; 12) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20892, USA; 13) Department of Nutrition, Harvard School of Public Health, Boston, MA, USA; 14) Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA; 15) Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, and Harvard Medical School, Boston, Massachusetts, USA; 16) Estonian Genome Center, University of Tartu, Tartu, Estonia; 17) Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 18) Genotyping Core Facility, Estonian Biocenter, Tartu, Estonia; 19) Molecular Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia; 20) Division of Preventive Medicine, Brigham and Women's Hospital, 900 Commonwealth Avenue East, Boston MA 02215, USA; 21) Harvard Medical School, Boston, Massachusetts, USA; 22) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 23) Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, Maryland, USA; 24) Faculty of Medicine, University of Iceland, Reykjavik, Iceland; 25) Department of Internal Medicine, Erasmus MC, Rotterdam, the Netherlands; 26) Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands; 27) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland; 28) The National Heart Lung and Blood Institute's Framingham Heart Study, Framingham, MA, USA; 29) Sections of General Internal Medicine, Preventive Medicine and Endocrinology, Department of Medicine, Boston University School of Medicine, Boston, MA, USA; 30) Department of Paediatrics, University of Cambridge, Cambridge, UK.

Background: Genome wide association studies have recently identified two loci associated with age at menarche. The first is the LIN28B locus on chromosome 6, which has been associated with adult height and is involved in regulation of the let-7 family of microRNAs which have been implicated in stem cell differentiation and tumourigenesis. The second menarche locus, at 9q31.2, is intergenic with no obvious candidate genes nearby. Aim: To identify additional loci for age at menarche, we extended the previous meta-analyses to include 32 genome-wide association studies involving 87,802 women of European descent. Results: We identified at least 32 novel genome-wide significant menarche loci (P<5x10-8), in addition to the two known loci at LIN28B (P=1.59x10-58) and 9q31.2 (P=4.4x10-33). The effect sizes ranged from 2.1 to 4.7 weeks per allele for the novel loci. Four loci have previously been associated with obesity (in/near FTO, SEC16B, TRA2B and TMEM18), with the BMI increasing allele associated with earlier age at menarche. Three loci were in/near genes implicated in body weight regulation and energy homeostasis in animal models (BSX, CRTC1, and MCHR2), but have not been identified as obesity-related by genome-wide scans. Of 16 published loci associated with obesity and related traits, 11 were associated with age at menarche (p<0.05). Three of the novel menarche-associated loci were in/near genes implicated in hormonal regulation (INHBA, PCSK2 and RXRG). The remaining 22 novel loci were in or near genes which have not previously been associated with puberty or related traits. Conclusions: Our findings suggest multiple biological pathways contribute to the timing of menarche, including genes involved in the regulation of body size/adiposity and hormone regulation. The role of many of the novel loci have yet to be confirmed and highlight the complexity of the control of pubertal timing.

274

Genome-wide Association Study reveals major gene for hypospadias. *L.F.M. van der Zanden[1], I.A.L.M. van Rooij[1], J. Knight[2], W.F.J. Feitz[1], S.H.H.M. Vermeulen[1], L.S. Baskin[3], A. Nordenskjöld[4], N. Roeleveld[1], B. Franke[1], N.V.A.M. Knoers[1].* 1) Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 2) King's College London, London, United Kingdom; 3) University of California, San Francisco, CA, United States; 4) Karolinska Institutet, Stockholm, Sweden.

Hypospadias is a common congenital malformation of the male external genitalia with a multifactorial etiology. To identify genetic variants contributing to susceptibility for hypospadias, we performed the first Genome-wide Association Study (GWAS) for this malformation using DNA pooling in the largest sample of hypospadias cases thus far reported. We included 436 Caucasian cases with isolated glanular or penile hypospadias from the AGORA project (Aetiologic research into Genetic and Occupational/environmental Risk factors for Anomalies in children) and 494 unaffected Caucasian male controls from the Nijmegen Biomedical Study and allelotyped 906,600 single nucleotide polymorphisms (SNPs) using Affymetrix 6.0 arrays. Allele frequencies were calculated using k-corrected signal intensities. The highest ranked SNPs that fulfilled quality-control criteria were selected for individual genotyping in the discovery sample. Furthermore, results were replicated in an additional Dutch sample of 133 hypospadias cases and their parents, using the Transmission Disequilibrium Test, and in a Swedish cohort of 266 hypospadias cases and 399 male controls. Seven of the highest ranked SNPs were located in the same X-chromosomal gene, *DGKK*. As most SNPs in this gene are in strong linkage disequilibrium with each other, we selected the (intronic) SNP that tagged most other SNPs and a potentially regulatory SNP in the 5' upstream region. Both SNPs showed highly statistically significant association results in the discovery sample (odds ratio (OR) = 2.46, $p=2.5 \times 10^{-11}$ and OR=2.25, $p=2.9 \times 10^{-9}$, respectively) as well as in the Dutch replication sample (OR=3.92, $p=5.2 \times 10^{-6}$ and OR=3.83, $p=8.0 \times 10^{-6}$, respectively) and in the Swedish replication sample (OR=2.48, $p=2.6 \times 10^{-8}$ and OR=2.16, $p=2.7 \times 10^{-6}$, respectively). Furthermore, nine SNPs in other genes were genotyped in the individual samples, of which eight showed statistically significant results in the discovery sample ($p<0.05$). These associations were not replicated in the replication samples. Experiments studying expression of *DGKK* in relevant human tissue are ongoing.

275

A Genome Wide Association Study Identifies an *IL17RC* Missense Mutation (S111L) as an Adolescent Idiopathic Scoliosis Locus. *J.P. Dormans[1], S.F.A. Grant[2,3,4], N. Rendon[1], F.D. Mentch[2], C.E. Kim[2], E.C. Frackelton[2], J.P. Bradfield[2], H. Zhang[2], R.M. Chiavacci[2], H. Hakonarson[2,3,4].* 1) Department of Orthopedic Surgery, Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, PA 19104; 2) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA 19104; 3) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA 19104; 4) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104.

Untreated scoliosis, particularly in more severe cases, has a detrimental influence on health throughout life plus a negative socioeconomic impact with respect to work and marital status. Approximately three quarters of structural scoliosis is clinically classified as idiopathic, which is the most common spine deformity arising during childhood. One of the main subforms of the disorder is adolescent idiopathic scoliosis (AIS), which presents in children aged 10 to 16 years old. There is strong evidence for a genetic component to idiopathic scoliosis. It often appears in several members of the same family; indeed, one in three cases is familial. Also studies on twins showing that concordance of monozygotic twins is greater than that of dizygotic twins. Nevertheless, in addition to genetic traits, a number of other environment-related factors have been described that variously contribute to the final expression/modulation of the phenotype in a given individual. Classical candidate gene studies have only achieved limited success in identifying genetic determinants of idiopathic scoliosis. As such we elected to perform a genome wide association study of AIS on subjects recruited from the Department of Orthopedic Surgery at the Children's Hospital of Philadelphia. We genotyped ~550,000 single nucleotide polymorphisms (SNPs) with the Illumina Human Hap550 Genotyping BeadChip on our study population of 137 AIS cases of European ancestry and 2,126 controls. Following adjustment for local ancestry, four SNPs on chromosome 3p25.3, rs708567, rs172155, rs279572 and rs455863, reached the threshold for genome wide significance. The top signal at this locus, rs708567, is a common missense mutation (S111L) within the "interleukin 17 receptor C" (*IL17RC*) gene ($P=1.18 \times 10^{-9}$). This risk associated allele C confers an odds ratio of 2.28. This locus resides in a 68 kb block (Build hg18: chr3:9876479-9944516) of linkage disequilibrium (LD). In addition to *IL17RC*, this region also harbors the genes *CIDEC*, *JAGN1* and *IL17RE*. Efforts are now underway to replicate this association further, test it for association in additional sub-forms of idiopathic scoliosis and to test its functional role in the pathogenesis of the trait.

276

Genome-wide meta-analysis identifies 67 loci associated with platelet count and platelet volume in the HaemGen consortium. *C. Gieger[1], N. Soranzo[2,3] for the HaemGen consortium.* 1) Institute of Epidemiology, Helmholtz Zentrum Munich, Neuherberg, Germany; 2) Human Genetics, Wellcome Trust Sanger Institute, Genome Campus, Hinxton, UK; 3) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK.

Platelet count (PLT) and mean platelet volume (MPV) are highly heritable quantitative traits that are widely used in the clinic. MPV has been associated in epidemiological studies with an increased risk of myocardial infarction and stroke. In a previous meta-analysis encompassing about 5,000 individuals we identified 12 and 3 loci associated with MPV and PLT, together accounting for approximately 8 percent of phenotypic variation in MPV. In a subsequent analysis we showed that one of the PLT loci is associated with coronary artery disease and myocardial infarction. Here we extend such previous analysis to up to 23 world-wide cohorts, including over 50,000 (PLT) and 20,000 (MPV) samples. We discovered 67 independent loci reaching genome-wide significance ($P<=5x10^{-8}$), of which 15 were significantly associated both with PLT and MPV. Of these, 15 were described previously and 52 are novel. Some of the top loci found in our present analyses are rs1354034 in ARHGEF3 and rs3184504 in SH2B3 and rs10914144 in DNM3. The further characterizations of the loci were carried out including pleiotropy in the haematopoietic pathway. We also did a network analysis of the genes associated with the lead loci. Finally, we supplemented our work in populations of European ancestry by extensive inter-ethnic comparisons. Taken together this meta-analysis with its follow-up analyses constitutes the first comprehensive picture of the genetic architecture of platelet count and volume.

277

Disease-associated loci are highly significantly over-represented among genes bound by TCF7L2 *in vivo*. *J. Zhao[1], J. Schug[2], M. Li[3], K.H. Kaestner[2], S.F.A. Grant[1,2,4].* 1) Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Genetics and Institute of Diabetes, Obesity and Metabolism, University of Pennsylvania School of Medicine, Philadelphia, PA; 3) Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA; 4) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA.

The transcription factor 7-like 2 (*TCF7L2*) gene has been strongly implicated in type 2 diabetes and cancer. To uncover its downstream targets *in vivo*, we performed chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-seq) in the colorectal carcinoma cell-line, HCT116, where TCF7L2 is abundantly expressed. We performed ChIP-seq experiments for TCF7L2 from three independent chromatin samples and found 1,095 discrete binding sites across the genome, of which a subset were within 5kb of 548 annotated RefSeq genes. Despite using a cancer cell-line, the most significant functions represented in this gene list, based on scoring from the Ingenuity Pathway analysis software package, were related to 'Diabetes' (P = 5.60x10^{-13}). 'Endocrine system disorder' (P = 1.10x10^{-12}), 'genetic disorder' (P = 1.68x10^{-11}), 'non-insulin-dependent diabetes mellitus' (P = 1.28x10^{-10}) and 'coronary artery disease' (P = 1.77x10^{-10}) and easily survived correction for multiple testing. As one of the enriched categories was related to genetic disorders, we queried our results against all published GWAS; unexpectedly, we also observed a highly significant over-representation of reported loci in this catalogue (P = 7.50x10^{-15}), primarily for metabolic (P = 7.31×10^{-9}) and cardiovascular (P = 4.57×10^{-7}) traits, and in particular for type 2 diabetes (P = 4.40x10^{-5}) and height (P = 7.56x10^{-8}) ; however there was no or minor enrichment of GWAS-derived loci for cancer, inflammatory or neurological diseases. When defining the distance from genes at 50kb or 500kb, this enrichment pattern persisted, with some additional evidence for enrichment of cancer-related loci. In summary, a highly significant proportion of genes bound by TCF7L2 are known disease-associated loci. These findings suggest that TCF7L2 is a central node in the regulation of human diabetes and other disease-associated genes.

278

Meta-analysis of genome-wide association data and independent replication identifies nine loci associated with circulating proinsulin levels. *E. Ingelsson[1], J. Dupuis[2,3], R. Strawbridge[4], I. Prokopenko[5,6], A. Barker[7], E. Ahlqvist[8], D. Rybin[9], G. Dedoussis[10], P. Schwarz[11], I. Barroso[12], J.B. Meigs[13,14], R.M. Watanabe[15,16], V. Lyssenko[8], C. Langenberg[7], A. Hamsten[4], J.C. Florez[14,17,18,19] on behalf of the MAGIC investigators.* 1) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 2) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 3) National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA; 4) Atherosclerosis Research Unit, Department of Medicine, Solna, Sweden; 5) Wellcome Trust Center for Human Genetics, Oxford, UK; 6) Oxford Centre for Diabetes, Endocrinology and Metabolism, Oxford, UK; 7) MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK; 8) Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, University Hospital Malmö, Malmö, Sweden; 9) Boston University Data Coordinating Center, Boston, MA; 10) Department of Nutrition-Dietetics, Harokopio University, Athens, Greece; 11) Department of Medicine III, Division Prevention and Care of Diabetes, University of Dresden, Dresden, Germany; 12) Metabolic Disease Group, Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 13) General Medicine Division, Massachusetts General Hospital, Boston, MA; 14) Department of Medicine, Harvard Medical School, Boston, MA; 15) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 16) Department of Physiology and Biophysics, Keck School of Medicine, University of Southern California, Los Angeles, CA; 17) in Medical and Population Genetics, Broad Institute, Cambridge, MA; 18) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 19) Diabetes Research Center, Diabetes Unit, Massachusetts General Hospital, Boston, MA.

Circulating proinsulin is a precursor of mature insulin and C-peptide. In normoglycemic individuals, higher proinsulin levels are associated with elevated glucose levels, increased insulin secretion, insulin resistance and increased risk of type 2 diabetes (T2D). Thus, studies of proinsulin and the insulin-processing pathway can provide new insights about mechanisms leading to T2D. The Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) conducted a meta-analysis of genome-wide association results between proinsulin and ~2.5 million imputed or genotyped SNPs in 10,701 non-diabetic adult men and women of northern European ancestry from 4 cohorts, with a targeted follow-up of 23 loci (P<10^{-5}) in up to 16,933 individuals from 12 cohorts. We performed additive genetic models adjusted for age, sex, population structure and fasting insulin. In the combined stage 1 and 2 meta-analysis, nine loci associated with proinsulin levels at genome-wide significance (P<5*10^{-8}) were identified, in or near *CENTD2* (two independent loci), *MADD*, *PCSK1*, *TCF7L2*, *VPS13C/C2CD4A/B*, *SLC30A8*, *LARP6* and *SGSM2*. Further adjustment for BMI and fasting glucose did not alter the results. Four of these loci have previously been associated with circulating proinsulin (*MADD*, *TCF7L2*, *VPS13C/C2CD4A/B*, *SLC30A8*), and one (*PCSK1*) has been implicated in obesity, whereas four loci (two at *CENTD2*, *LARP6*, *SGSM2*) have not been previously reported to be related to any metabolic trait. Interestingly, the proinsulin-raising allele of *CENTD2* was associated with a lower fasting glucose (0.019 mg/dL per A allele; P= 1.7*10^{-4}), and a lower risk of T2D (OR, 0.88; P=7.8*10^{-6}); a similar effect was seen in the second *CENTD2* locus. The other three novel loci demonstrated nominally significant associations with either fasting glucose (*PCSK1*, P= 4.8*10^{-4}) or fasting insulin (*LARP6* and *SGSM2*, P=0.020 and 0.040, respectively), but were not associated with T2D risk (all P>0.40). The previously identified obesity locus *PCSK1* encodes the protein prohormone convertase 1/3 (PC1/3) which catalyzes the conversion of prohormones into functional hormones, and cleaves proinsulin to 33,32-split proinsulin. In summary, we have identified nine loci associated with circulating proinsulin. This improves our knowledge of the biology underlying glucose homeostasis and T2D development in humans, can stimulate follow-up functional and genetic research, and may provide guidance in developing new therapies for T2D.

279

Identification of Novel Type 2 Diabetes Loci in African Americans using a Genome-wide Association Approach. *N. Palmer[1], C. McDonough[1], P. Hicks[1], S. An[1], J. Hester[1], J. Cooke[1], M. Bostrom[1], L. Lu[1], J. Ziegler[1], M. Sale[2], J. Divers[1], D. Shriner[3], A. Adeyemo[3], C. Rotimi[3], M. Ng[1], C. Langefeld[1], B. Freedman[1], D. Bowden[1].* 1) Wake Forest University School of Medicine, Winston-Salem, NC 27157; 2) University of Virginia School of Medicine, Charlottesville, VA 22908; 3) National Institutes of Health, Bethesda, MD 20892.

African Americans are disproportionately affected by type 2 diabetes (T2DM) yet few studies have examined T2DM using genome-wide association approaches in this ethnic group. We have performed a genome-wide association study (GWAS) to identify common variation associated with T2DM in the African-American population. In the initial stage, SNPs were genotyped on the Affymetrix 6.0 array in 965 African-American cases with T2DM and end-stage renal disease (T2DM-ESRD) and 1,029 non-diabetic, non-nephropathy controls. A total of 832K genotyped and 2.3M imputed SNPs were evaluated. Analysis included principal components-based adjustment for admixture. High scoring SNPs (*n*=754) were genotyped in an identically ascertained replication cohort (709 T2DM-ESRD cases and 690 controls). Consistent association and direction of effect was observed with 122 SNPs which were validated in three additional cohorts (1,458 T2DM cases and 1,598 controls). Forty-seven SNPs had evidence of association in the GWAS and replication (admixture-adjusted P-values=0.049-3.62x10^{-6}) and were associated with T2DM in subjects without nephropathy (P<0.05). Meta-analysis of the five study cohorts (3,121 cases and 3,317 controls) revealed eight novel loci that contribute to T2DM (P-values<5.0x10^{-8}). Among these, rs2722769 (P=2.96x10^{-13}, OR=0.63) is located between *GALNTL4*, a glycosyltransferase whose family member *GALNT2* has been implicated in cholesterol metabolism in a large meta-analysis, and *ZBED5*, a zinc finger protein. SNP rs8043824 (P=7.51x10^{-13}, OR=0.69) is located downstream of *MAF* a transcription factor shown to activate basal expression of the glucagon gene. Variants within this region have been implicated in obesity in Europeans and T2DM in Han Chinese. Notably, the association observed herein remained significant after adjustment for BMI, P=3.17x10^{-8}. In addition, SNPs in *PARD3* (P=2.27x10^{-9}; an inhibitor of insulin-stimulated glucose uptake and causing impaired translocation of Glut4, a glucose transporter), *SMYD3* (P=5.34x10^{-9}; a coactivator for estrogen receptor-mediated transcription) and several intergenic regions with P-values<2.16x10^{-5} may represent additional novel loci that contribute to T2DM. These results suggest that multiple loci underlie T2DM susceptibility in African Americans. Importantly, a review of the literature shows that these loci are distinct from those identified in other ethnic populations to date.

280

A Large Multinational Genome-Wide Association Study Identifies Novel Loci that Influence Adiponectin Levels. *Z. Dastani[1], J. Dupuis[2], M. Hivert[3], J. B Meigs[4,20], T. M. Frayling[5], K. Song[6], P. Pramstaller[7, 8], G. Davey-Smith[5], J. Florez[4,9], d. Evans[5], T. Winkler[11], B. Glaser[5], F. Kronenberg[13], P. Deloukas[10], M. Ladouceur[19], HE. Wichmann[21, 22], P. Henneman[15], C. Fuchsberger[7], J. Perry[19], I. M. Heid[11,12], A. Hicks[7], C. Van Duijn[15], Y. Aulchenko[16], D. Waterworth[6], T. Tanaka[17,18], N. Timpson[5], T. Spector[14], JB. Richards[14,23] On behalf of AGIPOGen Consortium.* 1) Departments of Epidemiology and Biostatistics, McGill University, Montréal, Canada; 2) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 3) Service d'Endocrinologie, Département de Médecine, Université de Sherbrooke, Quebec, Canada; 4) Department of Medicine, Harvard Medical School, Boston, MA, USA; 5) MRC Centre for Causal Analyses in Translational Epidemiology, Department of Social Medicine, University of Bristol, Bristol BS8 2BN, UK; 6) Genetics Division, GlaxoSmithKline, King of Prussia, PA, USA; 7) Institute of Genetics Medicine, European Academy Bozen/Bolzano (EURAC), Bolzano, Italy; 8) Department of Neurology, University of Lübeck, Lübeck, Germany; 9) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 10) Wellcome Trust Sanger Institute, Cambridge CB10 1SA, UK; 11) Department of Epidemiology and Preventive Medicine, Regensburg University Medical Center, Regensburg, Germany; 12) Institute of Epidemiology, Helmholtz Center Munich, German Research Center for Environmental Health, Neuherberg, Germany; 13) Division of Genetic Epidemiology; Department of Medical Genetics, Molecular and Clinical Pharmacology; Innsbruck Medical University, Innsbruck, Austria; 14) Department of Twin Research and Genetic Epidemiology, Kings College London, London SE1 7EH, UK; 15) Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands; 16) Department of Epidemiology, Erasmus MC Rotterdam, The Netherlands; 17) Medstar Research Institute, Baltimore, MD, USA; 18) Clinical Research Branch, National Institute of Aging, Baltimore, MD, USA; 19) Genetics of Complex Traits, Peninsula Medical School, Magdalen Road, Exeter, UK; 20) General Medicine Division, Massachusetts General Hospital, Boston, MA, USA; 21) Helmholtz Zentrum München - German Research Center for Environmental Health, Institute of Epidemiology, Munich-Neuherberg, Germany; 22) BE, Chair of Epidemiology, University of Munich, Munich, Germany; 23) Departments of Medicine and Human Genetics, McGill University, Montréal, Canada.

Adiponectin, produced exclusively by adipocytes, is highly heritable and its levels have been inversely associated with both type 2 diabetes mellitus (T2D) and coronary heart disease (CHD). To identify common genetic variants that influence adiponectin levels, we conducted a meta-analysis of 11 genome wide association (GWA) studies involving 21,880 individuals. Each study assessed the relationship of ~2.5 million common (MAF >1%) single nucleotide polymorphisms (SNPs) for their association with adiponectin levels. Additive SNP effects on adiponectin levels were estimated in each study separately adjusting for age, sex and BMI. We then performed fixed-effect meta-analysis to assess the combined effect of each SNP. Sex-specific results were also meta-analyzed. We identified SNPs located on three independent loci associated with adiponectin at genome-wide significant levels (P £ 5×10-8). The strongest association with adiponectin levels was found with the SNPs at the adiponectin-encoding ADIPOQ locus (P-combined = 2.53x10-60 for lead SNP, rs17366568, allele A, with effect size = -0.1497 per ln-transformed adiponectin levels). Several novel variants in the CDH13 (cadherin 13, H-cadherin (heart)) gene on chromosome 16 (16q23.2-24.1, P-combined = 6.30x10-12 for lead SNP, rs8047711, allele A, with effect size = -0.1278) and in PEPD (peptidase D) gene on chromosome 19 (19q13.11, P-combined = 6.08x10-11 for lead SNP, rs731839, allele G, with effect size = -0.0283) were associated with lower in circulating levels of adiponectin. The rs4311394 in ARL15 gene, which has been previously associated with adiponectin levels reached suggestive association at P= 1.19x10-5. Previous GWA studies for hypertension and CHD identified CDH13 as a susceptibility locus for these diseases. Furthermore, a GWA study in the Japanese population reported suggestive evidence on PEPD gene for T2DM. In summary, we confirmed strong support for the role of two novel loci in adiponectin physiology, which may impact upon cardiometablic disease.

281

Comprehensive pair-wise genome scans reveal considerable epistasis in the genetic control of BMI in population isolates. *W. Wei[1], G. Hemani[2], V. Vitart[1], C. Cabrera-Cardenas[1], P. Navarro[1], J. Huffman[1], C. Hayward[1], S. Knott[3], A. Hicks[4], I. Rudan[5,6], P. Pramstaller[4,7,8], S. Wild[5], J. Wilson[5], H. Campbell[5], N. Hastie[1], A. Wright[1], C. Haley[1].* 1) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Edinburgh, Scotland, United Kingdom; 2) The Roslin Institute and R(D)SVS, University of Edinburgh, Roslin, Midlothian, UK; 3) Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK; 4) Institute of Genetic Medicine, European Academy Bozen/Bolzano (EURAC), Bolzano, Italy. Affiliated Institute of the University of Lübeck, Germany; 5) Centre for Population Health Sciences, University of Edinburgh, Edinburgh, UK; 6) Croatian Centre for Global Health, University of Split, Split, Croatia; 7) Department of Neurology, General Central Hospital, Bolzano, Italy; 8) Department of Neurology, University of Lübeck, Lübeck, Germany.

Body mass index (BMI) is a heritable complex trait that has been widely studied as a measure of human obesity. A short list of common variants mapping (near) to 25 genes have been identified and replicated in several large scale genome-wide association (GWA) studies on BMI. However, the identified variants explained a rather small proportion of heritable BMI variation. One possible source of the unexplained heritability could be epistasis that remains largely unexplored in most BMI GWA studies. Using a computational effective algorithm based on regression, we performed exhaustive pair-wise genome scans to profile epistasis in BMI in a population of 2475 individuals combined from the populations used in the Scottish ORCADES study and the Croatian VIS and KORCULA studies, with a common set of 283,971 SNPs. A nested test approach was used to handle multiple testing for epistasis. BMI was corrected for sex and age and normalised in each individual population before combining the data; the combined data were corrected for population and polygenic effects using a mixed model approach and the resultant residuals were used as the trait to test for association and epistasis. While no single SNPs were associated with BMI at the genome-wide significance level, seven SNP pairs passed the Bonferroni adjusted genome-wide thresholds ($P < 1.24 \times 10^{-12}$) for epistasis: 1p13.2 with 12q24.33, 10p14 with 22q12.3, and 9q31.1 with 1p34.1, 3p14.1, 9p24.1, 15q26.3, and 16p11.2. These pairs jointly explained 9.2% of the variance of the trait; each had weak marginal effects hence would have been missed if the epistasis search were performed only for loci with significant marginal effects. In addition, 6 BMI associated genes (*NRXN3*, *MSRA*, *FTO*, *NEGR1*, *LRRC16A*, *CNR1*) were found interacting strongly with other loci ($P < 10^{-9}$) when tested in a limited search for candidate locus interactions. Twelve epistatic pairs involving the 6 genes jointly explained 8.2% of the variance of the trait. Nearly all the 19 epistatic pairs above were replicated ($10^{-6} < P < 10^{-3}$) in the form of interacting regions (scanning all SNPs within a gene or one-million-bp region for epistasis) in the Italian MICROS study population (1202 subjects with 293913 SNPs). The results suggest that epistasis plays an important role in the control of BMI population isolates and the BMI associated genes may actively interact with other loci. Large scale replication studies are needed to confirm those epistasis signals.

282

The effect of body mass index on the risk of ischaemic heart disease: using genomewide association results to estimate causal effects. *N. Timpson[1,2], T. Palmer[1,2], J. Zacho[3,4,5], M. Benn[3,4,5], A. Tybjærg Hansen[5,6], G. Davey Smith[1,2], B.G. Nordestgaard[3,4,5].* 1) MRC CAiTE Centre, University of Bristol, Bristol, United Kingdom; 2) Department of Social Medicine, University of Bristol, Bristol, United Kingdom; 3) Department of Clinical Biochemistry, Herlev Hospital, Demark; 4) The Copenhagen General Population Study, Herlev Hospital, Demark; 5) Copenhagen University Hospital, Faculty of Health Sciences, University of Copenhagen, Denmark; 6) Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, Faculty of Health Sciences, University of Copenhagen, Denmark.

The relationship between body mass index (BMI) and ischaemic heart disease (IHD) is one of clinical importance, yet one only explored by observational analyses. The Prospective Studies Collaboration (Lancet 2010) investigating the association of BMI with health outcomes in >900000 individuals illustrates this, however fails to make causal estimates. An accepted approach to the assertion of causality in epidemiology is Mendelian randomisation. In three large studies (The Copenhagen General Population Study (CGPS) n=55495[4020 cases], The Copenhagen City Heart Study (CCHS) n=10461[2004 cases] and the Copenhagen Ischaemic Heart Disease Study (CIHS) n=15810[5270 cases]), we used genotypes born of GWAS and reliably associated with BMI to reassess BMI/IHD relationships. In CGPS and CCHS, for one standard deviation (SD) increase in BMI there were crude increases in the risk ratios (RR) for disease of 1.20(95%CI 1.17,1.23) and 1.23(95%CI 1.18,1.28). In the CCHS, this was reflected in a RR for all cause mortality (ACM) of 1.11(95%CI 1.07,1.14). Instrumental variable (IV) analysis using genotypes at FTO(rs9939609), TMEM18(rs6548238) and MC4R(rs17782313) to re-assess crude relationships yielded estimates of the causal effect of a BMI on IHD risk of RR1.31(95%CI 0.74,2.33) and 1.41(95%CI 0.96,2.08) for CGPS and CCHS. In agreement with observational results, meta-analysis of these (n>65000) delivered a best estimate for the causal effect of a SD change in BMI on the risk of IHD; RR1.38(95%CI 1.00,1.90). In the absence of BMI data in CIHS, the direct per allele associations between BMI associated genotypes and IHD risk were consistent with IV results and the size of deviation in BMI afforded by these genotypes (rs9939609 RR1.05(95%CI 1.00,1.90), rs6548238 RR1.04(95%CI 0.99,1.10), rs17782313 RR1.04(95%CI 0.98,1.11). In the context of large observational studies investigating the link between BMI and IHD, we show an independent and causal effect of elevated BMI on the risk of disease. The availability of reliable genotype/BMI associations and the application of Mendelian randomisation (yielding unconfounded, non-biased effects estimates free from reverse causation) has allowed us to show that for every SD(4kg/m2) increase in BMI, IHD risk rises by approximately 38%. This has important, policy implications for public health and is a direct application of findings made following the genomewide era.

283

Genome-wide joint meta-analyses of SNP by BMI interaction on levels of insulin and glucose-related traits: a MAGIC study. *A. Manning[1], M.-F. Hivert[2], R. Scott[3], H. Chen[1], C.-T. Liu[1], N. Bouatia-Naji[4], L. Beilak[5], L. Rasmussen-Torvik[6], J.C. Florez[7,12,13], I. Prokopenko[8], C. Langenberg[3], R. Watanabe[9,10], J. Dupuis[1,11], J.B. Meigs[12,13], MAGIC.* 1) Biostatistics Department, Boston University, Boston, MA; 2) Centre de Recherche Medicale de l'Universite de Sherbrooke, Sherbrooke, Quebec, Canada; 3) Medical Research Council (MRC), Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK; 4) Centre National de la Recherche Scientifique-Unité Mixte de Recherche 8090, Pasteur Institute, Lille 2-Droit et Santé University, Lille, France; 5) Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI; 6) Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL; 7) Broad Institute, Cambridge, MA; 8) Oxford Centre for Diabetes, Endocrinology and Metabolism and Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 9) Department of Physiology and Biophysics, Keck School of Medicine, University of Southern California, Los Angeles, CA; 10) National Institute of Health and Welfare, Oulu, Finland; 11) National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA; 12) Harvard Medical School, Boston, MA; 13) Massachusetts General Hospital, Boston, MA.

Background: Insulin resistance (IR) and beta cell dysfunction are hallmarks of type 2 diabetes (T2D). Genome-wide association studies conducted by MAGIC (Meta-Analysis of Glucose and Insulin-related traits Consortium) identified 16 loci associated with fasting glucose (FG) or beta cell function (HOMA-B), but only 2 associated with fasting insulin (FI) levels or insulin resistance (HOMA-IR). Body mass index (BMI) is a key risk factor both for T2D and IR; variation in BMI could obscure or enhance genetic IR signals. **Methods:** To account for environmental heterogeneity induced by BMI we conducted single nucleotide polymorphism (SNP) by BMI (SNP×BMI, BMI modeled as a continuous covariate) interaction regression models as a strategy to strengthen existing and discover novel IR genetic signals. We used ~2 million imputed and genotyped SNPs and applied a new method to jointly meta-analyze SNP and SNP×BMI regression coefficients across 29 cohorts comprising up to 58,074 non-diabetic individuals. The joint meta-analysis (JMA) concurrently estimated and tested SNP and SNP×BMI regression coefficients with a 2-df test of the joint effect. We further tested SNPs with JMA $p<10^{-6}$ in meta-analyses stratified by low or high BMI (BMI $<$ or $\geq$ 28 kg/m^2). **Results:** Associations between FI and SNPs in or near *IRS1* were not significant in main effects meta-analyses ($p=1.0\times10^{-4}$). However, the association between FI and the joint effect of SNP and SNP×BMI interaction was highly significant (JMA $p=1.7\times10^{-12}$) with a strong, significant additive SNP effect on FI in the low BMI group ($p=1.95\times10^{-8}$). Novel FI associations reaching genome-wide significance included SNPs in or near *COBLL1* (JMA $p=6.1\times10^{-14}$) and *PDGFC* (JMA $p=1.5\times10^{-8}$) . The joint meta-analysis strengthened association signals of two known genetic loci associated with FI: *GCKR* (JMA $p=3.3\times10^{-10}$) and *IGF1* (JMA $p=1.2\times10^{-9}$). Both genes show strong additive genetic effects on FI in the high BMI group (*GCKR* $p=1.3\times10^{-8}$; *IGF1* $p=9.2\times10^{-6}$). The joint meta-analysis strengthened association signals in 11 of 14 genetic loci with known associations with FG and HOMA-B. **Conclusions:** Genetic associations with insulin resistance and beta cell dysfunction are modified by BMI. Accounting for interaction with BMI reveals novel FI and FG signals and demonstrates that the joint meta-analysis approach improves detection of SNP associations when heterogeneity in environmental factors exist.

284

Gene dosage at the 16p11.2 locus controls body mass index. *A. Reymond[1], F. Zufferey[2], L. Harewood[1], Z. Kutalik[3], D. Martinet[2], J. Chrast[1], R.G Walters[4,5], S. Bouquillon[6], A. Valsesia[3], L. Hippolyte[2], J. Andrieux[6], B. Delobel[7], A.I.F. Blakemore[4], P. Froguel[4,8], S. Jacquemont[2], J.S. Beckmann[2,3], 16p11.2 consortium members.* 1) Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; 2) Service de Génétique Médicale, CHUV, Lausanne, Switzerland; 3) Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland; 4) Department of Genomics of Common Disease, Imperial College London, London, United Kingdom; 5) Department of Epidemiology and Public Health, Imperial College London, London, United Kingdom; 6) Laboratoire de Génétique Médicale, CHRU Lille, Lille, France; 7) Centre de génétique Chromosomique, Hopital Saint Vincent de Paul, GHICL, Lille, France; 8) CNRS, 8090 - Institut de Biologie, Institut Pasteur, Lille, France.

The 16p11.2 deletion has been associated with childhood-onset developmental disorders, macrocephaly and autism in multiple cohorts. We reported an association between this deletion and obesity, regardless of the presence of cognitive or behavioral symptoms. This highly penetrant form accounted for 0.7% of morbid obesity cases from 8 European cohorts (body mass index, BMI $\geq$ 40 kg.m-2; p = 6.4x10-8, OR = 43.0), demonstrating the potential importance of rare variants exerting strong effects in common disease. The reciprocal duplication was associated with schizophrenia, bipolar disorder and microcephaly. We hypothesized that the gene dosage effects accountable for obesity in carriers of the 16p11.2 deletion may influence, in a converse manner, the phenotype of duplication carriers. In search of these reciprocal manifestations, we investigated the impact of the 16p11.2 duplication on BMI. We assessed its distribution in 105 carriers of the duplication regardless of their ascertainment (63 cases ascertained among developmental delay cohorts, 6 among the general population, 1 within a psychiatric cohort, 12 were first degree relatives of the aforementioned probands and 23 cases were from the literature). The BMI of these patients was compared to the BMI of the general population. This comparison took into account influencing factors, such as gender, age and country of origin (split into northern Europe, central Europe and North America). The BMI distribution among carriers of the duplication was shifted towards underweight overall (p = 0.04). Stratification by age and gender showed that adult females were driving this difference (p=0.003). Notably, two female cases were diagnosed with anorexia nervosa. We used quantitative real-time PCR to accurately measure the expression levels of genes mapping to the 16p11.2 rearranged interval or its flanks in lymphoblastoid cell lines from deletion and duplication carriers and age- and sex-matched controls. The relative expression levels of the genes mapping to the commonly rearranged interval were generally decreased in the deletion samples, but increased (sometimes by more than 2 fold) in the duplication samples. To our knowledge, this work represents the first association between a rare genomic variant and low BMI. We hypothesize that gene dosage at the 16p11.2 locus controls food intake, giving rise to obesity or low BMI. In severe case, the same mechanism(s) may also play a role in anorexia.

285

Mapping complex traits using an integrated 'omics' approach in multiple tissues in twins: the MuTHER study. *A.K. Hedman[1], E. Grundberg[2,3], K. Small[2,3], A. Nica[3,4], J. Tzenova Bell[1,3], T-P. Yang[2], D. Glass[3], J. Nisbet[2], A. Barrett[5], A. Wilk[2], M. Travers[5], N. Hassanali[5], L. Parts[2], S-Y. Shin[2], R. Durbin[2], K.T. Zondervan[1], N. Soranzo[2,3], C.M. Lindgren[1], K. Ahmadi[3], E.T. Dermitzakis[4], P. Deloukas[2], M.I. McCarthy[1,5], T.D. Spector[3] for the MuTHER consortium.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, UK; 2) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SA, UK; 3) Department of Twin Research and Genetic Epidemiology, King's College London, Westminster Bridge Road, London, SE1 7EH, UK; 4) Department of Genetic Medicine and Development, University of Geneva Medical School, Rue Michel Servet 1, 1211, Geneva 4, Switzerland; 5) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Old Road, Oxford, OX3 7LJ, United Kingdom.

Integrative approaches are needed to disentangle the genetic and environmental contributions to complex trait susceptibility and to identify underlying molecular mechanisms. We have initiated the MuTHER (Multiple Tissue Human Expression Resource) project, a comprehensive study of SNP-genotyping, re-sequencing, mRNA and miRNA expression, methylation and metabolite profiling data from a range of tissues (adipose, skin, skeletal muscle, lymphocytes and lymphoblastoid cell lines, LCL) collected from up to 900 well-phenotyped female twins (1/3 MZ and 2/3 DZ) from the TwinsUK resource. The twin design will allow accurate estimate of the genetic contribution to both clinical phenotypes and gene expression.

We estimated heritability of gene expression per tissue for all expressed transcripts and found 2346 (16%), 2800 (19%) and 4821 (32%) transcripts having $h^2>0.3$ in LCL, skin and adipose tissue, respectively. To map the underlying regulatory genetic effect tissue-shared heritable transcripts ($h^2>0.5$ in all three tissues) were selected for genome-wide association analysis. Notable, 79% of these transcripts were associated with *cis*-acting variants at study-wise significance level ($P<10^{-12}$). For the remaining transcripts only one (MAPK8IP1; chr. 11) was associated with *trans*-acting variants across tissues. Focusing on associations between transcripts and obesity-related phenotypes, 8.5% of the measured transcripts in adipose tissue were associated with percent abdominal fat mass (PAFM) and 6% with BMI ($P<10^{-10}$), indicating the added value of having sensitive measurements of adiposity. Ontology analysis showed top PAFM-associated transcripts being enriched for the functional process lipid metabolism ($P<10^{-5}$). In addition, adipose transcripts associated with obesity-related phenotypes were more heritable than all transcripts expressed in adipose tissue ($P<5\times10^{-8}$). Preliminary results from *cis*-eQTL analysis of the top PAFM-associated transcripts in adipose tissue highlighted strong *cis*-regulatory effects of ~10% of the transcripts, suggesting alternative control of gene expression in the remaining transcripts. Analyses are currently extended to include transcriptomic/eQTL data from all tissues in combination with related clinical phenotypes and environmental factors. This together with epigenomic, miRNA and metabolomic data gathered within the project will greatly facilitate mapping the genetic architecture of complex traits.

286

GENOME WIDE ASSOCIATION ANALYSIS IDENTIFIES VARIANTS THAT ASSOCIATE WITH NONALCOHOLIC FATTY LIVER DISEASE AND THAT HAVE DISTINCT EFFECTS ON SERUM LIPIDS. *E.K. Speliotes[1,2], L.M. Yerges-Armstrong[3], J. Wu[4], R. Hernaez[5,6], L.J. Kim[7], C.D. Palmer[8], T.B. Harris[7], G. Eiriksdottir[9], M.E. Garcia[7], L.J. Launer[7], M.A. Nalls[10], J.M. Clark[5,6,11], B.D. Mitchell[3], A.R. Shuldiner[3,12], J.L. Butler[2,8], U. Hoffmann[13,14], J.M. Massaro[15,16], C.J. O'Donnell[14,15,17], D.V. Sahani[13], J.J. Carr[18], M.F. Feitosa[4], V. Gudnason[9,19], C.S. Fox[15,17], A.V. Smith[9], W.H.L. Koa[5,11], J.N. Hischhorn[2,8,20], I.B. Borecki[4], The GOLD Consortium.* 1) Dept Gastroenterology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA; 2) Broad Institute, Cambridge, Massachusetts 02142, USA; 3) Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine , Baltimore, Maryland 21201, USA; 4) Division of Statistical Genomics, Washington University, Saint Louis, Missouri 63108, USA; 5) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21287, USA; 6) Division of General Internal Medicine, The Johns Hopkins Hospital, Baltimore, Maryland 21287, USA; 7) Laboratory of Epidemiology, Demography, and Biometry, Intramural Research Program, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA; 8) Divisions of Endocrinology and Genetics and Program in Genomics, Children's Hospital, Boston, Massachusetts 02115, USA; 9) Icelandic Heart Association, Kopavogur IS-201, Iceland; 10) Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA; 11) Welch Center for Prevention, Epidemiology and Clinical Research, Baltimore, Maryland 21287, USA; 12) Geriatric Research and Education Clinical Center (GRECC), Veterans Administration Medical Center, Baltimore, MD; 13) Department of Radiology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA; 14) Division of Cardiology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA; 15) National Heart, Lung, and Blood Institute (NHLBI), Framingham Heart Study, Framingham, Massachusetts 01702, USA; 16) Department of Biostatistics, Boston University School of Public Health, Massachusetts 02118; 17) Division of Intramural Research, National Heart, Lung and Blood Institute, Bethesda, Maryland 20817, USA; 18) Departments of Radiologic Sciences, Internal Medicine-Cardiology and Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, North Carolina, 27157, USA; 19) University of Iceland, 101 Reykjavik, Iceland; 20) Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115.

Nonalcoholic fatty liver disease (NAFLD) is a spectrum of disease that ranges from hepatic steatosis, to steatohepatitis to fibrosis/cirrhosis and is associated with dyslipidemia. Liver steatosis can be non-invasively measured using computed tomography (CT) in the general population. Heritability studies of NAFLD in large populations have not been done and further, the only known common genetic variants influencing NAFLD risk are near PNPLA3. Here, we determine that the heritability of computed tomography (CT) measured hepatic steatosis ranges from 21-30% in the Amish, Family Heart and Framingham Heart Studies (n=541 to 2386) using variance component methods in SOLAR. We identify genome wide significantly (p<5x10-8) associated variants in or near PNPLA3, NCAN and PPP1R3B by carrying out a fixed-effects meta-analysis using METAL of genome-wide association (GWA) results between CT hepatic steatosis and ~2.4 million imputed or genotyped SNPs in an additive genetic model in 6,629 individuals from the Amish, Age, Gene/Environment Susceptibility-Reykjavik study (AGES), Family Heart and Framingham Heart Studies. PPP1R3B is a regulator of glycogen metabolism, PNPLA3 functions to break down triglycerides, and NCAN family members may play a role facilitating the hepatic uptake of lipoprotein cholesterol esters. The variance in CT hepatic steatosis cumulatively explained by these 3 variants is 4.64% or 15-22% of the genetic component of the trait. The hepatic steatosis increasing variant in NCAN decreases serum LDL-cholesterol (p= 4.86 x 10-8) and triglycerides (p= 9.64 x 10-11) and those near PPP1R3B increase LDL-cholesterol (p= 1.9 x 10-3) and HDL-cholesterol (p=3.51 x 10-6) in data from over 19,000 individuals from the Global Lipids Genetics Consortium. In this, the largest genetic study of NAFLD to date, we determine the heritability of CT measured hepatic steatosis in three family based cohorts, expand the number of common genetic variants that associate with NAFLD, and show that the effects of these variants on hepatic steatosis and serum lipids can be dissociated. The presence of common genetic variants that affect hepatic steatosis in or near PNPLA3, NCAN and PPP1R3B suggests that variation in lipid/glycogen handling may influence development of NAFLD in the general population. These results gives us new insights into the biology and genetics of NAFLD and open up many avenues for biological, diagnostic, and therapeutic research of this condition in humans.

287

Identification of genetic variants associated with stuttering: Using inbred pedigrees to find rare genetic variants of large effect. *D. Drayna[1], S. Riazuddin[2], D. Krasnewich[3], P. Friedman[4], C. Kang[1].* 1) NIDCD, National Institutes of Health, Rockville, MD; 2) Alama Iqbal Medical College, Lahore, Pakistan; 3) NHGRI, National Institutes of Health, Bethesda, MD; 4) Clinical Center, National Institutes of Health, Bethesda, MD.

Stuttering is a common speech disorder present in all populations and language groups. Although the causes of this disorder have been unknown, twin and family studies have demonstrated that it is highly heritable. Segregation analyses and linkage studies, however, have failed to identify Mendelian inheritance of stuttering. We hypothesized that stuttering is caused by rare genetic variants of large effect, and we studied large, highly consanguineous families from Pakistan as a method to efficiently reveal such variants. We identified significant linkage on chromosome 12q in this population, and a search of the genes within this region identified a Glu1200Lys mutation in GNPTAB, which encodes the catalytic subunit of GlcNAc phosphotransferase, in affected members of several families in our collection. This mutation was identified in unrelated individuals who stutter from both Pakistan and India, and other mutations in this gene were identified in other cases but not in controls. Based on this, we evaluated the GNPTG gene, which encodes the recognition subunit of this enzyme. We identified a number of mutations in this gene in individuals who stutter that did not occur in matched controls. We then examined the NAGPA gene, which encodes the enzyme that acts immediately downstream of GlcNAc phosphotransferase in the same metabolic pathway, and identified a number of mutations in unrelated cases that did not occur in controls. These two enzymes act to generate the Mannose-6-phosphate lysosomal targeting signal that directs a diverse group of hydrolases to the lysosome. Deficits in GNPTAB and GNPTG have been associated with the mucolipidosis types II and III, which are rare recessive disorders with manifestations in the connective tissue, skeletal system, brain, liver, and spleen. No disorder has previously been associated with mutations in NAGPA. Detailed clinical examination of four stuttering subjects carrying mutations in these genes revealed no symptoms associated with mucolipidosis types II and III, and other than stuttering, these individuals were neurologically normal. We hypothesize that the mutations we've identified affect a distinct group of neurons in the brain that are uniquely dedicated to speech production and uniquely sensitive to the metabolic deficit caused by these mutations, and that inbred families present a generally powerful strategy for identifying rare mutations of large effect that underlie common complex disorders.

288

A potassium channel gene modulates natural variation in sleep duration in human populations. *K. Allebrandt[1], N. Amin[2], B. Müller-Myhsok[3], R. van der Maas de Azevedo[4], M. Teder-Laving[5], T. Esko[5], C. Hayward[6], J. van Mill[7], N. Vogelzangs[7], S. Melville[8], E. Wichmann[9], H. Campbell[10], J. Wilson[10], A. Hicks[8], P. Pramstaller[8], I. Rudan[11], M. Merrow[12], B. Penninx[7], C. Kyriacou[4], A. Metspalu[5], C. van Duijn[2], T. Meitinger[13], T. Roenneberg[1].* 1) Dept of Medical Psychology, University of Munich, Munich, Germany; 2) Dept of Epidemiology, ERASMUS MC, Rotterdam, The Netherlands; 3) Max-Planck-Institute of Psychiatry, Munich, Germany; 4) Dept of Genetics, University of Leicester, Leicester, UK; 5) Estonian Genome Project, Institute of Molecular and Cell Biology of University of Tartu and Estonian Biocentre, Tartu, Estonia; 6) IGMM, Medical Research Council, Human Genetics Unit, Edinburgh, Scotland; 7) Dept of Psychiatry, VU University Medical Center Amsterdam, The Netherlands; 8) Institute of Genetic Medicine, European Academy of Bolzano, Bozen, Italy; 9) Institute of Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany; 10) Centre for Population Health Sciences, University of Edinburgh, Edinburgh, Scotland; 11) Croatian Centre for Global Health, University of Split Medical School, Split, Croatia; 12) Dept of Chronobiology, University of Groningen, Haren, The Netherlands; 13) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany.

Average sleep duration shows a normal distribution in human populations, with extreme short and long sleepers lying on either side of its tails. Individual sleep duration changes throughout life, is gender- and body mass index-dependent, but inter-individual differences can still be very large after accounting for these confounding factors. To identify genetic factors regulating the predisposition towards being a long or short sleeper, we conducted a large-scale genome-wide association study for average sleep duration, comprising 4,260 subjects from 7 cohorts. Meta-analysis of these data revealed a significant association (P = 3.68 x 10-8) with an intronic variant of a gene encoding a potassium channel protein. Knocking down the expression of a conserved homologue of this gene pan-neuronally in *Drosophila* reduced exclusively sleep duration during the night, but had no effect on *Drosophila*'s day-time sleep. Our results indicate that natural variation in K+ channel mediated membrane excitability contributes to the modulation of sleep duration in general human populations.

289

A genetic network model of cellular responses to lithium treatment and cocaine abuse in bipolar disorder. *R. McEachin[1,2], H. Chen[1], M. Sartor[2], S. Saccone[3], B. Keller[4], A. Prossin[1], J. Cavalcoli[2], M. McInnis[1,2].* 1) Department of Psychiatry, Univ Michigan, Ann Arbor, MI; 2) National Center for Integrative Biomedical Informatics, Univ Michigan, Ann Arbor, MI; 3) Department of Psychiatry, Washington University, Saint Louis, MO; 4) Department of Computer Science, Eastern Michigan University, Ypsilanti, MI.

Background: Lithium is an effective treatment for Bipolar Disorder (BD) mania in ~70% of patients, significantly reducing suicide risk, though the molecular basis of lithium's effectiveness is not well understood. We seek to improve our understanding of this effectiveness by posing hypotheses based on new experimental data as well as published data, testing these hypotheses in silico, and posing new hypotheses for validation in future studies. Since lithium poses an environmental influence on cells, we initially hypothesized a gene-by-environment interaction, where lithium impacts signal transduction pathways leading to differential expression of genes important in the etiology of BD mania. Results: Using microarray and reverse transcription polymerase chain reaction assays, we identified candidate genes that are significantly differentially expressed with lithium treatment. We used a systems biology approach to identify interactions among these candidate genes and develop a network of genes that interact with the differentially expressed candidates. Notably, we also identified cocaine as having an important influence on the network, consistent with the observed high rate of comorbidity for BD and cocaine abuse. The resulting network represents a novel hypothesis on how multiple genetic influences on bipolar disorder are impacted by both lithium treatment and cocaine use. Testing this network for association with BD and related phenotypes, we find that it is significantly over-represented for genes that participate in signal transduction, consistent with our hypothesized-gene-by environment interaction. In addition, it models related pharmacogenomic, psychiatric, and chemical dependence phenotypes. Conclusions: We offer a network model of gene-by-environment interaction associated with lithium's effectiveness in treating BD mania, as well as the observed high rate of comorbidity of BD and cocaine abuse. We identified drug targets within this network that represent immediate candidates for therapeutic drug testing. Posing novel hypotheses for validation in future work, we prioritized SNPs near genes in the network based on functional annotation. We also developed a "concept signature" for the genes in the network and identified additional candidate genes that may influence the system because they are significantly associated with the signature.

290

A polymorphic dinucleotide repeat in the 5'- UTR of *DPYSL2* confers risk for schizophrenia. *Y. Liu[1], G. Burzynski[1], D. McGaughey[1], S. He[3], J. McGrath[2], P. Wolyniec[2], D. Fallin[4], M. Szymanski[1], D. Avramopoulos[1], A. McCallion[1], A. Pulver[2], D. Valle[1].* 1) Inst. Genetic Medicine; 2) Dept. of Psychiatry; 3) Dept. of Mol. Biol. and Genetics Johns Hopkins University School of Medicine; 4) Dept. of Epidemiology Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

Linkage and association results from our group performed on European Caucasian (CEU) and Ashkenazi Jewish (AJ) samples; and from others on CEU and Japanese (JAN) samples implicate *DPYSL2*, at 26.5Mb on chromosome 8p, as a candidate gene for schizophrenia (SZ). *DPYSL2* (a.k.a. *CRMP-2*) encodes a 62kDa cytosolic protein that regulates axonal growth. To identify additional variants in and around *DPYSL2*, we sequenced all 14 exons and 27 conserved non-coding regions (cNCRs) in and around *DPYSL2* in 48 CEU and 89 AJ SZ probands plus 56 AJ and 96 CEU controls. We identified a 3_SNP-haplotype (rs367948, rs400181, rs445678; pairwise r^2=0.97) in the proximal promoter region and a polymorphic dinucleotide repeat (DNR, rs3837184; r^2 with 3_SNP-haplotype: 0.91) in the 5'-UTR that were associated with SZ in the CEU sample (allele freq: 11 DNRs 0.7 in SZ, 0.85 in control; 13 DNRs 0.21 in SZ, 0.11 in control) (p-values ranging 0.003-0.012). To study the regulatory potential of the human sequence containing these variants, we utilized zebrafish (ZF) transgenesis. The human sequence directed pervasive expression of the reporter gene in the ZF CNS in a pattern largely overlapping with expression of the two ZF *dpysl2* orthologs. We next assessed the biological relevance of these 4 SZ-associated sequence variants using transient transfection and dual luciferase assays in HEK293 cells and primary E14.5 mouse cortical neurons. The risk 3_SNP haplotype with the control DNR moderately reduced expression while the risk DNR with the protective 3_SNP haplotype resulted in ~3.5X decreased expression in both primary E14.5 cortical neurons and 293 cells (P< 0.0001). We also asked if these variants were associated with altered *DPYSL2* expression in vivo using RT-qPCR of temporal lobe RNA from 200 control brains and found no correlation between the DNR genotypes and *DPYSL2* mRNA expression. To determine if the polymorphic 5'-UTR DNR influenced *DPYSL2* translation, we performed sucrose density gradient fractionation of cytoplasmic extracts of HEK293 cells transfected with luciferase constructs with the 5'-UTRs containing various DNR numbers. We found the fraction of mRNA in polysome with the SZ-risk 13 DNRs was reduced by 3.6 fold as compared to the control with 11 DNRs. Our results extend the evidence for a role of *DPYSL2* in SZ and identify an associated causative variant, a polymorphic DNR in the 5'-UTR of *DPYSL2*, that results in decreased translation of *DPYSL2* mRNA.

291

A novel locus associated with narcolepsy. *B.R. Kornum[1,2], M. Kawashima[1,3], J. Faraco[1], L. Lin[1], J. Hallmayer[1], E. Mignot[1].* 1) Center for Sleep Sciences and Department of Psychiatry, Stanford University School of Medicine, Palo Alto, CA; 2) Danish Center for Sleep Medicine, University of Copenhagen, Glostrup Hospital, Glostrup, Denmark; 3) Department of Human Genetics, University of Tokyo, Tokyo, Japan.

The human sleep disorder narcolepsy-cataplexy affects 1 in 2,000 individuals. Narcolepsy symptoms are primarily caused by the loss of approximately 70,000 hypocretin producing neurons located in the hypothalamus, and growing evidence supports the hypothesis that narcolepsy with cataplexy is an autoimmune disease targeting these neurons. Following on a recently published GWA study of 807 narcoleptic patients versus 1,074 DQB1*0602 positive controls, we conducted replication of 10 additional loci in 1,525 Caucasians (594 cases and 931 controls). Of these, only one loci replicated strongly ($p= 5.42x10^{-4}$; odds ratio 0.77). Replication was next attempted in other ethnic groups, but the original SNP had no effect in Asians or African Americans. Based on differential LD patterns for this marker across ethnic groups 6 SNPs were genotyped in 3,406 Caucasians (1,401 patients and 2,005 controls), 2,414 Asians (1,130 patients and 1,284 controls), and 302 African Americans (113 patients and 189 controls). A second linked SNP showed the highest association with narcolepsy in Caucasians and across all ethnic groups ($p= 6.1x10^{-10}$; odds ratio 1.28; associated allele frequency 0.664 in 2,522 cases, 0.603 in 3,167 controls). These findings identify a novel narcolepsy susceptibility factor and illustrate the value of transethnic mapping. Additional experiments indicate that this SNP modulates expression of a linked gene in white blood cells, most notably in $CD8^+$ T cells and NK cells where the disease associated allele show a significantly lower expression. We hypothesize that since this gene is involved immune cell chemotaxis, this finding reflects the importance of cell migration in brain-related autoimmune diseases, in view of the relative isolation of the brain from the peripheral immune system. We thank our other collaborators not listed here for contributing samples, cohort genotypes and participating in the genetic analysis. Funded by NS23724.

292

Genetic architecture of open-angle glaucoma and related outcomes. *C.M. Van Duijn[1], W.D. Ramdas[1,2], N. Amin[1], L.M.E. van Koolwijk[1,3], A.C.J.W. Janssens[1], A. Demirkan[1], P.T.V.M. de Jong[4,5], Y.S. Aulchenko[1], R.C.W Wolfs[1,2], A. Hofman[1], F. Rivadeneira[1,6], A.G. Uitterlinden[1,6], B.A. Oostra[7], H.G. Lemij[3], C.C.W. Klaver[1,2], J.R. Vingerling[1,2], N.M. Jansonius[1,8].* 1) Epidemiology & Biostatistics, Erasmus MC, Rotterdam, Netherlands; 2) Department of Ophthalmology, Erasmus Medical Center, Rotterdam, the Netherlands; 3) Glaucoma service, The Rotterdam Eye Hospital, Rotterdam, the Netherlands; 4) Department of Ophthalmogenetics, The Netherlands Institute for Neuroscience, RNAAS, Amsterdam, the Netherlands; 5) Department of Ophthalmology, Academic Medical Center, Amsterdam, the Netherlands; 6) Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands; 7) Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, the Netherlands; 8) Department of Ophthalmology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands.

Although various rare variants have been implicated in familial forms of open-angle glaucoma (OAG), the genetic origin of OAG remains for a large part unexplained. Genome-wide association (GWA) studies of endophenotypes have proven to be powerful to find common genetic variants with small effects on disease risks. There are several endophenotypes for research of genes involved in OAG including the vertical cup-to-disc ratio (VCDR) and intraocular pressure (IOP). Earlier, we found and replicated six loci (9q21, 14q22-23, 10q21.3-q22.1, 11q13, 13q13, and 22q12.1) that explain 2.4 percent of the VCDR variance and 1 locus (17p13) that explains 1 percent of IOP variance. Three of these genome wide significant loci were marginally significantly associated with OAG. In this study, we aim to explore the genetic architecture of OAG and OAG related endophenotypes further using genome wide profiling. This approach was successfully used to identify a polygenic mode of inheritance underlying both schizophrenia and bipolar depression. We used the GWA of VCDR and IOP conducted in the Rotterdam Study as discovery set. This study includes 10,972 Caucasian persons aged > 45 years. Different profiles were created based on p-values of SNPs in the discovery set. Some are based on genome wide significant ($p<5x10-8$) or borderline significant SNPs ($p<10-6$), while others include large numbers of non significant SNPs ($p>0.50$). The latter aim to evaluate polygenic effects. The risk scores were applied in an independent target study, the Erasmus Rucphen Family study (ERF), consisting of 1,646 participants (mean age: 46.8 years). In ERF, we calculated the explained variance in outcomes of interest based on the Rotterdam study profiles. We found that the explained variance of VCDR in the target set increased and remained significant when increasing the number of (non-significant) SNPs in the risk scores. This implies a large number of SNPs are involved in VCDR, each with a small effect on the risk of disease. The profiles based on VCDR were also significantly associated with OAG, suggesting the genetic etiology of VCDR and OAG at least in part overlap. We found no significant evidence for a polygenic mode of inheritance for IOP. A small, distinct number of SNPs may determine IOP. Our study suggests an overlapping polygenic background for VCDR and OAG and implies that VCDR is a useful endophenotype for the discovery of OAG genes.

293

Gene-gene interactions associated with Alzheimer disease replicate in independent Alzheimer Disease Genetics Consortium datasets. *T. Thornton-Wells[1], E.S. Torstenson[1], S.D. Turner[1], S.M. Dudek[1], W.S. Bush[1], M.D. Ritchie[1], M.A. Pericak-Vance[2], J.L. Haines[1], The Alzheimer's Disease Genetics Consortium.* 1) Center for Human Genetics Research, Molecular Physiology and Biophysics, School of Medicine, Vanderbilt University, Nashville, TN; 2) John P. Hussman Institute for Human Genomics, Miller School of Medicine, Unversity of Miami, Miami, FL.

Alzheimer disease is the leading cause of dementia in the elderly and has a complex genetic etiology, involving heterogeneity and gene-gene interactions. Recent genome-wide association studies (GWAS) have led to the discovery of novel genetic risk factors; however, the investigation of gene-gene interactions in GWAS has been limited. We conducted a gene-gene interaction analysis in three datasets (one for discovery; two for replication) from the Alzheimer's Disease Genetics Consortium (ADGC) funded by the National Institute on Aging. We first analyzed SNP data from Illumina Human660W BeadChip ('Miami-Vanderbilt' ADGC dataset) and then attempted to replicate our findings in two independent ADGC GWAS datasets ('TGen' and 'Mayo' ADGC datasets). We used a biological knowledge-driven approach (Biofilter software; Bush et al., Pacific Symposium on Biocomputing 14:368-79,2009) to select SNP-SNP models for which there was a priori evidence that their respective genes interact or participate in common biological pathways or processes (Biofilter implication index >= 3; SNPs known to be in LD with APOE were excluded). We analyzed the over 25 million SNP-SNP models using the parallel multifactor dimensionality reduction (pMDR) method and selected all models with a testing balanced accuracy (BA) ≥60% for replication testing, resulting in 1448 models. Because our replication datasets used the Affymetrix Genome-Wide Human SNP Array 5.0, which has a different selection of SNPs from the Illumina platform, we first mapped RefSeq numbers back to their genomic positions and then to all genes within 50kb of those SNPs. Using Biofilter, we generated all possible combinations of SNP-SNP models from that list of genes, and we analyzed those models using pMDR in the TGen and Mayo datasets separately. We then selected all models with a testing BA≥60% in at least 2 of 3 datasets. Two gene-gene models fit this criterion across all 3 datasets, each with an average testing BA=60%: (1) NMDA receptors GRIN2A and GRIN2B and (2) sortilin-related receptors SORCS1 and SORCS2. Seventeen additional gene-gene models replicated in 1 dataset. Some genes appeared in more than one model (e.g., ERBB4 was in 2 models, one with EGFR and the other with INSR), such that 28 unique genes were involved in these replicated interaction models. Efforts are ongoing to examine the LD structure of the implicated SNPs in CEU HapMap data and to provide a full interpretation of results.

294

What can GWAS association tell us about rare variants in autism? *L. Weiss.* Dept Psychiatry, Univ California, San Francisco, San Francisco, CA.

The advent of genome-wide association studies (GWAS) brought great excitement to the field of human genetics. GWAS efforts, including those in autism, have revealed mostly associations with modest effects which do not have an obvious pathological mechanism or copy number variants (CNVs) that are individually difficult to assign pathological significance. However, these results have the potential to provide critical clues into the underlying biology and genetic architecture of autism. Here, we have followed up a modest GWAS result with eQTL mapping in a control population which revealed an extensive expression regulation network. Primary eQTLs in controls were identified in *cis* (12 SNPs $P < 10^{-4}$, $P_{min} = 10^{-12}$) and in *trans* (>900 SNPs in ~230 regions with $P < 10^{-4}$). Secondary eQTLs (eQTL^{2}s; master regulators controlling >3 primary eQTLs) were identified in 14 regions. Common polymorphism within this expression network (excluding the primary GWAS-identified locus) is strongly associated with autism *per se* (eQTLs genome-wide $P < 0.005$; eQTL^{2}s genome-wide $P < 10^{-4}$). We have also identified rare copy number variants and nucleotide mutations in autism families at eQTLs and eQTL^{2}s. Therefore, our results provide a context to other autism genetic results by implicating rare variants within this network as likely to be pathogenic and common polymorphisms likely to influence expression traits. This network provides a biologically-relevant subtype of autism in which pathogenic mechanism and phenotypic commonalities can be investigated and translational potential for diagnosis and treatment can be explored.

295

A two-stage genomewide scan for common alleles affecting risk for autism. *J. Buxbaum for the Autism Genome Project Consortium (AGP).* Psychiatry, Mount Sinai Sch Med, New York, NY.

While autism spectrum disorders (ASDs) have a substantial genetic basis, most of the known genetic risk has been traced to rare variants, principally copy number variants (CNVs). To identify common risk variation, the Autism Genome Project (AGP) Consortium has carried out two staged analyses of ASD trios. In the first stage, the AGP genotyped 1,558 rigorously defined ASD families for one million single nucleotide polymorphisms (SNPs) and analyzed these SNP genotypes for association with ASD. In four primary association analyses, the association signal for marker rs4141463, located within MACROD2, crossed the genome-wide association significance threshold of P<5x10-8. When a smaller replication sample was analyzed, the risk allele at rs4141463 was again over-transmitted; yet, consistent with the winner's curse, its effect size in the replication sample was much smaller; and, for the combined samples, the association signal barely fell below the P<5x10-8 threshold. Exploratory analyses of phenotypic subtypes yielded no significant associations after correction for multiple testing. They did, however, yield strong signals within several genes, PLD5, ST8SIA2, POU6F2, and in KIAA0564. These proteins are involved in neuronal development, including synapse formation, playing roles broadly similar to those played by proteins previously implicated on the basis of rare variants. In the second stage, the AGP genotyped an additional 1,301 ASD families and analyzed these SNP genotypes for association with ASD. Analyses of this cohort and of the combined cohort will be completed within 8 weeks and will be reported. As all samples are also being assessed for copy number variation (see the abstract of D Pinto et al. and the AGP, this meeting) and >500 samples are being assessed for whole-exome sequencing (see the abstract of C Stevens et al., this meeting), we will ultimately look at genetic risk for ASD from multiple levels in the same samples. In addition, as there is detailed phenotypic information associated with the AGP samples, we have the opportunity to examine genotype-phenotype relationships.

296

Copy Number Variants Segregate in Extended Families with Autism Spectrum Disorder. *D. Salyakina[1], H.N. Cukier[1], D. Ma[1], J.M. Jaworski[1], M.L. Cuccaro[1], J.R. Gilbert[1], S.M. Williams[2], R.K. Menon[3], M.A. Pericak-Vance[1].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 3) Department of Epidemiology and Department of Obstetrics & Gynecology, Rollins School of Public Health, Emory University, Atlanta, GA.

Within the past decade, copy number variations (CNVs) have been revealed to be a major cause of genetic variation with far more nucleotides being altered by duplications and deletions than by single nucleotide polymorphisms. In the multifaceted etiology of autism spectrum disorder (ASD), CNVs must now be incorporated into our understanding of this complex disease. We sought to take advantage of our unique resource of 46 extended ASD families to detect potential CNVs that segregate to all autism family members and thus may contribute to ASD susceptibility. Each family had from two to four affected individuals with sibling, avuncular and cousin pairs. Using the genotyping of over 1 million sites from the Illumina Human genotyping array, we applied the PennCNV algorithm to recognize deletions and duplications. Families were then evaluated for co-segregation of CNVs in ASD patients. We were able to identify and validate nine deletions and seven duplications that were segregating in multiple affected individuals within 13 extended families. Eight genes were disrupted by 6 of these CNVs; deletions were identified on chromosomes 1p34.1 (ZSWIM5), 4q31.3 (LRBA) and 6q11.1(KHDRBS2) and duplications were located on 4p16.3 (TNIP2), 7p21.2 (ICA1, NXPH1) and 10q 23.2 (BTAF1,FGFPB3). Our results support the hypothesis that neuronal cell adhesion molecules are involved in ASD etiology. The duplication located on chromosomes 7p21.2 implicates the NXPH1 (neurexophilin 1) gene. The NXPH1 protein interacts with neurexin 1α, which has been shown to be mutated in autistic patients. In addition, our CNV findings overlap with larger, previously reported genomic rearrangements in patients with ASD, ADHD, developmental delay, mental retardation, and various congenital defects at 3p26.3-p26.2, 3q26.1, 4p16.3, 10p12, 12q24.31àqpter, 13q , 15q11.2-q.13.3. Our data support the hypothesis that the misregulation of genes either by altering their dosage, affecting their regulatory elements, or creating novel functional regions, may be an important genetic mechanism in ASD.

297

Molecular identification of a canine model for X-linked centronuclear myopathy. *J. Bohm[1], A.H. Beggs[2], E. Snead[3], M. Kozlowski[2], M. Maurer[4], K. Minor[5], M.K. Childers[6], S.M. Taylor[3], C. Hitte[7], L.T. Guo[8], A.P. Mizisin[8], A. Buj-Bello[9], L. Tiret[4], J. Laporte[1], G.D. Shelton[8].* 1) IGBMC, Strasbourg, France; 2) The Manton Center for Orphan Disease Research at Children's Hospital and Harvard Medical School, Boston, USA; 3) Western College of Veterinary Medicine, University of Saskatchewan, Canada; 4) Ecole Nationale Vétérinaire d'Alfort, Maisons-Alfort, France; 5) School of Veterinary Medicine, University of Minnesota, USA; 6) Wake Forest University, Winston-Salem, USA; 7) Institut de Génétique et Développement, Rennes, France; 8) School of Medicine, University of California, San Diego, La Jolla, USA; 9) Génethon, Evry, France.

Centronuclear myopathies (CNM) are a group of congenital disorders characterized by hypotonia and skeletal muscle biopsies typically showing small rounded fibers with central nuclei. The most severe form is linked to mutations in the phosphoinositides phosphatase myotubularin (MTM1), milder etiopathologies are associated to mutations in BIN1 and DNM2. There is no therapy to date and animal models are a prerequisite to test potential approaches. We report here the molecular identification of a canine model for X-linked centronuclear myopathy. Seven male Labrador retrievers, from 14 to 26 weeks of age, were clinically evaluated for generalized weakness and muscle atrophy. Cryostat sections and ultrastructural studies on muscle biopsies showed variability in fiber size, centrally placed nuclei resembling fetal myotubes, abnormal perinuclear structure and mitochondrial accumulations. Triads were infrequent with an abnormal orientation of T-tubules and immunofluorescence staining using antibodies against T-tubules (DHPRα1) and the adjacent sarcoplasmic reticulum (RYR1) confirmed an abnormal distribution of these structures. DNA analysis of the exonic sequences from the MTM1 gene revealed a unique variant in exon 7 in all affected males causing the non-conservative missense change c.465C>A (N155K) in the linker region between the GRAM-PH and phosphatase domains of myotubularin. Haplotype analyses suggest a derivation from a recent founder in the local population and sequencing of a world-wide panel of 237 unrelated and unaffected Labrador retrievers, and 59 additional control dogs from 25 breeds, failed to identify this variant, strongly indicating it is the pathogenic mutation responsible for the canine myopathy. Immunoblot and immunofluorescence on muscle extracts from affected dogs showed an abnormal myotubularin protein level and localization and expression of GFP-MTM1p.N155K in COS1 cells revealed that the mutant protein was sequestered in proteasomes where is was presumably miss-folded and prematurely degraded. These data are not only important for the scientific understanding of the pathomechanism, but also demonstrate that XLMTM in Labrador retrievers is a faithful genetic model of the human condition and represents the only large animal model available for preclinical trials of potential therapies.

298

Functional analysis of *AFG3L2* mutations causing spinocerebellar ataxia type 28 (SCA28). *V. Fracasso[1], S. Magri[1], F. Lazzaro[2], M. Plumari[1], C. Gellera[1], C. Mariotti[1], P. Plevani[2], M. Muzi-Falconi[2], D. Di Bella[1], F. Taroni[1].* 1) Genet Neurodegen Metab Dis, IRCCS Ist Neurol Carlo Besta, Milan, MI, Italy; 2) Dept. Biomolecular Sciences and Biotechnology, University of Milan, Italy.

Autosomal dominant spinocerebellar ataxias (SCA) are a heterogeneous group of neurological disorders characterized by cerebellar dysfunction. We recently showed that *AFG3L2* mutations cause dominant ataxia SCA28 (Di Bella *et al*, 2010). AFG3L2 and its partner protein paraplegin, which causes recessive spastic paraparesis SPG7, are components of the *m*-AAA complex, a highly-conserved protease involved in mitochondrial protein quality control, degradation of damaged inner-membrane proteins, and protein maturation. We used an *m*-AAA-deficient yeast cellular model to evaluate the functional effects and the pathogenic role of the *AFG3L2* mutations identified in patients with spinocerebellar ataxia. We expressed 22 different *AFG3L2* variants (20 missense mutations and 2 small in-frame deletions). Most of the AFG3L2 mutations are located in the ATPase or in the protease functional domains of the protein involving highly conserved residues, while 3 variants have been identified in the poorly-conserved N-ter region. Expression of mutant AFG3L2 homocomplex in *m*-AAA-deficient yeast cells demonstrate that the mutations located in the functional domains cause respiratory deficiency and defective processing of *m*-AAA substrates, while the 2 missense mutations in the N-ter region do not show any defective phenotype, suggesting that they are rare polymorphic variants. Moreover, functional studies allow to classify AFG3L2 mutants in two different groups according to the phenotype induced by coexpression of wild-type AFG3L2 or paraplegin. In a first group of mutants, coexpression of paraplegin does not rescue the defective phenotype, while coexpression of $AFG3L2^{WT}$ results in reduced growth rate, thus indicating a dominant-negative mechanism for the mutation. These mutations are expected to be highly penetrant, with several affected subjects in the families. Interestingly, the majority of mutations were found to be "paraplegin-responsive" as paraplegin coexpression is able to rescue the defective phenotype. Furthermore, coexpression of $AFG3L2^{WT}$ normalizes the growth rate of $AFG3L2^{MUT}$-harbouring cells. In these cases, the mechanism is likely to be haploinsufficiency or a weak dominant negative effect, which would result in variably reduced penetrance and/or expressivity in affected people. In conclusion, this study expands the spectrum of *AFG3L2* mutations and indicates that they can act through distinct pathomechanisms at the molecular level. [Telethon grant GGP09301 to FT].

299

Lysosonal targeting of CLN7, a Major Facilitator Superfamily transporter underlying variant late-infantile neuronal ceroid lipofuscinosis. *M. Kousi[1], A. Sharifi[2], C. Sagne[2], GC. Bellenchi[3], M. Darmon[4], R. Ruivo[2], C. Debacker[2], AE. Lehesjoki[1], A. Jalanko[5], B. Gasnier[2], A. Kyttala[5].* 1) Folkhalsan Institute of Genetics, Dept of Medical Genetics and Neuroscience Center, Univ Helsinki, Helsinki, Finland; 2) Institut de Biologie Physico-Chimique, Centre National de la Recherche Scientifique UMR 8192, Universite Paris Descartes, Paris, France; 3) Istituto di Genetica e Biofisica "Adriano Buzzati Traverso", CNR, Napoli, Italy; 4) INSERM U894, Centre de Psychiatrie et Neuroscience, Paris, France; 5) National Institute for Health and Welfare (THL) Biomedicum Helsinki and FIMM, Institute for Molecular Medicine Finland, Finland.

The neuronal ceroid lipofuscinoses (NCLs) comprise a group of neurodegenerative disorders involving epilepsy, psychomotor decline and visual loss. NCLs are the most common neurodegenerative disease encountered in childhood, although rare adult cases have also been described. To date defects in eight different genes are known to cause NCLs (*CLN1-CLN3*, *CLN5-CLN8*, *CLN10*). *CLN7* is the most recently identified NCL gene, underlying a variant form of the disease (vLINCL). Localization studies and sequence homology analyses have shown that CLN7 is a polytopic protein acting in the lysosomes as a transporter protein of unknown substrate specificity. In this study we show that of the cell types in the central nervous system CLN7 has highest expression in neurons. Moreover, it is more abundant in hippocampus compared to other brain regions.. To characterize trafficking of CLN7, CD8-CLN7 chimeric and full-length CLN7 constructs carrying mutations at all the putative lysosomal targeting motifs were constructed. An N-terminal dileucine motif (9-EQEPLL-14) at the N-terminal cytosolic tail of CLN7 was identified as the major, but non-exclusive determinant of the protein's lysosomal targeting. This motif was found to bind specifically to heterotetrameric AP-1 adaptor in GST pull-down experiments. In endocytosis blocking experiments only a minor fraction of the protein was found at the plasma membrane, providing further evidence that the lysosomal targeting of CLN7 occurs mainly via a direct trafficking route. The present study provides primary information for redirection of CLN7 to the plasma membrane where it could be used to study the precise transporter function of the protein and thus shed light into the pathways affected in this severe neurological disease.

300

The non-coding RNA Vax2OS1 is involved in mouse retinogenesis. *N. Meola[1], M.T. Pizzo[1], G. Alfano[2], E.M. Surace[1], S. Banfi[1].* 1) TIGEM, Telethon Inst Gen/Medicine, Naples, Italy; 2) Molecular Genetics Dept, University College London, London, UK.

Regulatory "noncoding RNAs" (ncRNAs) represent a significant component of eukaryotic genomes. In particular, long ncRNAs, which are both spliced and polyadenylated, are emerging as central players in the regulation of gene expression at multiple levels. A large proportion of these transcripts correspond to intergenic "opposite strand" transcripts (OSTs), which are localized in the vicinity of developmentally regulated genes. However the role of such transcripts is still far to be elucidated. Here, we report on the functional analysis of an intergenic transcript, *Vax2OS1*, which is transcribed in the opposite direction with respect to its neighboring gene, *Vax2*, a homeobox gene involved in mouse eye development. The Vax2OS transcriptional unit is characterized by the presence of three main spliced isoforms, each of which lacks obvious coding potentials, and by an expression pattern mostly restricted to the ventral part of the developing and adult mouse retina. In the present study, the analysis of the temporal expression of all the three Vax2OS isoforms reveales that the Vax2OS isoform 1 (*Vax2OS1*) displays the highest levels of expression during retina development in the retinal progenitor cell layer and in the layer of mature photoreceptor cells in adult mouse retina. Transient overexpression of *Vax2OS1* in mouse photoreceptor-like immortalized cells leads to a delay of differentiation processes. More in detail, we observe an enrichment of *Vax2OS1*-overexpressing cells treated for 48 hours in both the S and G2 phases of the cell cycle. We suggest that these cells may undergo a prolongation of the S and the G2 phases as determined also by, respectively, BrdU (5-bromo-2-deoxyuridine) studies and phospho-Histone H3 staining. Furthermore, in vivo delivery of *Vax2OS1* in the developing postnatal mouse retina by Adeno-Associate Viral (AAVs) constructs causes a delayed differentiation of photoreceptor progenitor cells. In conclusion, our studies suggest that the misexpression of *Vax2OS1* may cause a delay of cell cycle exit of progenitor photoreceptor cells toward their final differentiation. To the best of our knowledge, this is the first example of a long non-coding RNA putatively involved in the fine regulation of cell cycle of photoreceptor progenitors.

301

Evaluation of *RET* transcription modulation by estrogen in vitro and in vivo. *Z.E. Stine[1], D.M. McGaughey[1], S. Maragh[1], S.L. Bessling[1], L. Taher[3], I. Ovcharenko[3], A.S. McCallion[1,2].* 1) McKusick - Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205; 2) Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine, Baltimore, MD 21205; 3) Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA.

Estrogen mediated transcriptional modulation plays an important role in development and disease, notably in breast cancer. Physical interaction among regulatory elements is also proposed to contribute to transcriptional control, supported by the recent efforts to map estrogen mediated chromatin interactions. *RET*, a receptor tyrosine kinase critical for normal renal and neuronal development is implicated in multiple tumor types, and is one of the hundreds of genes reported to be estrogen modulated. However, the transcriptional regulatory sequences responsible for the estrogen modulation of *RET* have yet to be described. Several whole genome chromatin immunoprecipitation experiments have recently identified putative Estrogen Receptor α (ESR1) binding sites within and flanking the *RET* locus. We completed *in vitro* analyses of sequences encompassing the putative ESR1 binding sites in MCF-7 estrogen responsive breast cancer cells, demonstrating that one ESR1 binding site, RET Estrogen Response Element 1 (RERE1), is strongly estrogen responsive. We report the identification of several motifs required for full RERE1-mediated estrogen response, including multiple ESR1 and FOXA1 motifs. We have also demonstrated by siRNA-based methods that ESR1 is absolutely required for estrogen induced response directed by RERE1. Furthermore, ectopic ESR1 expression in estrogen receptor negative MDA-MB-231 cells was sufficient to induce RERE1 directed reporter expression. Consistent with the idea of physical interactions between multiple enhancers, RET locus ESR1 binding sequence intervals have been reported to be in close proximity upon estrogen stimulation. We have observed non-additive RERE1 dependent cooperation between multiple RET locus ESR1 binding sites *in vitro*. Importantly, RERE1 also directs *RET/ret* appropriate reporter expression in the central and peripheral nervous system in zebrafish. However, although a poorly conserved mouse orthologous region, mRere1 directs closely overlapping reporter control *in vivo*, it does not exhibit estrogen responsiveness in MCF-7 cells. We will discuss our efforts to elucidate the sequence origins of estrogen-induced *RET* modulation and its potential relevance to development and disease.

302

Cerebellar neurodegeneration in SCA7 requires the expression of polyglutamine expanded ataxin-7 in both neurons and glia. *S. Furrer[1], C. Chang[1], B.L. Sopher[1,2], V.A. Damian[2], T. Ashe[3], G.A. Garden[1], A.R. La Spada[3,4,5].* 1) Neurology, Univ of Washington, Seattle, WA; 2) Laboratory Medicine, Univ of Washington, Seattle, WA; 3) Pediatrics, UCSD, La Jolla, CA; 4) Cellular & Molecular Medicine, UCSD, La Jolla, CA; 5) Institute for Genomic Medicine, UCSD, La Jolla, CA.

Spinocerebellar ataxia type 7 (SCA7) is a dominantly inherited disorder characterized by progressive degeneration of neurons in the cerebellum and brainstem, and a cone-rod dystrophy phenotype. SCA7 is caused by the expansion of a CAG/polyglutamine (polyQ) repeat in the ataxin-7 gene. We have previously shown that directed expression of polyQ-ataxin-7 to Bergmann glia (BG) in transgenic mice leads to ataxia and non cell-autonomous degeneration of cerebellar Purkinje cells (PCs). To determine the cellular basis of SCA7 neurodegeneration, we derived a conditional inactivation model for SCA7 by inserting a LoxP flanked ataxin-7 cDNA with 92 repeats into the translational start site of the murine prion protein (PrP) gene contained in a bacterial artificial chromosome (BAC) that promotes transgene expression in all CNS cell types (including PCs and BG). Two independent lines of PrP-Lox-SCA7-92Q BAC mice were found to model SCA7 on a time course similar to our previously published SCA7 murine PrP-promoter-driven standard transgenic mice. Additionally, PrP-Lox-SCA7-92Q BAC mice exhibited a measurable reduction in the number of BG processes within the cerebellar molecular layer, as demonstrated in our previous model. To conditionally inactivate polyQ-ataxin-7 expression in BG, we crossed PrP-Lox-SCA7-92Q BAC mice with Gfa2-Cre transgenic mice to direct expression of Cre-recombinase to BG in the cerebellum. Bigenic PrP-Lox-SCA7-92Q BAC; Gfa2-Cre mice exhibited excision of ~70% of mutant ataxin-7 expression in BG. Based upon a composite neurological phenotyping assessment and accelerating rotarod analysis, we found that reduced expression of polyQ-ataxin-7 in BG significantly ameliorates, but does not eliminate SCA7 neurological dysfunction. We also crossed PrP-Lox-SCA7-92Q BAC mice with Pcp2-Cre transgenic mice to direct expression of Cre-recombinase to PCs in the cerebellum. Performing a similar series of experiments, we again documented a significant amelioration of SCA7 neurodegeneration in bigenic PrP-Lox-SCA7-92Q BAC; Pcp2-Cre mice. Our results indicate that expression of mutant ataxin-7 protein in both neurons and glia are required for SCA7 cerebellar degeneration, and reinforce the view that multiple cell types contribute to the neuronal demise characteristic of neurodegenerative proteinopathies.

303

Following up a GWAS hit: *in-vitro* and *in-vivo* over-expression models of electrocardiographic QT-interval and sudden cardiac death associated gene, *NOS1AP*. *A. Kapoor[1], R.B. Sekar[2], D.R. Miller[1], J. Simmers[1], D. DiSilvestre[2], A.S. Barth[2], R. Cohn[1], G.F. Tomaselli[2], A. Chakravarti[1].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, U.S.A; 2) Department of Medicine, Division of Cardiology, Johns Hopkins University School of Medicine, Baltimore, MD, U.S.A.

Electrocardiographic (EKG) QT-interval, a measure of cardiac repolarization, is a genetically influenced quantitative trait and its excessive prolongation and shortening can increase the risk for ventricular tachycardia/ventricular fibrillation and sudden cardiac death (SCD). We have used GWAS to identify multiple common polymorphisms at the *NOS1AP* locus associated with QT-interval and effects on SCD. However these variants are non-coding with no obvious consequence on *NOS1AP* mRNA or protein. In order to assess a possible role of *NOS1AP* in regulating QT-interval duration and risk for SCD, we created cellular and animal models of *NOS1AP*-overexpression and assessed them for altered electrical phenotypes. For *in-vitro* analyses, cDNAs for long (*NOS1AP-L*) and short (*NOS1AP-S*) isoforms of human *NOS1AP* were over-expressed in neonatal rat ventricular myocytes (NRVMs) using lentiviral vectors. Optical mapping studies using voltage-sensitive fluorescent dyes were performed on monolayer cultures of NRVMs 6-7 days post transduction to assay action potential duration (APD) and conduction velocity (CV). Over-expression of *NOS1AP* shortened APD (*NOS1AP-L* 118±14 ms; *NOS1AP-S* 109±10 ms; Control 180±12 ms) and increased CV (*NOS1AP-L* 25.2 cm/s; *NOS1AP-S* 24.1 cm/s; Control 19.4 cm/s) in infected cells relative to untreated control cells. These observations demonstrate the role of *NOS1AP* in the electrical function of the heart. For *in-vivo* analyses, mouse *Nos1ap* cDNA was cloned into the pCLIP vector, which utilizes Cre-loxP for conditional expression, and the linearized vector was used for transgenic mouse production by pronuclear injections. One of the transgenic founders was crossed with *MLC2v-Cre* knock-in line in order to induce ventricular-restricted *Nos1ap* over-expression. Resting surface EKGs demonstrated spontaneous bradyarrhythmias in 2 out of 9 transgenic animals positive for both transgene and *MLC2v-Cre*, while no arrhythmias were detected in wild-type littermate controls (n=9). However, across the two groups no significant difference in various EKG parameters was observed at the resting level. Collectively, these results suggest a critical role of *NOS1AP* for cellular electrophysiology, by modifying cardiac conduction and APD *in-vitro* and by increasing the risk of arrhythmias *in-vivo*.

304

Overexpression of the Trps1 Transcription Factor in Odontoblasts Results in a Dentinogenesis Imperfecta-like Phenotype. *D. Napierala[1], Y. Sun[2], I. Maciejewska[2], E. Munivez[1], T. Bertin[1], B. Dawson[3], R. D'Souza[2], C. Qin[2], B. Lee[1,3].* 1) Baylor College of Medicine, Houston, TX; 2) Baylor College of Dentistry, Texas A&M Health Science Center, Dallas, TX; 3) Howard Hughes Medical Institute.

Dentinogenesis imperfecta (DI) is the most common hereditary defect of dentin. DI is a genetically and clinically heterogeneous disorder. DI type I is associated with osteogenesis imperfecta (OI) caused by mutations in one of the type I procollagen genes, *COL1A1* or *COL1A2*. The more severe forms of DI (type II and III) are caused by mutations in the *DSPP* gene. While mutations in the *COL1A1*, *COL1A2* and *DSPP* genes (coding for the most abundant dentin proteins) are responsible for the majority of DI, the etiology of isolated hereditary dentin disorders in many affected families is still unknown. Dentin is a calcified tissue that is one of the four major components of teeth. Dentin is produced by neural crest-derived cells, odontoblasts, that form a continuous single cell layer at the periphery of the dental pulp. Here we demonstrate that mice overexpressing Trps1transcription factor in odontoblasts (*Col1a1-Trps1* transgenic mice) present with a phenotype resembling DI. Trps1 is a GATA-type transcription factor that has been shown to play role in endochondral bone formation. Our previous studies suggested that Trps1 may be a negative regulator of mineralization. To understand the function of the Trps1 transcription factor in mineralizing tissues we generated transgenic mice overexpressing Trps1 under the control of a 2.3 kb fragment of collagen 1a1 promoter. *Col1a1-Trps1* mice develop severe post-weaning growth retardation and lethality which are secondary to malnutrition. MicroCT and histological analyses revealed tooth fragility due to diminished dentin layer in the teeth of *Col1a1-Trps1* mice. Biochemical analyses of non-collagenous dentin matrix proteins demonstrated decreased levels of both Dsp and Dpp proteins in *Col1a1-Trps1* transgenic mice. Dsp and Dpp are the most abundant non-collagenous dentin matrix proteins, and are both encoded by the *Dspp* gene. Further analyses shown that overexpression of *Trps1* in odontoblasts results in inhibition of the *Dspp* expression in these cells. Additionally, by using chromatin immunoprecipitation (ChIP) assay, we have shown that Trps1 can bind GATA consensus sites in the *Dspp* promoter. Interestingly, during tooth development *Trps1* is expressed in preodontoblasts, but not in mature odontoblasts secreting dentin matrix. These data collectively demonstrate that Trps1 is a negative regulator of dentin formation and serves this function, at least in part, through repression of the *Dspp* gene.

305

Schimke immuno-osseous dysplasia: A Mendelian complex trait arising from impaired DNA maintenance and global alterations in gene expression. *A. Baradaran-Heravi[1,2], M. Morimoto[1,2], B. Tolhuis[3], C. Shaw[4], M. Sanyal[5], A. Raams[6], A. Bokenkamp[7], K.S. Cho[8], C. Myung[1], D. Leung[1], A. Fam[1], K.H. Choi[1], Y. Huang[1], S. Lou[4], C. Huang[4], L. Elizondo[1], M. van Lohuizen[3], N. Jaspers[6], D. Lewis[5], C. Boerkoel[1,2].* 1) Child and Family Research Institute, Department of Medical Genetics, University of British Columbia, Vancouver, Canada; 2) Rare Disease Foundation, Vancouver, Canada; 3) Division of Molecular Genetics and the Centre for Biomedical Genetics, The Netherlands Cancer Institute, Amsterdam, The Netherlands; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 5) Department of Pediatrics, Stanford University School of Medicine, Stanford, CA; 6) Department of Genetics, Erasmus Medical Center, Rotterdam, the Netherlands; 7) Department of Pediatrics, Vrije University Medical Center, Amsterdam, The Netherlands; 8) Department of Biological Sciences, Konkuk University, Seoul, Republic of Korea.

Schimke immuno-osseous dysplasia (SIOD) is an autosomal recessive, highly pleiotropic, incompletely penetrant multisystem childhood disorder. It is characterized by skeletal dysplasia, renal dysfunction, and defective cellular immunity. SIOD is associated with biallelic loss-of-function mutations of *SMARCAL1* (swi/snf-related matrix-associated actin-dependent regulator of chromatin, subfamily-a-like-1) gene. This gene encodes the DNA annealing helicase, SMARCAL1. In cell culture, SMARCAL1 is recruited to stalled replication forks by direct interaction with Replication Protein A (RPA) to stabilize genome structure at those regions. To understand better the pathophysiology of SIOD, we have studied the role of SMARCAL1 homologues in humans, mice (Smarcal1) and fruit flies (Marcal1). We hypothesized that SIOD is a complex disease arising from altered DNA structure and RNA transcription. Using phenotypic, genetic and molecular analyses of *Marcal1-del/del*, *Smarcal1del/del* and *SMARCAL1del/del* organisms and cells, we found that organisms and cells deficient for their respective SMARCAL1 homologue have global alteration in genome structure. This results in increased sensitivity to genotoxic agents and environmental insults as well as global alterations in gene expression. Additionally, SMARCAL1 deficiency changes the promoter structure of the deregulated genes. Interestingly, deficiency for SMARCAL1 homologues is insufficient to cause disease in the absence of another insult. Our observations suggest that SIOD arises as a multifactorial trait when SMARCAL1 deficiency permits crossing of a gene expression or genomic integrity threshold. Such a model explains the pleiotropism and incomplete penetrance of SIOD.

306

Identification of cell specific novel splice variants and allele specific expression using RNA-seq. *B.P. Fairfax, S. Makino, K. Plant, J.C. Knight.* Wellcome Trust Centre for Human Genetics, Oxford, Oxfordshire, United Kingdom.

Chronic inflammation is a characteristic hallmark of multiple pathologies, ranging from atherosclerosis to cancer. This inflammation involves both adaptive and innate arms of the immune system. Whilst a detailed picture of innate immune cell specific expression patterns has emerged from microarray analysis, our understanding of cell specific splicing patterns is less clear. We have set out to identify, using highly purified primary cell populations, splice variation and cell specific expression in antigen presenting and other innate immune cells. We have used RNA amplification and RNA-seq to identify patterns of differential splicing confined to specific cell subtypes across healthy volunteers of European ancestry. Furthermore, we have identified transcripts whose expression shows previously undescribed cell specificity. We aim to use this data to identify allelic specific expression in a cell specific manner and to shed new light onto the functional impact of splicing in the innate immune response.

307

Population differentiated copy-number variants. *C.D. Campbell[1], N. Sampas[2], A. Tsalenko[2], P. Sudmant[1], J.M. Kidd[1], M. Malig[1], T. Vu[1], L. Vives[1], L. Bruhn[2], 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 variants (CNVs) can be polymorphic in human populations and have been shown to influence human phenotypes. We designed a custom microarray to assay 4041 polymorphic CNVs in humans. Of these variant regions, 648 are in segmental duplication-rich regions of the genome and 1269 are insertions of sequence not in the reference genome assembly. 39% of these CNVs are not represented on SNP microarray platforms commonly used to interrogate CNVs. We sought to identify CNVs that had large frequency differences across populations as these variants may be candidates for recent positive selection or may underlie ethnic predilection to disease. After applying quality control metrics, we obtained microarray data for 487 individuals from five HapMap populations: 159 European-American (CEU) (109 unrelated), 40 Han Chinese (CHB), 41 Japanese (JPT), 164 Yoruba (YRI) (109 unrelated) and 83 Maasai (MKK) (54 unrelated). By using single channel intensity data from the microarray, we were able to determine discrete copy numbers for 1155 CNVs. In addition, we accurately estimated copy number for 303 CNVs that do not form discrete copy number classes - variants often excluded in CNV analyses. Using V_{ST} and F_{ST} as metrics of differentiation, we found that CNVs in segmental duplications are more likely to be population differentiated than CNVs in unique regions of the human genome (p=0.003). We also found biallelic CNVs showed slightly greater stratification when compared to frequency-matched SNPs (p=0.065). These results suggested that CNVs (especially CNVs in duplication-rich regions) are an important source of human genetic diversity whose population characteristics must be comprehensively understood before embarking on disease association studies. As part of this analysis, we have identified several interesting CNVs with remarkably large frequency differences between populations including a number of novel sequence insertions, which would not be genotyped by existing SNP microarrays. We have generated genotype data for a subset of the most differentiated CNVs in a cohort of 1200 human diversity samples and report differences in their allele frequency among human populations. These include a CNV in the *OCLN* gene, which is a tight junction protein involved in hepatitis C viral entry, and several novel insertions that contain conserved sequence elements. Our results highlight the importance of analyzing CNVs in segmental duplications and polymorphic novel insertions.

308

Assessment of copy-number variation in a family using both whole genome sequencing and array CGH: context, transmission and mechanisms. *C. Gonzaga-Jauregui[1], F. Zhang[1,4], C.M. Carvalho[1], P. Stankiewicz[1], J.R. Lupski[1,2,3].* 1) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX, USA; 2) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 3) Texas Children's Hospital, Houston, TX, USA; 4) State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai, China.

In recent years our understanding of human genetic variation has been expanded to include structural variation, which comprises large genomic copy-number variants and inversions. Copy-number variants (CNVs) account for a more variable number of base pairs between individuals than single nucleotide polymorphisms (SNPs). In addition, CNVs that occur in gene regions can change the dosage and regulation of the genes they comprise or neighboring ones, contributing to the phenotypic variability between individuals. Therefore it is interesting to study the dynamics of this type of variation, the mechanisms that generate them and their genomic context.

Here we attempt to provide a comprehensive view of CNVs in a family of four, including both parents and two siblings. We have performed array comparative genomic hybridization (aCGH) in all the members of this family, as well as whole genome sequencing (WGS) in the two siblings in order to survey the number and location of CNVs throughout their genomes. We detected more than 200 non-redundant CNVs in this family. Most of these variants have already been identified in other studies and can be found in the structural variation databases; however, we also identified a few not previously reported CNVs. We assessed the transmission states and *de novo* rates of these CNVs by comparing the CNVs in the offspring versus those in the parents. We identified approximately 3% of the variants in the offspring to be potential *de novo* events not found in either of the parents. In addition we have refined the breakpoints of some of these CNVs in order to elucidate the possible mechanisms that have given rise to CNVs and that generate *de novo* CNVs from generation to generation. In addition to NAHR events between low-copy repeats or segmental duplications in the genome, replication based mechanisms seem to play an important role in the generation of CNVs throughout the human genome.

309

Cross-neuropsychiatric disorder comparisons of rare copy number variation identify risk genes for attention deficit and hyperactivity. *A.C. Lionel[1,2], J. Crosbie[3], N. Barbosa[3], T. Goodale[3], B. Thiruvahindrapduram[1], J. Rickaby[1], J.L. Howe[1], Z. Wang[1], J. Wei[1], L. Zwaigenbaum[4], B.A. Fernandez[5], W. Roberts[6], P.D. Arnold[3], P. Szatmari[7], C.R. Marshall[1], R. Schachar[3], S.W. Scherer[1,2].* 1) The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 3) Department of Psychiatry, Neurosciences and Mental Health, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) Department of Pediatrics, University of Alberta, Edmonton, Alberta, Canada; 5) Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St. John's, Newfoundland, Canada; 6) Autism Research Unit, The Hospital for Sick Children and Bloorview Kids Rehabilitation, University of Toronto, Toronto, Ontario, Canada; 7) Offord Centre for Child Studies, Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario, Canada.

Background: Attention deficit hyperactivity disorder (ADHD) is a common and persistent psychiatric disorder characterized by developmentally atypical and impairing inattention, hyperactivity and impulsiveness. Despite substantial evidence for a strong underlying genetic basis for ADHD, previous findings from genome wide linkage scans and hypotheses based candidate gene studies account for only a small proportion of genetic variance. ADHD traits are commonly observed in other neuropsychiatric conditions like autism spectrum disorder (ASD), schizophrenia and the DiGeorge and Williams-Beuren microdeletion syndromes. Given these observations, and the growing recognition of the role of rare copy number variants (CNVs) in these and other neuropsychiatric disorders, we sought to determine if rare-inherited and/or *de novo* CNVs contributed to the etiology of ADHD. We also investigated potential genetic overlap, in the form of rare CNVs, between ASD and ADHD. Methods: We conducted a genome-wide CNV scan of 204 unrelated ADHD patients genotyped on the Affymetrix Human SNP Array 6.0. A multi-algorithm approach incorporating Birdsuite, Affymetrix Genotyping Console and iPattern was utilized for CNV detection. High confidence CNVs detected by this approach in the ADHD cases were compared with CNVs found in 2,357 population based controls and 350 unrelated ASD patients, using the same microarray platform and CNV calling strategy. Results: We found and confirmed *de novo* CNVs in 3 of 145 (2.1%) ADHD families, for which array data was available from the proband and both parents. We also detected rare inherited CNVs in 22 of 204 (11%) probands that were absent in controls and overlapped previously implicated-ADHD loci (e.g. *DRD5*), or identified new candidate susceptibility loci for ADHD (e.g. *GABRG1*), which have been reported in rare CNV scans of other neurodevelopmental disorders. These rare CNV findings also highlighted several shared loci in our ADHD and ASD datasets (e.g. 16p11.2). Our results provide support for the role of rare CNVs in ADHD risk, and suggest that some of the affected genes may also be etiologic factors in other neuropsychiatric conditions.

310

Functional impact of global rare copy number variation in autism spectrum disorders. *D. Pinto, S.W. Scherer, Autism Genome Project Consortium (AGP).* The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children, Toronto, Ontario, Canada.

The autism spectrum disorders (ASDs) are a group of early onset conditions which affect 0.6% of the general population, characterized by impairments in reciprocal social interaction and communication, and the presence of restricted and repetitive behaviors. Individuals with an ASD vary greatly in cognitive development, which can range from above average to intellectual disability. Although ASDs are known to be highly heritable (~90%), the underlying genetic determinants are still largely unknown. The Autism Genome Project Consortium (AGP) was formed to facilitate identification of genes involved in ASDs by uniting investigators and family data. The AGP Phase I involved genome-wide analysis of linkage and copy number variation of 10K SNP data from >1100 multiplex families, revealing promising loci on 11p and 15q and a striking degree of copy number variation (CNV). Here we report on the genome-wide characteristics of rare (<1% frequency) CNV in ASD using the high density Illumina 1M array, as part of the current AGP Phase II. When comparing 996 ASD cases of European ancestry to 1,287 matched controls, cases were found to carry a higher global burden of rare, genic CNV (1.19 fold, empirical $P = 0.012$), especially so for loci previously implicated in either ASD and/or intellectual disability (OR= 1.69 fold, 95% CI= 1.17-2.4, $P = 3.4 \times 10^{-4}$). Among the CNVs there were numerous *de novo* and inherited events, sometimes in combination in a given family, implicating many novel ASD genes such as *SHANK2*, *SYNGAP1*, *DLGAP2* and the X-linked *DDX53-PTCHD1* locus. We also discovered an enrichment of CNVs disrupting functional gene sets involved in cellular proliferation, projection and motility, and GTPase/Ras signalling. Our results reveal many new genetic and functional targets in ASD that may lead to final connected pathways. Finally, CNV analysis of an independent collection of 1,000 ASD trios is ongoing. We expect that integration of the two CNV sets coupled with deeper phenotyping will help in establishing genotype-phenotype correlations.

311

Genome Wide Copy Number Variation Study Reveals Novel Loci Impacting Childhood Obesity. *S.F.A. Grant[1,2,3], J.T. Glessner[1], J.P. Bradfield[1], K. Wang[1], N. Takahashi[4], H. Zhang[1], P.M. Sleiman[1], F.D. Mentch[1], C.E. Kim[1], C. Hou[1], K.A. Thomas[1], M.L. Garris[1], S. Deliard[2], E.C. Frackelton[1], F.G. Otieno[1], J. Zhao[1], R.M. Chiavacci[1], M. Li[5], J.D. Buxbaum[4], R.I. Berkowitz[6,7], H. Hakonarson[1,2,3].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA; 4) Laboratory of Molecular Neuropsychiatry, Department of Psychiatry, Mount Sinai School of Medicine, New York, NY; 5) Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA; 6) Behavioral Health Center and Department of Child and Adolescent Psychiatry, Children's Hospital of Philadelphia, Philadelphia, PA; 7) Center for Weight and Eating Disorders, Department of Psychiatry, University of Pennsylvania, Philadelphia, PA.

The prevalence of obesity in children and adults in the United States has increased dramatically over the past decade. Genomic copy number variations (CNVs) have already been strongly implicated in extreme syndromic obesity with developmental delay. Conversely, but to complement these previous studies, we addressed CNVs in common non-syndromic childhood obesity by examining children in the upper 5th percentile of BMI but excluding any subject greater than 3 standard deviations from the mean to reduce the potential of syndromic cases being present in the cohort. We performed a whole-genome CNV survey of our cohort of 1,080 defined European American (EA) childhood obesity cases and 2,500 lean controls (<50th percentile of BMI) who were genotyped with 550,000 SNP markers. Positive findings were evaluated in an independent African American (AA) cohort of 1,479 childhood obesity cases and 1,575 lean controls. We identified 17 CNV loci which were unique to at least three EA cases that both were not previously reported in the public domain and were validated using quantitative PCR. 8 of these loci (47.1%) also replicated exclusively in AA cases (6 deletions and 2 duplications). Replicated deleted loci consisted of *EDIL3*, *S1PR5*, *FOXP2*, *TBCA*, *ABCB5* and *ZPLD1* while replicated duplications at loci consisted of *KIF2B* and *ARL15*. Although these variants may be individually rare, our results indicate that CNVs contribute to the genetic susceptibility of non-syndromic childhood obesity in subjects of both European and African ancestry.

312

A Large-Scale Analysis of Shared Loci Between Systemic Lupus Erythematosus (SLE) and 16 Other Autoimmune Diseases Identifies Novel SLE Loci. *P.S. Ramos[1], P.M. Gaffney[2], L.A. Criswell[3], M.E. Comeau[1], A.H. Williams[1], R.R. Graham[4], S.A. Chung[3], R. Zidovetzki[5], J.A. Kelly[2], K.M. Kaufman[2], C.O. Jacob[6], R.P. Kimberly[7], B.P. Tsao[8], M.E. Alarcón-Riquelme[2,9,10], T.J. Vyse[11], J.B. Harley[2], K.L. Moser[2], C.D. Langefeld[1], International Systemic Lupus Erythematosus Genetics (SLEGEN) Consortium.* 1) Dept Biostatistical Sciences, Wake Forest Univ Health Sciences, Winston Salem, NC, USA; 2) Arthritis and Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA; 3) Rosalind Russell Medical Research Center for Arthritis, University of California, San Francisco, CA, USA; 4) Biomarkers Group, Genentech, South San Francisco, CA, USA; 5) Dept of Cell Biology and Neuroscience, University of California, Riverside, CA; 6) Keck School of Medicine, Univ of Southern California, Los Angeles, CA, USA; 7) Dept Medicine, University of Alabama at Birmingham, Birmingham, AL, USA; 8) School of Medicine, Univ of California, Los Angeles, CA, USA; 9) Dept of Genetics and Pathology, Uppsala University, Uppsala, Sweden; 10) Center for Genomic and Oncological Research, Pfizer-Universidad de Granada-Junta de Andalucia, Granada, Spain; 11) Faculty of Medicine, Imperial College, London, United Kingdom.

In spite of the known clustering of multiple autoimmune disorders (ADs) in families, evidence of specific shared variants is sparse, and consequently the genetic mechanisms that may explain the patterns of disease aggregation remain unclear. In order to assess which variants are common and different between SLE and other ADs, we compared the variants reported as associated in GWAS of ADs to those identified in our large-scale analysis of SLE. We compiled a list with 381 non-MHC variants identified as significant in 49 genome-wide association studies (GWAS) of 16 ADs, available at the GWAS catalog (www.genome.gov/gwastudies). We combined the genotypic and imputed data from three published Caucasian cohorts (total of 1500 cases and 5706 controls)(Seligman et al, 2001; Remmers et al, 2007; Graham et al, 2008; Harley et al, 2008) and performed a joint- and a meta-analysis. When available, we used data from a previously described independent replication cohort of 2085 SLE cases and 2854 controls (Harley et al, 2008). All analyses were adjusted for admixture. Of the 213 SNPs that met quality control criteria, 44 survived a multiple comparisons adjustment with P<0.05 in the joint-analysis. Interestingly, the loci most shared between GWAS of ADs include IL23R, TNFAIP3, PTPN22, IL12B, IL12RA and PFKFB3-PRKCQ. Based on the number of genome-wide significant loci, we observe that SLE shares the most loci with Crohn's disease (TNFAIP3, ATG5, and 7p12.2), followed by both rheumatoid arthritis (TNFAIP3, BLK) and psoriasis (TNFAIP3, TNIP1), reflecting the number of published GWAS and associated SNPs available from the GWAS catalog. Several autoimmune variants were herein confirmed to be associated with SLE. These include SNPs in the TNFAIP3 (rs2230926, P=1.39x10-20; rs5029939, P=1.51x10-14; rs6920220, P=5.34x10-06), BLK (rs2736340, P=3.59x10-07; rs2618476, P=1.10x10-07), IL10 (rs3024505, P=6.45x10-05; rs3024493, P=4.38x10-05), and TYK2 (rs2304256, P=2.44x10-08) regions. Novel SLE loci include the VTCN1 (rs12046117, P=2.02x10-06), CD40 (rs1569723 P=1.26x10-03), IRGM (rs11747270, P=1.12x10-03) and IL12A (rs4680534, P=1.78x10-03) regions. This study expands the number of candidate loci associated with SLE and further dissects the extent of genetic overlap between SLE and other ADs.

313

Solute Carrier 2 (SLC2A9) on Chromosome 4 is associated with Uric Acid in African Americans. *B. Charles, D. Shriner, A. Doumatey, J. Zhou, A. Adeyemo, C. Rotimi.* CRGGH, NHGRI/NIH, Bethesda, MD.

Purpose: Uric acid is the primary byproduct of purine metabolism. Hyperuricemia is associated with body mass index (BMI), sex, hypertension (HTN), renal disease, and other complex diseases, including the metabolic syndrome and type 2 diabetes (T2D). Multiple genome-wide association studies (GWAS) conducted in individuals of European ancestry (EA) have reported associations between uric acid and specific genomic loci. The purpose of this study was to identify novel susceptibility loci for serum uric acid levels in African Americans and to replicate previous GWAS finding for uric acid in European ancestry populations. Methods: A total of 1,017 African Americans who participated in the Howard University Family Study conducted in Washington, DC where included in this study. Genotyping was conducted using Affymetrix® Genome-wide Human SNP Array 6.0 with genotyping calls determined by Birdseed, v2. Imputation was conducted using MACH and the HapMap reference panels for CEU and YRI. A total of 2,400,542 SNPs were assessed for association with uric acid using PLINK under the additive model with adjustment for age, sex, BMI, glomerular filtration rate, HTN, T2D status, and the 2 principal components identified in the assessment of population stratification. Results: Three variants in the gene SLC2A9 achieved genome-wide significance for association with serum uric acid (p-values ranging from 3.66x10-9 to 1.79x10-10). Conclusions: The most strongly associated locus for serum uric acid levels in individuals of European ancestry was also the most strongly associated locus in this African American sample. This finding provides significant evidence for the potential role of SCL2A9 in uric acid homeostasis across human populations.

314

Genome-wide association study identifies a susceptibility locus at 21q21 for ventricular fibrillation in acute myocardial infarction. *R. Pazoki[1,2,17], C.R. Bezzina[2,17], A. Bardai[2,17], R.F. Marsman[2,17], J.S.S.G. de Jong[3,17], M.T. Blom[2], B.P. Scicluna[2], J.W. Jukema[4,5], N.R Bindraban[3,6], P. Lichtner[7], A. Pfeufer[7,8], N. Bishopric[9,10,11], D.M. Roden[12], T. Meitinger[7,8], S.S. Chugh[13], R.J. Myerburg[9], X. Jouven[14], S. kääb[15], L.R.C. Dekker[3,16], H.L. Tan[2,3], M.W.T. Tanck[1], A.A.M. Wilde[2,3].* 1) Epidemiology, Academic Medical Center , Amsterdam, Noord Holland, Netherlands; 2) Heart Failure Research Center, Department of Experimental Cardiology, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands; 3) Heart Failure Research Center, Department of Cardiology, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands; 4) Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands; 5) Durrer Center for Cardiogenetic Research, Amsterdam, the Netherlands; 6) Department of Social Medicine, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands; 7) Institute of Human Genetics, Klinikum Rechts der Isar Technische Universität, München, Munich, Germany; 8) Institute of Human Genetics, Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt, Neuherberg, Germany; 9) Division of Cardiology, Department of Medicine, University of Miami Miller School of Medicine, Miami, FL, USA; 10) Departments of Pediatrics, University of Miami Miller School of Medicine, Miami, FL, USA; 11) Molecular and Cellular Pharmacology, University of Miami Miller School of Medicine, Miami, FL, USA; 12) Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA; 13) The Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA; 14) Université Paris Descartes, AP-HP, Hôpital Européen Georges Pompidou, Paris. France; 15) Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universität, Munich, Germany; 16) Department of Cardiology, Catharina Hospital, Eindhoven, the Netherlands; 17) The first 5 authors contributed equally to this study.

Sudden cardiac death is a leading mode of death in adults in the Western world. The largest proportion of these deaths are caused by ventricular fibrillation(VF) during acute myocardial infarction(MI). The identification of genetic variants associated with this phenotype is extremely challenging due to obvious ascertainment limitations in these patients. We report here the first genome-wide association study addressing this problem. Genome-wide association analysis in a discovery set of 972 patients with a first acute MI, 515 of whom had VF and 457 did not, identified 8 SNPs, all at chromosome 21q21 and in moderate to complete linkage disequilibrium (LD) with each other, associated with VF. Of these, the most significant association was found for rs2824292 and rs2824293 within 338bp of each other and in complete LD (OR=1.78;95%CI,1.48-2.13;P=3.3e-10). Association of rs2824292 with VF was then replicated in an independent case-control set consisting of 146 out-of-hospital cardiac arrest patients with MI and VF and 391 MI survivor controls (OR=1.49;95%CI,1.14-1.95;P=0.004). The gene closest to rs2824292 is CXADR, which encodes a viral receptor implicated in myocarditis and dilated cardiomyopathy, and which has recently been identified as a modulator of cardiac conduction. This locus has not previously been implicated in arrhythmia susceptibility.

315

GWAS for tissue plasticity: loci associated with response to resistance training in young adults. *FE. Orkunoglu-Suer[1], JM. Devaney[1], B. Harmon[1], H. Gordish-Dressman[1], P. Clarkson[2], P. Thompson[3], T. Angelopoulos[4], P. Gordon[5], N. Moyna[6], L. Pescatello[7], P. Visich[8], R. Zoeller[9], EP. Hoffman[1].* 1) Research Cntr for Genetic Med, Children's National Medical Center, Washington, DC; 2) Department of Kinesiology, University of Massachusetts, Amherst, MA; 3) Division of Cardiology, Henry Low Heart Center, Hartford Hospital, Hartford, CT; 4) Center for Lifestyle Medicine and Department of Health Professions, University of Central Florida, Orlando, FL; 5) Department of Physical Medicine and Rehabilitation, School of Medicine, University of Michigan, Ann Arbor, MI; 6) School of Health and Human Performance, Dublin City University, Dublin 9, Ireland; 7) Department of Kinesiology & Human Performance Laboratory, University of Connecticut, Storrs, CT; 8) Human Performance Laboratory, Central Michigan University, Mount Pleasant, MI; 9) Department of Exercise Science and Health Promotion, Florida Atlantic University, Davie, FL.

Skeletal Muscle is an organ system involved in some of the most prevalent and chronic human health problems, such as weakness during aging (sarcopenia, frailty), insulin resistance in metabolic syndrome and type 2 diabetes, and functional disability after injury or critical care. Muscle is also among the most adaptable tissues, remodeling in response to environmental cues, such as strength training. However, the degree of response is quite variable between individuals, suggesting genetic modifiers of the remodeling response. To study the genetics underlying variability in muscle remodeling, we enrolled young adults (av. age 24 yrs) into an unilateral upper arm supervised resistance training at seven university recruitment sites (12 wks training, non-dominant arm). In the 784 subjects completing the intervention with all phenotyping information, strength changes ranged from 0 to +250% (0 to +27.5 kg) for 1-repetition maximum, and from -42 to +189% (-47.9 to +116.0 kg) for maximum voluntary contraction. To identify genetic loci underlying differences in response to training, a genome-wide association with the change in isometric strength was tested in males in the FAMUSS cohort using a responder/non-responder subset, and Affymetrix 6.0 SNP chips. Change in strength was modeled as a continous dependent variable. The top ranked associations were then validated in the original 189 responder/non-responder subset. The probability level for statistical significance was set at $p<6.49x10-8$ to correct for the increased Type I error due to the number of SNPs that were tested. Suggestive significance was taken to be $p<1.0x10-4$. Eight loci (6 haplotype blocks, and 2 singleton SNPs) with associations between p = 10-5 - 10-7 were studied in the entire FAMUSS cohort. All eight loci were confirmed in the males but none of the loci were significant in the females. A number of the transcript units within the genomic loci have a cogent functional argument for their involvement in response to training and muscle strength. CoQ10 is critical for muscle metabolism, PAIP2B is a poly-A binding protein where other protein family members cause muscular dystrophy, TRDN (triadin) is a component of the neuromuscular junction, EHMT1 is important for epigenetics, and NRG3 (neuregulin) is involved in nerve/muscle interaction. This data has potential significance for medical rehabilitation, disease etiology, and understanding of the basic biology of muscle.

316

Meta-analysis of genome-wide association studies for bone-mineral-density (BMD) identifies 34 loci regulating BMD and a novel locus for fracture risk. *K. Estrada[1], E. Evangelou[2], Y.H. Hsu[3], U. Styrkarsdottir[4], C.T. Liu[5], A. Moayyeri[6], S. Kaptoge[6], E. Duncan[7], N. Amin[1], D. Kiel[3], D. Karasik[3], O. Albagha[8], M. Brown[7], R.L. Minster[9], M.C. Zillikens[1], C. Ohlsson[10], G. Thorleifsson[4], J. Reeve[6], L. Vandenput[10], T. Spector[16], A.G. Uitterlinden[1], L.M. Yerges-Armstrong[11], J.B. Richards[12], N. Glazer[13], A. Kung[14], D. Koller[15], K. Stefansson[4], J. Ioannidis[2], S. Ralston[8], F. Rivadeneira[1] for the AOGC and the GEFOS consortia.* 1) Erasmus University MC, Rotterdam, Netherlands; 2) University of Ioannina School of Medicine, Ioannina, Greece; 3) Hebrew SeniorLife, Harvard Medical Scool, Boston, Massachusetts, USA; 4) deCODE Genetics, Reykjavík, Iceland; 5) Boston University, Boston, Massachusetts, USA; 6) Strangeways Research Laboratory, Cambridge England; 7) University of Queensland Diamantina Institute; 8) School of Molecular and Clinical Medicine, Western General Hospital, Edinburgh, UK; 9) University of Pittsburgh Graduate School of Public Health; 10) University of Gothenburg, Gothenburg, Sweden; 11) University of Maryland; 12) McGill University, Montréal, Québec, Canada; 13) University of Washington, Seattle, Washington, USA; 14) University of Hong Kong, Hong Kong, China; 15) Indiana University School of Medicine; 16) Kings College London, London, UK.

Risk of fracture (the clinically most important outcome in osteoporosis) is frequently assessed using BMD, a heritable (h2»0.60) complex trait. Genome-wide association studies (GWAS) have identified >20 loci associated with BMD at a genome-wide significant (GWS) level, while no fracture loci have been identified associated at GWS level in Northern European populations. As part of the Genetic Factors of OSteoporosis (GEFOS) Consortium we performed inverse-variance fixed-effects meta-analysis for 2.5 million SNPs in 17 GWAS of Lumbar spine (LS)- and Femoral neck (FN)-BMD in 32,000 individuals including both men and women. The meta-analysis also included individuals from extreme truncate designs (ETD) in women of Caucasian (n=1,800) and Han Chinese (n=800) ancestry using an unbiased estimates approach. We declared GWS at $P<5x10^{-8}$ and suggestive association at $5x10^{-6} > P > 5x10^{-8}$ after overall genomic control correction (λ=1.1). We detected 34 GWS and 48 suggestive associated loci with either FN or LS-BMD (82 in total); 4 GWS and 12 suggestive loci were identified after inclusion of 2,600 ETD samples. The 34 GWS hits map within or nearby genes involved in pathways relevant to bone biology; e.g. genes in or interacting with WNT pathway (*WLS, WNT4, WNT16, AXIN1, LRP5, CTNNB1, DKK1, FOXC2, SOX4, SOX6, SOX9*). We then tested SNPs in the 82 loci for association with osteoporotic fracture risk in a subset of 7 GWAS comprising 5,900 fracture cases and 18,000 controls. Eleven out of 82 loci were nominally associated (P<0.05) with fracture risk of which 2 endured Bonferroni correction including one of the BMD GWS loci at 7q21.3 (OR=1.1 [95%CI 1.06-1.16], $P=1.7x10^{-5}$) and one of the 48 suggestive loci at 18p11.2. The latter locus was found associated with fracture risk even beyond GWS level (OR= 1.17 [95%CI 1.12-1.22], $P=5.9x10^{-11}$) and maps within the intron of an uncharacterized gene with a lipoprotein-like receptor domain. Evaluation of secondary signals and testing for non-additive effects is underway together with de-novo genotyping replication of SNPs in 82 loci. This follow-up effort comprises ~50,000 additional individuals with BMD and > 10,000 fracture cases. In summary, this large scale meta-analysis improved power by including different study designs and ethnic backgrounds. We identified 34 loci involving known and novel biological pathways underlying BMD variation and report for the first time a locus robustly associated at GWS with the risk of fracture.

317

Novel pathogenic mutations identified in Mendelian disorders by exon-level array CGH in 3,000 clinical cases. *S. Aradhya, R. Busin, T. Bonaga, N. Nwokekeh, B. Boggs, K.S. Hruska, N. Smaoui, J.G. Compton, S. Bale, S. Suchy, G. Richard.* GeneDx, Gaithersburg, MD.

Most mutations causing Mendelian disorders can be identified by DNA sequencing of the associated genes while detecting pathogenic intragenic copy number changes requires other technologies, such as Southern blot, quantitative PCR, or multiplex ligation-dependent probe amplification analysis. Although these methods are used for a limited number of disorders with a high rate of deletion/duplication mutations (dels/dups), they can be inflexible, costly, and time-consuming. For other disorders the frequency of dels/dups may be unknown or underestimated since they are not routinely investigated. We have developed an array CGH design that contains oligonucleotide probes concentrated specifically in exonic regions of 560 disease-related genes. We used this array to evaluate 3000 clinical cases, many of which were negative by DNA sequencing of one or more genes, including those associated with 133 autosomal dominant (AD), 55 autosomal recessive (AR), and 34 X-linked (XL) disorders. This testing identified 98 partial or whole-gene deletions and 2 duplications in 55 genes, corresponding to a detection rate of 3.3%. Thirty small deletions involved only one exon and might have been missed by other methods, indicating that the presence of multiple heterozygous polymorphisms should not preclude a search for a small deletion elsewhere in a gene. The previously reported high frequency of deletions was confirmed for several AD disorders, including Peutz-Jeghers, aniridia, Axenfeld-Rieger, Rubinstein-Taybi, anophthalmia, and multiple exostoses syndromes. In addition, we identified novel dels/dups in 25 other AD genes. For AR disorders, exon-level array CGH was used when only a single mutation was found by sequencing. This effort revealed 11 partial or whole-gene deletions as the 'missing' second mutation in 8% of AR cases tested. In XL genes, seven partial deletions, three whole-gene deletion, and one partial gene duplication were identified in carriers and affected individuals. Our results demonstrate that the clinical application of exon-level array CGH is a valuable tool for del/dup testing. Notably, these data affirm that some genes that are traditionally not examined by MLPA or qPCR should also be tested for copy number mutations. Exon-level array CGH is therefore suitable for multiple clinical indications, increases the sensitivity of diagnostic testing, and thus should become an integral part of diagnostic testing.

318

Acquired Uniparental Isodisomy as a Common Somatic 2nd-hit Explains Multifocality of Glomuvenous Malformations. *M. Amyere[1], V. Virginie[1], P. Brouillard[1], F. Duhoux[2], M. Wassef[3], O. Enjolras[4], J.B. Mulliken[5], H. Antoine-Poirel[2], O. Devuyst[6], L.M. Boon[1,7], M. Vikkula[1].* 1) Lab Human Molec Genetics, de Duve Inst, Brussels, Belgium; 2) Center for Human Genetics, Cliniques universitaires Saint-Luc, Université catholique de Louvain, Brussels, Belgium; 3) Service d'anatomie et de cytologie pathologiques, Hôpital Lariboisière, Paris, France; 4) Pediatric Vascular Clinic, Departement of Maxillofacial and Plastic Surgery, Hôpital d'enfants Armand Trousseau, Paris, France; 5) Vascular Anomalies Center, Division of Plastic Surgery, Children's Hospital and Harvard Medical School, Boston, MA, USA; 6) Unit of nephrology, Université catholique de Louvain, Brussels, Belgium; 7) Centre for Vascular Anomalies, Division of Plastic Surgery, Cliniques Universitaires Saint-Luc, Université catholique de Louvain, Brussels, Belgium.

Inherited vascular anomalies are commonly characterized by autosomal dominant inheritance with high penetrance, highly variable expressivity regarding size, number and localization of malformations, and small size of postnatal lesions. In 1994, when we identified the genomic locus linked to inherited cutaneomucosal venous malformations (VMCM), we hypothesized that the Knudson's double-hit model for retinoblastoma could be applicable to inherited vascular anomalies. Rarity of accessible resected inherited vascular malformations has hindered such studies. In 2002, we reported a somatic second-hit in one glomuvenous malformation (GVM). The current study, we screened 25 GVM tissues for somatic 2nd-hit mutations using DNA extracted from whole tissue or laser capture microdissected tissue, or cDNA made of total tissular RNA. We identified a somatic 2nd-hit in 13 lesions (52%). Three were intragenic changes leading to altered mRNA splicing, one was a 1p21-22 deletion and nine were acquired uniparental isodisomies (aUPIDs) of the almost complete short arm of chromosome 1, identified only by pairewise Affymetrix SNP-chip analyses. In 53 control lesions from other types of resected vascular anomalies and polycystic kidney tissues, 1p aUPID was not observed, even with pairewise analysis. Interestingly, all the identified 1p aUPIDs start from 1p13.1 region, suggesting a labile DNA area in it. Difficulty of identification, enrichment of somatic mutations by laser capture microdissection and cDNA studies, and the need for pairewise copy number analysis suggest important tissular heterogeneity for the 2nd-hits. All the identified cellular double-hits lead to localized complete glomulin loss-of-function. Thus, the inherited mutations are phenotypically recessive and need a co-existing somatic mutation, a hallmark of paradominant mode of inheritance. The data demonstrate, for the first time, that 1p aUPID is involved in non-malignant disorders and although the 1p aUPID was specific to GVMs, somatic UPID may be an important modifier of phenotypic variability in single gene genetic disorders. (miikka.vikkula@uclouvain.be) (http://deduveinstitutue.be/vikkula).

319

Mutations in the *OTX2* gene are involved in Otocephaly-Dysgnathia. *N. Chassaing[1], S. Sorrentino[2], D. Martin-Coignard[3], A. Iacovelli[2], W. Paznekas[2], P. McGuire[2,4], B. Kirmse[2,4], O. Faye-Petersen[5], F. Encha-Razavi[6], L. Lequeux[7], A. Vigouroux[1], P. Loget[8], D. Carles[9], C. Sergi[10], S. Puvabanditsin[11], C. Chen[12], G. Santiago[13], P.J. Taub[4,13], A.M. Greenberg[14], H.C. Etchevers[1,15], E.H. Rose[14], C.L. Mercer[16], P. Calvas[1], E.W. Jabs[2,4,17].* 1) Department of Medical Genetics, CHU Purpan, Inserm U563, Toulouse III University, Toulouse, France; 2) Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY, USA; 3) Department of Medical Genetics, University Hospital, Le Mans, France; 4) Department of Pediatrics, Mount Sinai School of Medicine, New York, NY, USA; 5) Department of Pathology, University of Alabama at Birmingham, Birmingham, Alabama, USA; 6) Department of Histology-Embryology, Hôpital Necker, Paris, France; 7) CHU Toulouse, Purpan Hospital, Department of Ophthalmology, Toulouse, France; 8) Department of Pathology, University Hospital of Rennes, Rennes, France; 9) Pathology Laboratory, Bordeaux University Hospital, Bordeaux, France; 10) Department of Paediatric Pathology, St. Michael's University Hospital, Bristol BS2 8EG, United Kingdom; 11) Department of Pediatrics, University of Medicine and Dentistry of New Jersey, Newark, New Jersey, USA; 12) Department of Obstetrics and Gynecology, Mackay Memorial Hospital, Taipei 104, Taiwan; 13) Department of Surgery, Mount Sinai School of Medicine, New York, NY, USA; 14) Department of Dentistry, Mount Sinai School of Medicine, New York, NY, USA; 15) INSERM, U781, Hôpital Necker-Enfants Malades, Paris Cedex 15, 75743 France; 16) Academic Unit of Genetic Medicine, Division of Human Genetics, University of Southampton, Southampton, United Kingdom; 17) Department of Developmental and Regenerative Biology, Mount Sinai School of Medicine, New York, NY, USA.

Otocephaly-Dysgnathia (OMIM #202650) is characterized by mandibular hypoplasia or agenesis, ear anomalies, microstomia and microglossia. Molecular basis of this developmental defect is unknown in humans, but three different mice knock out models for *TWSG1*, *PGAP1*, or *OTX2* genes may lead to otocephaly depending on the genetic background. In humans, *OTX2* mutations have been noted in patients with micro/anophthalmia and pituitary malformations. We have identified a deleterious *OTX2* mutation in a large family where three cousins with micro/anophthalmia each gave birth to at least one child with otocephaly. We thus screened *OTX2* in 10 unrelated otocephalic patients and found two additional mutations. Molecular analysis of *PGAP1*, *TWSG1* and *MSX1* (a gene involved in mandible embryonic formation), performed in otocephalic patients without detectable *OTX2* mutations, did not identify mutations of these genes. We also investigated a possible modifier effect of *PGAP1*, *TWSG1* and *MSX1* variants in patients with OTX2-linked otocephaly. We demonstrate that OTX2 regulates *MSX1* expression. We thus propose that abnormal regulation of *MSX1* expression in the developing mandible secondary to *OTX2* mutations cause agnathia. Our findings support genetic heterogeneity of otocephaly, and identify *OTX2* as the first gene implicated in human otocephaly.

320

Exome sequencing identifies novel gene for recessive Retinitis pigmentosa. *S. Zuchner[1], G.W. Beecham[1], A.C. Naj[1], M.A. Kohli[1], P.L. Whitehead[1], E.H. Powell[1], S.H. Blanton[1], D. Seo[1], J.D. Buxbaum[2], R. Wen[3], J.M. Vance[1], B.L. Lam[3], M.A. Pericak-Vance[1].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Department of Psychiatry, Mt. Sinai School of Medicine, NY, NY; 3) Bascom Palmer Eye Institute, University of Miami, Miami, FL.

Whole genome approaches, such as exome sequencing, have the potential to revolutionize gene testing and gene identification in Mendelian diseases. We have applied whole exome sequencing to a family with apparent autosomal recessive (AR) retinitis pigmentosa (RP). RP comprises a large number of genetically distinct forms of inherited degeneration of the retina leading to progressive loss of vision. Over 40 non-syndromic RP genes and chromosomal loci have been mapped thus far, yet they explain only ~50% of the genetic risk. The remaining genes likely represent rare forms with only few families affected worldwide. Because classic linkage approaches were not applicable, we performed whole exome sequencing in three affected and one unaffected siblings of an Ashkenazi Jewish (AJ) RP family that screened negative for all known AR RP mutations. Across the four individuals we identified 30,911 functionally significant single nucleotide variants. In all, 238 significant variants, including 27 indels, co-segregated with the phenotype, but only one change was novel (i.e. not reported in dbSNP). This novel variant was present in only 5/24,000 mixed Caucasian control chromosomes (3/5 were AJ, 1/5 non-AJ and 1 unknown as determined by Eigenstrat analysis); the allele frequency in a separate, ethnically matched AJ sample, was 5 out of 574 control chromosomes or 0.0087, resulting in a homozygote frequency of ~1/13,200 in the AJ population. Both parents carried the ancestral haplotype. The observed Lys>Glu amino acid change is predicted to be "damaging" by Polyphen and received a high 5.25 GERP conservation score. The novel RP candidate gene encodes a key enzyme in the terpenoid/ carotenoid backbone synthesis. Carotenoids are antioxidants concentrated in the macula lutia of the human eye protecting the retina from ultraviolet radiation. The in vitro effects of the identified gene mutation are currently being tested in functional assays. In summary, we demonstrate the first gene identification in a single small human recessive family suffering from a genetically highly heterogeneous phenotype. We will illustrate a viable strategy and its challenges for gene identification using whole exome targeted sequencing.

321

Genetic Basis of Van Den Ende-Gupta Syndrome. *N. Anastasio[1], A. Teebi[2], T. Ben-Omran[2], K.C.H. Ha[1,3], E. Lalonde[1,3], J. Liu[1], V. Der Kaloustian[1,4], D.S Rosenblatt[1], L. Jerome-Majewska[1,4], J. Majewski[1,3].* 1) McGill University Department of Human Genetics Room N5-13, Stewart Biology Building, 1205 Dr Penfield Avenue Montreal, Quebec H3A 1B1; 2) Weill Cornell Medical College Hamad Medical Corporation P.O.BOX. 3050 Doha-Qatar; 3) McGill University and Genome Quebec Innovation Centre 740, Dr. Penfield Avenue, Room 7104 Montreal, Quebec H3A 1A4; 4) McGill University Department of Pediatrics Montreal Children's Hospital, 2300 Tupper Street Montreal, Quebec H3H 1P3.

Van Den Ende-Gupta Syndrome (VDEGS) is an extremely rare autosomal recessive syndrome characterized by distinctive craniofacial and skeletal manifestations with normal growth and intelligence. The craniofacial anomalies include: blepharophimosis, malar and/or maxillary hypoplasia, narrow and beaked nose, and everted lower lip.

We present molecular data on 3 VDEGS patients from 3 consanguineous Qatari families belonging to the same highly inbred Bedouin tribe. These 3 patients were genotyped using SNP microarrays and a 2.4 Mb homozygous region was found on chromosome 22q11, in an area overlapping the DiGeorge critical region. This region contains 44 genes, including *SCARF2*, a gene that is expressed during development in a number of mouse tissues relevant to the symptoms previously described. Sanger sequencing identified a missense change, c.773G>A (p.C258Y), in exon 4 in the two closely related patients and a 2 base-pair deletion in exon 8, c.1328_1329delTG (p.V443DfsX82), in the unrelated individual. A fourth unrelated patient from another consanguineous marriage within the same tribe was subsequently also found to carry the c.1328_1329delTG mutation. In parallel with the candidate gene approach, complete exome sequencing was used to confirm the mutation in *SCARF2* as the genetic basis of this syndrome. SCARF2 protein contains putative EGF-like domains in its extracellular domain along with a number of positively charged residues in its intracellular domain, indicating it may be involved in intracellular signaling. However, the function of SCARF2 has not been previously characterized, and this is the first report of the phenotypic effects associated with defects in the scavenger receptor type F family of genes.

322

Disruption of the Podosome Adaptor Protein TKS4 (SH3PXD2B) causes Frank-Ter Haar Syndrome. *z. Iqbal[1,13], p. Cejudo-Martin[3,13], A. de Brouwer[1], B. van der Zwaag[4], P. Ruiz-Lozano[3], M. Cecilia Scimia[3], J. D.Lindsey[5], R. Weinreb[5], B. Albrecht[6], A. Megarbane[7], Y. Alanay[8], Z. Ben-Neriah[9], M. Amenduni[10], R. Artuso[10], J. A.Veltman[1], E. van Beusekom[1], A. Oudakker[1,2], J. Luis Millan[3], R. Hennekam[11,12], B. Hamel[1], S. A.Courtneidge[3], H. van Bokhoven[1,2].* 1) Human Genetics, Radboud University, Nijmegen centre1Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands for molecular life sciences (NCMLS), Nijmegen, Gelderland,; 2) Burnham Institute for Medical Research, 10901 N. Torrey Pines Road, La Jolla, CA 92037, USA; 3) Department of Neuroscience and Pharmacology. University Medical Center Utrecht, Utrecht, The Netherlands; 4) Hamilton Glaucoma Center, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA; 5) Institut für Humangenetik, Universitätsklinikum, Universität Duisburg-Essen, Germany; 6) Medical Genetics Unit, Saint Joseph University, Beirut, Lebanon; 7) Pediatric Genetics Unit, Department of Pediatrics, Hacettepe University Faculty of Medicine, Ankara, Turkey; 8) Center for Human Genetics, Hadassah Medical Center, Hebrew University of Jerusalem, Jerusalem, Israel; 9) Medical Genetics, Department of Molecular Biology, University of Siena, Italy; 10) Department of Cognitive Neurosciences, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands; 11) Department of Pediatrics, Academic Medical Centre, Amsterdam, the Netherlands; 12) Institute of Child Health, Great Ormond Street Hospital for Children, UCL, London, UK; 13) These authors contributed equally to the work described in this manuscript.

Frank-Ter Haar syndrome (FTHS), also known as Ter Haar syndrome, is an autosomal recessive disorder characterized by skeletal, cardiovascular and eye abnormalities, such as increased intraocular pressure, prominent eyes and hypertelorism. We have conducted homozygosity mapping on patients representing twelve FTHS families. A locus on chromosome 5q35.1 was identified for which patients from ten families shared homozygosity. For one family, a homozygous deletion mapped exactly to the smallest region of overlapping homozygosity, which contained a single gene, SH3PXD2B. This gene encodes the TKS4 protein, a PX and SH3 domain-containing adaptor protein and Src substrate. This protein was recently shown to be involved in the formation of actin rich membrane protrusions called podosomes or invadopodia, which coordinate pericellular proteolysis with cell migration. Mice lacking Tks4 also show pronounced skeletal, eye and cardiac abnormalities and phenocopied the majority of the defects associated with FTHS. These findings establish a role for TKS4 in FTHS and embryonic development. Mutation analysis revealed five different homozygous mutations in SH3PXD2B in seven FTHS families. No SH3PXD2B mutations were detected in six other FTHS families, demonstrating the genetic heterogeneity of this condition. Interestingly however, dermal fibroblasts from one of the individuals without an SH3PXD2B mutation nevertheless expressed lower levels of the TKS4 protein, suggesting a common mechanism underlying disease causation.

323

Mapping of a novel Ménière's disease locus to chromosome 1q32.1-q32.3. *C.A. Campbell[1, 2], B. Boese[3], N.C. Meyer[1], L.T. TeGrootenhuis[1], J.A. Webster[4], C. Li[5], D.A. Stephan[4], C.C. Della Santina[6], J.P. Carey[6], L.B. Minor[6], M.R. Hansen[1], B.J. Gantz[1], T.T. Harkins[7], R.J.H. Smith[1, 2].* 1) Otolaryngology, University of Iowa, Iowa City, IA; 2) Interdepartmental PhD Program in Genetics, University of Iowa, Iowa City, IA; 3) 454 Life Sciences, Branford, CT; 4) Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ; 5) Dept of Biostatistics, Harvard School of Public Health, Boston, MA; 6) Dept of Otolaryngology-Head & Neck Surgery, The Johns Hopkins University, Baltimore, MD; 7) Roche Applied Science, Indianapolis, IN, USA.

Ménière's disease (MD) is a complex idiopathic disorder of the inner ear of unknown etiology characterized by the symptom triad of vertigo, sensorineural hearing loss and tinnitus. Most cases of MD are sporadic although 5-14% of individuals will report a family history which segregates in an autosomal dominant fashion with incomplete penetrance. To date a genetic component that contributes to sporadic and/or familial MD has not been identified.

In this study we completed a genome-wide linkage scan using the Affymetrix GeneChip® Mapping 50K array on a Chilean family segregating MD over three generations. Multi-point parametric linkage analysis identified a candidate interval on 1q32 which spanned 8.3 Mb and included 117 genes. Targeted exon capture and pyrosequencing was performed on all known and hypothetical exons in the interval. Filtering of 3600 variants from three affected samples identified two variants, one each in *PCTK3* and *SLC45A3* which segregate with disease in the family. We found both genes to be expressed in the ear lending support to a role for these genes in the development of MD.

To identify the 'correct' gene, we hypothesized that a causative gene for the rare familial form of MD would also have rare and common variants involved with the development of the more common sporadic form of the disease. We therefore screened the coding sequences of both genes in the largest cohort of sporadic MD patients (n=250) and matched controls studied to date. Our data suggest that common and rare variants in both genes are associated with disease in the sporadic population.

This study is exciting as it combines linkage data with targeted capture sequence data to identify two causative genes for familial and sporadic MD demonstrating the utility of this experimental design. The genes identified in this study provide novel avenues for MD research and may clarify aspects of disease etiology in both familial and sporadic disease leading to better treatment strategies. (This research was supported in part by a grant from the American Otological Society (RJHS).).

324

SOX3 analysis in a large cohort of patients with syndromic and non-syndromic forms of pituitary deficiency: Allelic and phenotypic heterogeneity - Functional studies. *M.L. Sobrier[1], I. Giurgea[1,3], K. Machinis[1], M.P. Vie-Luton[1], S. Rose[2], G. Pinto[4], B. Mignot[5], A.M. Bertrand[5], C. Naud-Saureau[6], F. Kurtz[7], S. Odent[8], F. M'Bou[9], S. Cabrol[10], M. Houang[10], J.C Carel[11], J. Léger[11], P. Czernichow[11], M. Legendre[1,2], S. Amselem[1,2].* 1) INSERM U933, Paris, France; 2) AP-HP, Hôpital Trousseau, Service de génétique et d'embryologie médicales,Paris, France; 3) AP-HP, Hopital Henri Mondor, Service de biochimie et génétique, Créteil, France; 4) AP-HP, Hopital Necker Enfants Malades, Service d'endocrinologie, Paris, France; 5) Hopital Saint-Jacques, Service d'endocrinologie, Besançon, France; 6) Hopital de Lorient, Service de pédiatrie, Lorient, France; 7) Hopital de Saint-Avold, Service de pédiatrie, Saint-Avold, France; 8) Service de génétique, CHU de Rennes,France; 9) Service de pédiatrie et néonatologie, CH du Lamentin, Martinique, France; 10) AP-HP, Hopital trousseau, Service d'explorations fonctionnelles endocriniennes, Paris, France; 11) AP-HP, hopital Robert Debré, Service d'endocrinologie, Paris, France.

SOX3 is a transcription factor involved in the early control of pituitary development and function. So far, only 4 mutations have been identified in the SOX3 gene, which maps to Xq27.1. To assess SOX3 involvement in human pathology, we investigated 168 male patients from 166 independent families with pituitary dysfunction and with or without midline central nervous system (CNS) defects. Mutations were identified in 13 patients from 11 unrelated families. Expansions of the polyalanine tract (+8 and +11 Ala) were found in 4 boys (from 2 families); 3 of them displayed isolated growth hormone deficiency (IGHD), whereas one had combined pituitary hormone deficiency (CPHD) involving GH and TSH. A contraction of the polyalanine tract (-9 Ala) was found in a patient with CPHD and callosal anomalies. Missense mutations were observed in 8 independent patients: p.Ala102Ser in a patient with blindness related to septo-optic dysplasia and severe midline CNS anomalies; p.Arg5Pro in one with IGHD and behaviour troubles; p.Arg5Gln in a patient with CPHD and mental retardation; p.Pro3Ser in a patient with CPHD and hypertelorism and p.Pro103Thr in 4 independent patients with non-syndromic CPHD. A majority of patients had morphological pituitary anomalies: anterior pituitary hypoplasia (9/13), abnormal pituitary stalk (9/13), and ectopic posterior pituitary (10/13). The impact of each identified mutation on the subcellular location of SOX3 and its ability to transactivate a target construct were assessed through in vitro expression studies (i.e immunofluorescence and luciferase tests). The missense mutations and the contraction of the polyalanine tract (-9Ala) do not modify the expression pattern of this nuclear protein; however expansions of the polyalanine tract (+8 or +11Ala) lead to the formation of SOX3 aggregates in the nucleus or both in the nucleus and the cytoplasm. Depending on the type of mutation, transcriptional activity was found to be diminished (+11Ala), increased (p.Pro3Ser, p.Arg5Pro, p.Pro103Thr, -9Ala) in keeping with the previously described critical importance of SOX3 gene dosage in proper development, or normal (p.Arg5Gln, p.Ala102Ser, +8Ala). Two sequence variations (p.Arg5Gln and p.Ala102Ser) displayed properties identical to wild type SOX3. This study, which identifies SOX3 mutations (among which 4 are new) in 7% of this large cohort, broadens the clinical spectrum of SOX3-related disorders.

325

DUX4 over-expression recapitulates FSHD-associated phenotypes in vivo. *L. Wallace[1,3], S. Garwick[3], W. Mei[2,4], A. Belayew[5], F. Coppee[5], K. Ladner[6], D. Guttridge[6], J. Yang[2,4], S. Harper[1,2,3].* 1) Molecular, Cellular, and Developmental Biology Graduate Program, The Ohio State University, Columbus, OH; 2) Department of Pediatrics, The Ohio State University College of Medicine, Columbus, OH; 3) Center for Gene Therapy, The Research Institute at Nationwide Children's Hospital, Columbus, OH; 4) Center for Molecular and Human Genetics, The Research Institute at Nationwide Children's Hospital, Columbus, OH; 5) Laboratoire de Biologie Moleculaire, Universite de Mons-Hainaut, Academie Universitaire Wallonie-Bruxelles, Mons, Belgium; 6) Department of Molecular Virology, Immunology and Medical Genetics, The Ohio State University, Columbus, OH.

The pathogenic mechanisms underlying facioscapulohumeral muscular dystrophy (FSHD) are unclear. Predominant pathogenesis models support FSHD is caused by genetic and epigenetic abnormalities that ultimately increase expression of genes with myotoxic potential. Several putative 'FSHD genes' have been identified, but none are conclusively linked to FSHD development. Here, we provide in vivo evidence that over-expression of the FSHD candidate gene DUX4, which encodes a transcription factor, recapitulates features of muscular dystrophy in zebrafish and mice. In addition, we show that DUX4-associated toxicity requires its DNA binding ability, since over-expression of a DUX4 DNA binding domain mutant (DUX.HOX1) produced no abnormalities in vitro or in vivo. These results suggested that DUX4 transactivates downstream genes that are incompatible with normal muscle development and/or maintenance. Using real-time PCR arrays, we found that numerous genes in the p53 pathway were elevated in muscles over-expressing DUX4, but not mutant DUX4.HOX1, supporting that DUX4 transcriptional activity directly or indirectly activates apoptosis in vivo. Importantly, we show that the myopathic effects of DUX4 are p53-pathway dependent, since muscles from p53 null mice were resistant to the DUX4-induced damage we observed in wild-type mice. These results are consistent with previous observations that some p53 pathway components are activated in muscles from FSHD patients. Together, our data support that DUX4 over-expression contributes to FSHD development, and justify further investigation of p53 pathway involvement in FSHD.

326

Impaired expression of p63 and activated inflammatory response in a mouse model of premature aging. *D. Grochova, Y. Rosengardten, T. McKenna, M. Eriksson.* Department of Bioscience and Nutrition, Karolinska Institutet, Hälsovägen 7, Huddinge 141 83, Sweden.

Hutchinson-Gilford Progeria Syndrome (HGPS) is a rare genetic disorder clinically characterized by early onset of several features seen in normal aging, i.e. hair loss and wrinkling of the skin, skeletal abnormalities, dilated cardiomyopathy and atherosclerosis. Most cases of HGPS are caused by a de novo mutation in the LMNA gene (1824C>T, G608G). The mutation leads to the activation of a cryptic splice site and the production of a truncated prelamin A protein called progerin. Progerin expression has also been seen during in vitro and in vivo physiological aging suggesting a common mechanism between HGPS and normal aging. How progerin cause the phenotype seen in children with HGPS might be explained by the theory of stem cell depletion, which suggests that a premature decrease in the adult stem cell pool and subsequent impaired maintenance of tissue homeostasis account for the phenotype seen in the various organs (Halaschek-Wiener and Brooks-Wilson, J Gerontol., 2007). To test this hypothesis we used our transgenic mouse model of premature aging (Sagelius et al., J Cell Sci., 2008). These mice express a minigene of human lamin A with the most common HGPS point mutation in the epidermis and replicate several features of the HGPS skin phenotype. Preliminary results from this model suggest progressive loss of epidermal stem cells. To explore the underlying molecular mechanism of HGPS we performed global genome expression analysis using Mouse exon 1.0 ST arrays of primary keratinocytes extracted from our mouse model. Expression of the HGPS mutation resulted in an up-regulation of 35 genes and a down-regulation of 743 genes. Down-regulated genes included stem cell markers a6- and b1- integrin, Lrig1, CD34 and Krt15 as well as genes involved in two of the main signaling pathways involved in stem cell maintenance, the Wnt and Notch signaling pathways. Furthermore the level of p63 protein, which is the transcription factor important for maintaining the proliferative state of epidermal transit amplifying cells, was down-regulated on both RNA and protein level. Up-regulated genes included different inflammatory cytokines, Il1F8, Il1a, Il1b, S100a8, S100a9 and Sprr2e. In summary our results show an impaired expression of key factors responsible for epidermal stem cell maintenance and activated inflammatory response, supporting the theory of stem cell depletion in HGPS.

327

Integrating genetic and environmental risk in common disease: insights from genome-wide analysis of vitamin D receptor binding. *J.C. Knight[1], S.V. Ramagopalan[1,2,3], A. Heger[4], A.J. Berlanga[1,2], N.J. Maugeri[1], C.P. Ponting[4], G.C. Ebers[1,2].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford UK; 2) Department of Clinical Neurology, University of Oxford, Oxford UK; 3) Blizard Institute of Cell and Molecular Science, Queen Mary University of London, London UK; 4) MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford UK.

Advances in our understanding of the genetic architecture of susceptibility to common disease have implicated many new genomic loci but the resolution of specific functional variants has remained challenging. A clearer appreciation of how genetic risk relates to environmental modulators of disease may facilitate such analyses. Consideration of the ligand-activated transcription factor vitamin D receptor (VDR) provides an opportunity for such studies in the context of autoimmune disease given the growing epidemiological evidence supporting vitamin D deficiency as a significant risk factor in diseases such as multiple sclerosis (MS), type 1 diabetes (T1D) and rheumatoid arthritis (RA). Here we describe how a genome-wide map of VDR binding sites for human lymphoblastoid cells was generated by chromatin immunoprecipitation analysed by high-throughput sequencing (ChIP-seq). We define 2776 sites of VDR occupancy following calcitriol stimulation and show VDR binding sites were significantly enriched near autoimmune and cancer associated genes and disease intervals identified from genome-wide association (GWA) studies including for MS, T1D, Crohn's disease (CD), systemic lupus erythematosus (SLE), RA, chronic lymphocytic leukaemia and colorectal cancer. Enrichment was also seen for hair colour, tanning and height. Notable genes with VDR binding included IRF8, associated with MS, and PTPN2 associated with CD and T1D. Furthermore, a number of SNP associations from GWA were located directly within VDR binding intervals, for example rs13385731 associated with SLE and rs947474 associated with T1D. ChIP-seq determination of transcription factor binding, in combination with GWA data, provides a powerful approach to further understanding the molecular basis of common multifactorial diseases.

328

Identification of high-risk genes for metabolic disorders using large-scale sequencing-based GWAS. *YR. Li, J. Wang, HM. Yang, RQ. Li.* Beijing Genomics Institute at Shenzhen,Shenzhen 518083, China.

Type II diabetes is comorbid with a variety of diseases, including hypertension and cardiovascular disease. One of the primary techniques to gain an understanding of the genetic basis of these polygenic complex diseases is the use of genome-wide association studies (GWAS). Although this array-based technique has been successfully applied to identify a large number of disease-associated genotypes, an emerging issue in the medical genomics field is that there clearly remain a significant number of undiscovered genetic variations that impact these disease states, especially for metabolic disorders. The theory is that genetic variations of intermediate frequency (between 0.001-0.05) may have a large impact on the disease phenotype, and current genotyping arrays are not designed to detect such variations. We previously sequenced 200 Caucasian exomes and demonstrated that an excess of deleterious variations of intermediate-frequency do exist and that these may partially explain the missing hereditary components underlying complex diseases. We have now sequenced the exomes of 1,000 patients who had comorbidity of type II diabetes, hypertension, and cardiovascular disease and the exomes of 1000 matched long-term health-tracked super-normal controls. We discovered approximately 150,000 SNPs in the 2000 exomes, many of which had a frequency <0.05, that showed a significant difference between cases and controls. We then applied a customized genotyping array to validate these association signals in 15,000 cases and 5,000 controls, and sorted the validated signals into type II diabetes-, hypertension- and cardiovascular disease-attributable categories. Our study provides a comprehensive catalog of metabolic disorder-associated variations and also shows the power of using sequencing-based GWAS to get at these more refractory variants to improve our understanding of the biological basis of these and other polygenic complex disorders, providing a substantial advancement in the field medical genomics.

329

SOMATIC MITOCHONDRIAL DNA MUTATIONAL PATTERN IN TYPE 2 DIABETES CONCORDANT AND DISCORDANT MONOZYGOTIC TWINS. *M.Y. Buchshtav[1], S. Dadon[1], D. Glass[2], T. Andrew[2], T. Spector[2], D. Mishmar[1].* 1) Life Sciences, Ben-Gurion University of the Negev, Beer Sheva, Israel; 2) Twin Research & Genetic Epidemiology, Genetics and Molecular Medicine, Kings College, London, UK.

De novo somatic mitochondrial DNA (mtDNA) mutations accumulate during one's lifetime 10 times faster than in the nuclear genome, thus creating a mixed population of mtDNAs per individual (heteroplasmy). Heteroplasmic mutations were shown to alter mitochondrial function as well as the risk to develop complex disorders. Accordingly, we hypothesized that mtDNA somatic mutations could associate with the susceptibility to age related diseases, using Type 2 diabetes mellitus (T2DM) as a model. Specifically, we aimed at investigating patterns of de novo mutations in the mtDNA control region in a unique sample of identical twins that are concordant or discordant for the T2DM phenotype. To this end, we generated and analyzed multiple bacterial clones (>800) harboring the mtDNA control region from blood cells of each of 22 identical twin pairs, including 11 concordant and 11 discordant T2DM patients. Additionally, we are currently analyzing mtDNA control regions from muscle cells, available from 6 of the above-mentioned discordant and 3 concordant twins. Whereas our initial analysis of discordant twins did not reveal notable differences among the blood mtDNA sequences, a larger repertoire of mutations was revealed in the patients' muscle as compared to their healthy twins. Moreover, the heteroplasmic mutation repertoire was significantly different between muscle and blood tissues, with only little overlap. Interestingly, we identified 8 different multiple mutations clones, found in 7 different unrelated T2DM patients and one healthy individual. In 3 patients, the de novo somatic mutational pattern reconstituted a combination of mutations defining a known genetic background (haplotype, haplogroup) of the human mtDNA, which was different than the haplogroup the patients inherited. Recently, we showed that such a 'haplogroup switch' was a common feature of de novo cancer mutations, so it is tempting to suggest a possible functional potential for this phenomenon. The importance of this finding to the T2DM phenotype is discussed.

330

The use of whole exome sequencing and linkage analysis to identify novel candidate loci for pediatric mitochondrial disorder. *J.P. Casey[1], J. Conroy[1], R. Regan[1], N. Shah[1], E.B. Crushell[2], S.A. Lynch[3], S. Ennis[1,3].* 1) Health Sciences Centre, University College Dublin, Dublin 4, Ireland; 2) National Centre for Inherited Metabolic Disorders, Children's University Hospital, Temple Street, Dublin 1, Ireland; 3) National Centre for Medical Genetics, Our Lady's Hospital for Sick Children, Dublin 12, Ireland.

Mitochondrial diseases are a group of disorders that are caused by impairment of the mitochondria that leads to downstream effects in a number of different organs. The disorder affects 10-15 individuals per 100,000. Mitochondrial disorders have strong clinical and genetic heterogeneity and case reports of specific gene defects have been crucial to our understanding of clinical phenotypes. We report genetic analysis of 3 children presenting with mitochondrial disease of unknown aetiology. The children are first cousins from an endogamous nomadic Irish group. Parental consanguinity indicates a shared possibly recessive genetic aetiology for their condition. Patient 1 presented with severe developmental delay, sensorineural hearing loss, encephalopathy and seizures. Patient 2 presented with moderate developmental delay and dysmorphic features. Furthermore patient 2 has two phenotypically normal brothers suggesting that a nuclear-encoded gene, and not a mitochondrial gene, may underlie the disease phenotype. Patient 3 presented with mild developmental delay and seizures. All 3 patients suffer from liver and renal dysfunction. Genetic testing for Pearsons deletion and known pathogenic mitochondrial mutations proved normal. We applied a combination of SNP homozygosity mapping (HM) and next-generation-sequencing methods to investigate if a shared recessive mutation was common to this pedigree. Whole exome sequencing identifies a large number of variants making it difficult to distinguish the causative mutation from non-pathogenic variants. Our strategy is to use the homozygous blocks identified by SNP HM to narrow the search for candidate disease causing genes. The 3 individuals with mitochondrial disease and 5 unaffected relatives were genotyped on the Illumina platform using the 1million SNP array. SNP HM identified 46 homozygous segments (ranging in size from 12 to 996kb) shared by the 3 affected children but not shared by unaffected relatives. The candidate loci contain 134 genes, of which 6 have previously been implicated in mitochondrial function. Whole exome capture and sequencing was performed by GATC biotech services. Mutation analysis is currently underway. The results of this study which evaluates the effectiveness of combining whole exome sequencing and linkage analysis in the identification of rare disease causing variants in mitochondrial disorder will be presented at the ASHG 2010.

331

Exome sequencing of autism cases at the 'extreme'. *T. Turner[1], V. Pihur[1,2], S. Yegnasubramanian[3], A. Chakravarti[1].* 1) McKusick - Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Bloomberg School of Public Health, Department of Biostatistics, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Sidney Kimmel Comprehensive Cancer Center Johns Hopkins University, Baltimore, MD.

Autism has a sex ratio of 4; consequently, under the multifactorial model of inheritance, females are more extreme than males in the liability distribution. Thus, a search for deleterious 'autism' alleles in families enriched for female cases may be more efficient. Such families with high risk either have more mutations or more deleterious mutations among affecteds. We hypothesize, based on the generally higher risk associated with coding than non-coding variants, that there will be an accumulation and enrichment of rare coding variants in such females. In this study, we have sequenced the exomes of five affected females from different female-affected only families. These five exomes are a subset of the fifty-four affected females from unrelated families whom we plan to sequence. For this genomic analysis, we have used the Agilent SureSelect whole exome capture technology (captures ~38 Mb of DNA) and AB SOLiD 3 Plus sequencing platform. Variants derived from sequencing will be filtered using high quality HapMap control exomes. Remaining variants after HapMap filtering will be applied to an algorithm we are developing that both considers nucleotide conservation and amino acid conservation. All variants of interest that remain will then be characterized for the features they disrupt: missense, nonsense, and mutations affecting splicing. Initial analysis of the five autism exomes shows average coverage depth of ~53X with percent aligned targets ±200 bp (on average) being ~87%. Characterization of interesting variants in individuals examined missense and nonsense variants of which there were ~4,400 and ~35, respectively. Of all the variants in each individual, about 94% are represented in dbSNP. When analyzing each exome for variants in well established autism associated genes no damaging mutations were found. When analyzing all five exomes together there were a total of 86,423 non-reference coding and non-coding variants identified: 47,317 coding variants in the five individuals were identified. When removing duplicate variant positions across individuals the total unique coding variant loci were 21,321. Of these, 664 variants have been predicted to be damaging. Lastly, when considering only the variants which have identical position and genotype across all five exomes, 3 nonsense mutations as well as 5 predicted damaging missense mutations were discovered. These mutations were found in the following genes: *ANAPC1*, *ZNF117*, *CDC27*, *HMCN1*, *FRG1*, *PON1*, *OR4C3*.

332

Overlap and effective size of the human CD8+ T-cell receptor repertoire. *C. Carlson, H. Robins, S. Srivastava, P. Campregher, J. Andriesen, S. Riddell, E. Warren.* Pub Hlth Sci, Fred Hutchinson Cancer Res Ctr, Seattle, WA.

Diversity in T-lymphocyte antigen receptors is generated by somatic rearrangement of T-cell receptor (TCR) genes. The antigenic specificity of αβ T-lymphocytes, which primarily recognize peptide antigens presented by class I and class II molecules of the Major Histocompatibility Complex (MHC), is in large part determined by the amino acid sequence in the hypervariable complementarity-determining region 3 (CDR3) regions of the α and β chains of the T-cell receptor (TCR). The nucleotide sequences that encode the CDR3 regions are generated by somatic rearrangement of noncontiguous variable (V), diversity (D), and joining (J) region gene segments for the β chain, and V and J segments for the α chain. The existence of multiple V, D, and J gene segments in germline DNA permits substantial combinatorial diversity in receptor composition, and receptor diversity is further augmented by the deletion of nucleotides adjacent to the recombination signal sequences (RSS) of the V, D, and J segments, and template-independent insertion of nucleotides at the Vβ-Dβ, Dβ-Jβ, and Vα-Jα junctions. We sequenced the CDR3 regions from millions of rearranged TCR β chain genes in naïve and memory CD8+ T-cells of seven adults. The CDR3 sequence repertoire realized in each individual was strongly biased toward specific Vβ-Jβ pair utilization, dominated by sequences containing few inserted nucleotides, and drawn from a defined subset comprising less than 0.1% of the estimated 5 x 1011 possible sequences. Surprisingly, the overlap in the naïve CD8+ CDR3 sequence repertoires of any two of the individuals was approximately 1000-fold larger than predicted, and essentially independent of the degree of HLA matching. This overlap suggests that public TCR clonotypes may play a more important role in the adaptive immune system than has previously been expected.

333

Modeling schizophrenia pathophysiology using the Snap-25 mutant, *blind-drunk*. *K.E. Davies[1], M.V. Sobczyk[2], P.L. Oliver[1], J.W. Smith[3], K. Wulff[2], R.G. Foster[2].* 1) Physiology, Anatomy & Genetics, Univ Oxford, Oxford, Oxon, United Kingdom; 2) Nuffield Laboratory of Opthalmology, University of Oxford, John Radcliffe Hospital, Oxford, OX3 0DU, UK; 3) Dept of Psychiatry, Eli Lilly & Co. Ltd, Erl Wood Manor, Windlesham, Surrey, GU20 6PH, UK.

SNAP-25, a SNARE protein that is essential for synaptic vesicle release from presynaptic nerve terminals, has been implicated in schizophrenia from a range of genetic, pathological and functional studies. The blind-drunk (*Bdr*) mutant mouse has a point mutation in Snap-25 that affects exocytotic vesicle recycling due to altered affinity within the SNARE complex. *Bdr* mice show behavioural deficits including reduced sensorimotor gating (PPI), a schizophrenic endophenotype. As SNARE knockouts are neonatal lethal, *Bdr* is a valuable model linking neurotransmission to complex behaviour. Environmental factors are also a key element in the pathophysiology of psychiatric disease, and we investigated this by examining the influence of variable pre-natal stress on *Bdr* mice. Sensorimotor gating deficits were markedly enhanced by pre-natal stress; significantly, these effects could be reversed with the application of anti-psychotic drugs. Moreover, social interaction abnormalities were observed only in *Bdr* animals from stressed dams but not in wild-type littermates or mutants from non-stressed mothers. These results show for the first time that combining a synaptic mouse point mutant with a controlled pre-natal stressor paradigm produces both modified and previously unseen phenotypes, generating new insights into the interactions between genetics and the environment relevant to psychiatric disease. Sleep and circadian disruption is a common feature of schizophrenia; significantly, these observations pre-date the use of typical antipsychotics. Although widely reported, the genetic and mechanistic links between sleep/wake disturbance and neuropsychiatric disorders remain poorly explored. Therefore, to explore the possible mechanistic links between sleep/wake disturbance, neurotransmission and schizophrenia we have examined both rest/activity behaviour and the expression of molecular elements of the circadian system in *Bdr* mice. Allowing for nocturnal vs. diurnal activity patterns, there is a striking similarity in circadian disruption between *Bdr* mice and human subjects with schizophrenia; *Bdr* therefore provides the first direct behavioural link between disruption of sleep-wake timing and the neural physiology underlying psychiatric disorders. Our findings also suggest that sleep/wake disruption in schizophrenia can occur independently of pharmaceutical or environmental stress and results from abnormalities in the genes linked to the disorder.

334

A genome-wide map of meiotic double strand break hotspots in mouse. *K. Brick[1], I. Gregoretti[1], F. Smagulova[2], P. Khil[1], G. Petukhova[2], R.D. Camerini-Otero[1].* 1) Genetics & Biochemistry Branch, NIDDK, National Institutes of Health, Bethesda, MD; 2) Department of Biochemistry and Molecular Biology, Uniformed Services University of the Health Sciences, Bethesda, MD.

Homologous recombination is the key process that drives genetic diversity and ensures faithful segregation of homologous chromosomes during meiosis. While it is known that recombination events occur most frequently at discrete genomic loci known as recombination hotspots, our understanding of these hotspots has been limited by the lack of scalable, genome wide assays. The most detailed map to date has been inferred from HapMap data in the human genome, however it is limited to SNP-level resolution.

One of the initial steps in recombination is the formation of DNA double-stranded breaks (DSBs) by the meiotic topoisomerase II-like protein, SPO11. Following DSB formation, both 5' ends are resected, allowing the recombinase, RAD51 and its' meiosis specific paralog, DMC1 to be loaded on the single stranded DNA. In this work, we used chromatin immunoprecipitation with antibodies against DMC1 and RAD51, followed by Solexa sequencing (ChIP-Seq) to generate the first direct physical map of DSB hotspots in a mammalian genome. By a conservative estimate, we identified 9,874 mouse DSB hotspots at high resolution, including a large hotspot cluster in the pseudo-autosomal region, the site of an obligate recombination event.

Mouse DSB hotspots are over-represented in genic regions, have a distinct DNA skew pattern and are associated with certain repeats. Nucleosome occupancy, as predicted from DNA sequence is also high at hotspot centers, and ChIP-Seq against tri-methylated histone H3 lysine-4 (H3K4me3), showed that testis-specific, non promoter-associated H3K4me3 peaks are cocentered with 94% of DSB hotspots. We also identified a 19bp motif associated with the central 1kb of 45% of hotspots. This motif is highly similar to the predicted binding site for PRDM9, a H3K4-tri-methyltransferase recently implicated in hotspot formation. Intriguingly, the motif quality in hotspots positively correlates with hotspot strength. These data allude to a central role for this motif, and for H3K4me3 in DSB formation.

335

LINE-1 expression and retrotransposition is reinstated in induced pluripotent stem cells. *J.L. Garcia-Perez, P. Catalina, I. Gutierrez-Aranda, A. Macia, R. Rodriguez, P.E. Leone, P. Menendez, M. Munoz-Lopez.* Andalusian Stem Cell Bank, University of Granada, Armilla, Granada, Spain.

Long Interspersed Element-1s (LINE-1s or L1s) are a family of active non-LTR retrotransposons that predominate in the human genome. Active LINE-1 elements encode proteins required for their mobilization (a process known as retrotransposition). L1-encoded proteins can also act in trans to mobilize Short Interspersed Elements (SINEs), such as Alu elements. L1 and Alu insertions have been implicated in many cases of human genetic disease, and their retrotransposition provides an ongoing source of human genetic diversity. As selfish DNA elements, L1 and Alu are expected to ensure their transmission to subsequent generations by retrotransposing either in germ cells or during early embryonic development. Human Embryonic Stem Cells (hESCs) are characterized for active L1 expression and for supporting a low level of L1 retrotransposition using engineered L1 elements. On the other hand, recent work from several laboratories has allowed the reprogramming of terminally differentiated human cells into Induced Pluripotent Stem Cells (iPSCs) by the forced expression of a cocktail of reprogramming factors. Indeed, iPSCs are indistinguishable from hESCs in a number of characteristics including gene expression profile, pluripotency potential, etc. Here, we analyzed L1 expression and retrotransposition in two iPSC lines. First, we generated a new iPSC line by ectopic expression of the transcription factors Oct4, Klf4, Sox2 and C-myc using human fibroblasts. Next, we analyzed L1 expression (mRNA and encoded ORF1p) and L1 mobilization using engineered human L1s in iPSCs. Our results indicate that, upon reprogrammation, the promoter of L1 is hypomethylated with respect to the parental fibroblasts. Accordingly, the level of L1 mRNA in iPSCs is around 10-20 fold higher than in the parental fibroblasts, reaching similar levels as detected in hESCs. In addition, using engineered human L1s, we detected a significant increase in L1 retrotransposition with respect to the parental fibroblasts. Thus, we propose that L1 expression and likely retrotransposition is a characteristic of pluripotent stem cells, and suggest that both hESCs and iPSCs are an excellent model system to study L1 biology in humans.

336

Dosage of the X chromosome: sex-specific differences in gene expression in brain, immune cells and reproductive organs. *D. Nguyen, C. Disteche.* Department of Pathology, University of Washington, Seattle, WA.

In mammals, X-linked gene expression is adjusted by two regulatory mechanisms - upregulation of the active X in both sexes and X inactivation in females - to compensate for the evolutionary loss of Y-linked genes. 15-25% of human X-linked genes escape dosage compensation, potentially leading to expression differences between males and females. A comprehensive analysis of more than 4500 expression arrays representing 65 human tissues showed a male bias for pseudoautosomal gene expression and a female bias for other genes that escape X inactivation. Significant sex-specific bias were particularly pronounced in certain regions of the brain, specific subtypes of immune cells, muscle, skin, and reproductive organs. Interestingly, many escape genes implicated in mental retardation showed higher expression in women's brain compared to men's. Furthermore, some genes previously classified as genes subject to X inactivation showed strong bias in some tissues such as T-cells, suggesting imcomplete inactivation in this cell type. Our data support multiple functional roles of escape genes in Turner patients who have unique cognitive deficiencies, autoimmune diseases and ovarian dysfunctions. We propose that escape from X inactivation is positively selected to contribute unique female biological functions.

337

A search for novel genes involved in cholesterol metabolism by combining exome sequencing with exclusion linkage analysis data in nuclear families. *M.M. Motazacker[1,2], G.K. Hovingh[1], J.J.P. Kastelein[1], E.S.G. Stroes[1], J.A. Kuivenhoven[1,2].* 1) Department of Vascular Medicine, Academic Medical Center, Meibergdreef 9, 1105AZ, Amsterdam, The Netherlands; 2) Department of Experimental Vascular Medicine, Academic Medical Center, Meibergdreef 9, 1105AZ, Amsterdam, The Netherlands.

Exome sequencing has recently become a feasible method to identify genetic variations in all coding exons in the human genome. However, identifying disease causing variations among hundreds of mutations in every human genome is a challenge. For complex disorders, this is even a bigger challenge since pooling/comparing different patients is not as feasible as in Mendelian disorders due to heterogeneity. Selecting from the extremes of normal distribution curve for plasma HDL cholesterol concentrations, we have performed exome sequencing in index cases of 3 families with recurrent very low or very high plasma HDL-c levels (<5th and >95th percentile for age and gender, respectively) with autosomal dominant pattern of inheritance after mutations in well-characterized candidate genes were excluded. To narrow down the window of identified variations by exome sequencing, we have performed classical linkage analysis using whole genome SNP data from index cases as well as their first degree relatives. We then excluded those variations that reside in haplotype blocks with LODscores below -2, i.e. regions that are excluded from being linked to the phenotype in the family. With this approach, we were able to exclude more than 98% of the variations identified by exome sequencing in the 3 families under study. Co-segregation and functional characterization of the remaining suggestive variations is currently under investigation. We believe that combination of exome sequencing and exclusion linkage analysis will greatly facilitate identification of disease causing variations in core families with complex disorders.

338

Genetic modifiers of cardiac outflow tract anomalies in humans with velo-cardio-facial/DiGeorge/22q11.2 deletion syndrome. *T. GUO[1], D. McGinn[2], A. Blonska[1], A. Bassett[3], E. Chow[3], M. Bowser[2], M. Sheridan[2], F. Beemer[4], K. Devriendt[5], M.C. Digilio[6], B. Marino[7], B. Dallapiccola[6], C. Carpenter[1], A.M. Higgins[8], N. Philip[9], T. Simon[10], K. Coleman[11], W. Kates[12], M. Devoto[2], E. Zackai[2], Y. Shen[13], Z. Liu[13], J. Ott[13], R. Shprintzen[9], B. Emanuel[2], B. Morrow[1], the International Chromosome 22q11.2 Consortium.* 1) Department of Genetics, Albert Einstein College of Medicine, Bronx, NY; 2) Division of Human Genetics CHOP and Department of Pediatrics, U Penn, SOM; 3) Clinical Genetics Research Program, Centre for Addiction and Mental Health and Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; 4) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 5) Center for Human Genetics, University of Leuven, Leuven, Belgium; 6) Medical Genetics, Bambino Gesù Hospital, Rome, Italy; 7) Department of Pediatrics, La Sapienza University of Rome, Rome, Italy; 8) Velo-Cardio-Facial Syndrome International Center, Department of Otolaryngology and Communication ScienÙce, SUNY Upstate Medical University, Syracuse, NY, USA; 9) Department of Medical Genetics, AP-HM and University of Mediterranee, Timone Children's Hospital, Marseille, France; 10) M.I.N.D. Institute & Department of Psychiatry and Behavioral Sciences, University of California, Davis, California, USA; 11) Children's Healthcare of Atlanta, Atlanta, Georgia, USA; 12) Department of Psychiatry and Behavioral Sciences, and Program in Neuroscience, SUNY Upstate Medical University, Syracuse, NY, USA; 13) Statistical Genetics Group Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China.

We have used the GWAS strategy to determine the molecular basis of variable cardiovascular expressivity in a genomic disorder. Approximately 70% of 22q11.2 deletion syndrome (velo-cardio-facial/DiGeorge) patients have varied conotruncal heart defects, while 30% have normal anatomy. A genome-wide association study was performed to find genetic modifiers of cardiovascular anomalies, on 677, 22q11DS subjects using Affymetrix 6.0 microarrays. Some loci were associated with multiple anomalies, while others were specific for single defects, helping build genotype-phenotype correlations. For example, one locus on 7q31.32, encompasses *PTPRZ1* (Receptor-type tyrosine-protein phosphatase zeta) and *AASS* (Aminoadipate-semialdehyde synthase) were associated with combined cardiovascular defects ($p=1.9x10^{-5}$, OR=0.61). A cluster of SNPs on 20q13, located in an intron of *EYA2* (Eyes absent homolog 2) associated with RAA ($p=1.1\ x10^{-6}$, OR= 0.39). Another on 14q31.1, in an intron of *C14orf145* showed association with TOF ($p=1.4x10^{-7}$, OR=2.48) and VSD ($p=1.4x10^{-6}$, OR=2.11) and TOF versus all the other samples ($p=7.5x10^{-6}$,OR=1.91). However, two rare SNPs from 16q23.2-*PKD1L2* (Polycystic kidney disease protein 1-like 2) and 15q22.31-*CSNK1G1* (Casein kinase I isoform gamma-1) showed much stronger association with TOF versus all the other samples ($p=1.0\text{-}2.4x10^{-8}$). Of interest, there were two loci (4q28.3- *PCDH10* (Protocadherin-10), 9q22.32-*PTCH1* (Protein patched homolog 1)) were associated to isolated VSDs ($p=1.3x10^{-8}$, OR=2.81; $p=1.5x10^{-7}$, OR=0.39, respectively), but not with VSDs associated with TOF or PTA, suggesting existence of different risk factors. PCDH10 is a calcium-dependent cell-cell adhesion molecule and was reported to be involved in the paraxial mesoderm development, and PTCH1 is a negative regulator of hedgehog signaling. Replication and fine mapping of top signals is underway. Haploinsufficiency of *TBX1* on 22q11.2 is responsible for conotruncal defects in the syndrome. When taken together the unbiased approaches suggest novel genes, many of them as of yet, not directly related to TBX1. The *TBX1* gene was re-sequenced on the remaining allele of 22q11.2 and found slight association to ASDs with/out other cardiac anomalies. Rare variants, occurring in 1-2 subjects each may alter function of the protein. Several known candidate genes were ruled out by this approach. Some were not covered on the arrays, requiring candidate gene approaches to find common/rare variants.

339

Evidence that sarcomere mutations account for a significant proportion of 'idiopathic' cardiomyopathy in a large pediatric cohort. *S.C. Bowdin[1,2], C. Manlhiot[2], B.W. McCrindle[2], Y. Etoom[2], E.A. Stephenson[2], L.N. Benson[2], P.F. Kantor[2].* 1) Clinical & Metabolic Gen, Hosp Sick Children, Toronto, ON, Canada; 2) Division of Cardiology, Hospital for Sick Children, Toronto, ON, Canada.

Pediatric cardiomyopathy is recognised to have a genetic cause in a subgroup of patients. Clinical testing is available for mutations in over 20 genes on a commercial basis. To date, cardiologists and geneticists at our institution have ordered genetic testing for cardiac sarcomere mutations in over 100 cardiomyopathy patients. The specific aims of our study were:- 1.To determine the proportion of patients in whom the etiology is primarily genetic 2.To correlate genetic results with the outcome measures of remodeling or death/transplant. We performed a retrospective cohort study of patients diagnosed with hypertrophic, dilated, restrictive cardiomyopathy (HCM, DCM,RCM respectively) and left ventricular noncompaction (LVNC)in our institution over a 9 year period, excluding patients with acquired disease. Our final cohort numbered 196 patients, of which 23% had a mutation in a cardiac sarcomere gene (SAR), 21% had metabolic disease (MET), 12% had neuromuscular disease (NM), 14% had a known diagnosis not classified above and 40% had no known diagnosis ('idiopathic',IDI). The sarcomere genes accounting for the greatest number of cases were MYBPC3, MYH7 and TNNI3. In the cohort of patients with HCM, 35% had a sarcomere mutation. In those with DCM, this figure was 7%, and in LVNC 24% had a sarcomere mutation. Freedom from death or transplantation up to age 16 in the SAR group was 80%, in the MET group 35% and in the NM group was 85%. From our data, we conclude that commercially available testing for sarcomere mutations in children with idiopathic cardiomyopathy allows a diagnosis to be made in a significant proportion of those with HCM and LVNC, and to a lesser degree in those with DCM. Our data shows that the outcome is significantly better in the SAR group compared to the MET group, suggesting that genetic testing could provide clinicians with some evidence upon which to base prognosis in this group. The limitation of this study is that it is a retrospective, observational study. There was no systematic genetic testing of children in the idiopathic group, therefore the percentage of cases that we report as being explained by sarcomere mutations is likely to be an underestimate.

340

A genome-wide study of copy number variation identifies a novel gene for dilated cardiomyopathy. *N. Norton, J.D. Siegfried, A. Morales, D. Li, R.E. Hershberger.* Cardiovascular Division, University of Miami Miller School of Medicine, Miami, FL.

Dilated cardiomyopathy (DCM), a common primary myocardial disease, causes systolic dysfunction and heart failure, a major public health problem. Mode of transmission is mostly autosomal dominant but the disease is extremely heterogeneous with point mutations identified in over 30 genes. However, only ~30% of DCM genetic cause can be explained by point mutations in the known DCM genes, leaving most genetic cause unknown. To further identify genetic causation, we performed the first genome-wide screen of copy number variation (CNV) in DCM in the genomes of 12 unrelated DCM patients with NimbleGen 2.1M comparative genomic hybridization arrays. We then validated putative novel variants (not present in the database of genomic variants) with custom NimbleGen 135K arrays in the patient and both parental DNAs. We validated 66 putative novel CNVs with the custom arrays (39 deletions and 27 duplications) ranging 2,132 to 67,058bp, (average 9,806bp, median 6,030bp) demonstrating the ability of the NimbleGen 2.1M genome-wide arrays to detect small CNVs. None of the validated CNVs occurred in any of the published DCM genes. 41 validated CNVs spanned exons or encompassed whole genes, 6 of which were inherited from an affected parent and 2 of which were de novo and occurred in sporadic cases. An 8.7kb deletion spanning exon 4 of the *BAG3* gene was present in the proband, who was also previously shown to be point mutation negative for 15 of the most common known DCM genes. 6/6 other affected family members, including the proband's father, carried the deletion, which was not present in 3/4 unaffected family members. The youngest member, who carried the deletion, was 25 years old. The average age of disease onset in this family is 43.5 years and the age of the unaffected, mutation-negative individuals ranged from 49-60 years, making the observed pattern of segregation consistent with disease causation. The deletion was not observed in a set of 90 controls. Point mutations in *BAG3* are associated with myofibrillar myopathies including severe autosomal dominant childhood muscular dystrophy. Cardiomyopathy is a frequently associated feature making *BAG3* an excellent novel candidate for DCM. Confirmatory studies are in process. We conclude that copy number variants must be searched for alongside point mutations as a cause for DCM and that genome-wide studies will identify novel genes in this disease.

341

Recurrent Duplications of 16p13.1 Confer Susceptibility to Thoracic Aortic Aneurysms and Dissections. *S.K. Prakash[1], S.Q. Kuang[2], D.C. Guo[2], R.J. Johnson[2], E.S. Regalado[2], L. Russell[1], J. Cao[2], C.S. Kwartler[2], K. Fraivillig[2], J.S. Coselli[1], H.J. Safi[2], A.L. Estrera[2], S.M. Leal[1], S.A. LeMaire[1], J.W. Belmont[1], D.M. Milewicz[2], the GenTAC Consortium.* 1) Baylor College of Medicine, Houston, TX; 2) University of Texas Health Science Center at Houston, Houston, TX.

Deletions of chromosome 16p13.1 are associated with autism, while reciprocal duplications of the same region are associated with a three-fold increased risk of schizophrenia. In this study, we report 16p13.1 duplications in adult patients with Thoracic Aortic Aneurysms and Dissections (TAAD), a major cause of death in the United States. We conducted a genome-wide SNP array analysis of 800 sporadic TAAD (STAAD) patients who underwent treatment at the University of Texas Health Science Center and Baylor College of Medicine (100% Caucasian, 35% female, average age 63, 36% Stanford Type A dissections). Fourteen recurrent rare copy number variant (CNV) regions were significantly enriched in STAAD patients in comparison with ethnically-matched controls from the Database of Genotypes and Phenotypes. The most abundant of these were large recurrent duplications of 16p13.1, which were enriched more than ten-fold in STAAD cases (9 of 800, 1.03%) compared with ethnicity matched controls (4 of 4261, 0.09%, $P=2.3x10^{-5}$, OR=12.0, CI=3.9-37.1). We replicated these findings in an independent cohort of 466 Caucasian patients with thoracic aortic disease (P= 0.017, OR 11.6). Patients with 16p13.1 duplications were more likely to harbor a second rare CNV (Fisher's exact P=0.017) and present with aortic dissections (P=0.00058). We found no evidence of other cardiovascular or neuropsychiatric disorders in individuals with duplications. *MYH11*, which encodes the smooth muscle specific isoform of β-myosin heavy chain and is mutated in 1-2% of familial TAAD cases, is one of nine genes within the 0.84 Mb common duplicated region of 16p13.1. Levels of *MYH11* transcripts were significantly increased in aortic tissues from patients with 16p13.1 duplications compared with aortic tissue from TAAD patients without the duplication and control aortas. However, the duplications failed to segregate with TAAD in familial cases, and unaffected family members harbored 16p13.1 duplications. We propose that 16p13.1 duplications confer a substantially increased risk for TAAD by perturbing smooth muscle contractility, but are not sufficient to cause aortic disease without additional genetic variants or environmental risk factors. Additionally, 16p13.1 duplications confer a risk for more than one disease, specifically schizophrenia and aortic dissections, a finding that may be common to other recurrent CNVs involving multiple genes.

342

Aortic dilatation rate in Marfan syndrome is associated with the expression of HLA-DRB genes. *T. Radonic[1,4], P. de Witte[2], M.C. van Eijk[3], M.J.H. Baars[4], B.J.M Mulder[2], M. Groenink[2], A.H. Zwinderman[1].* 1) Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center Amsterdam, Netherlands; 2) Department of Cardiology, Academic Medical Center Amsterdam, Netherlands; 3) Department of Biochemistry, Academic Medical Center Amsterdam, Netherlands; 4) Department of Clinical Genetics, Academic Medical Center Amsterdam, Netherlands.

Objectives: Because losartan is hypothesized to be beneficial in reducing aortic root growth of Marfan syndrome (MFS) patients we investigated the changes of whole transcriptome gene expression (WTGE) in skin biopsies before and after 4 weeks of treatment with losartan. We correlated the baseline WTGE with aorta dilatation rate measured over 12 years prior to treatment start. Methods: Punch skin biopsies were obtained in consenting participants of the COMPARE trial before therapy (baseline), after 4 weeks and one year of losartan therapy. In 88 samples RNA was isolated and WTGE was measured using Human Exon 1.0 ST Arrays (Affymetrix). WTGE changes and splicing events caused by losartan were investigated in patients before and 4 weeks after losartan therapy. Results were validated using rtPCR. Baseline WTGE measurements were correlated with aortic dilatation rate and aortic distensibility change over past 12 years measured by means of MRI. Results were validated in another group of 13 patients. Results: Paired analysis of gene expression changes after 4 weeks of losartan therapy revealed 20 differentially expressed genes (delta=0.54, $q<10^{-7}$). Two genes of the TGF-β pathway were significantly changed after losartan therapy: CIDEA and ENG. Losartan therapy changed splicing of ACSM3 and ADCY6 (delta=0.3, $q<10^{-7}$). When correlated with the aortic dilatation rate, baseline expression of 2 genes was significant: HLA-DRB5 and HLA-DRB1 (r=0.46, r=0.42; $q<10^{-7}$). Validation revealed p values of 0.0002 and 0.014 respectively again, suggesting an inflammation process in the connective tissue modifying the aortic dilatation rate. In order to understand the type of the inflammation and its role, we measured levels of 48 cytokines and TGF-beta in 170 blood samples of MFS patients. Correlation of WTGE with the aortic distensibility change revealed 1 significant gene Rai14 (delta= 0.5, $q<10^{-7}$). Conclusion: The modifying role of inflammation on the aortic dilatation rate is a novel finding in the MFS. It may be used as a biomarker of progressive aortic dilatation and offers new perspectives for the treatment of a progressive aortic dilatation. Clinical dosage of losartan seems to decrease the TGF-β signaling pathway by modulating the expression in CIDEA and ENG genes in Marfan patients. The exact role of TGF β pathway in inflammation and the modifying role of losartan in these patients are yet to be defined.

343

Clinical phenotyping of a new cohort of 105 probands with mutations in the TGFbeta receptor genes. *B. Loeys, C. Van Hemelrijk, E. D'Haese, M. Renard, L. Van Laer, J. De Backer, P. Coucke, A. De Paepe.* Dept Med Gen, Ghent Univ Hosp, Gent, Belgium.

TGFBR1/2 mutations are associated with Loeys-Dietz syndrome (LDS), a pleiotropic connective tissue disorder characterized by particularly aggressive vascular disease including aortic root dissections at small dimensions and at young age and occurrence of aneurysms throughout the arterial tree. Debate exists on whether conditions that are usually associated with less aggressive thoracic aortic disease such as subsets of individuals with Marfan syndrome (MFS) or familial thoracic aortic aneurysm (FTAA) can be caused by TGFBR1/2 mutations. We studied a hitherto unreported cohort of 105 probands with TGFBR1/2 mutations, respectively 39 patients with a TGFBR1 mutation and 66 patients with a TGFBR2 mutation, including the first intragenic TGFBR2 deletion that was detected only after MLPA. Over 90% of all mutations represent missense mutations in the intracellular serine-threonine kinase domains of both receptors. Male to female distribution was about equal (54% versus 46%). The overall mean age at diagnosis was 19.1 years. About 35% of the probands have a family history of LDS whereas the remaining were sporadic cases. Aortic aneurysms or dissections were the most consistent finding in this patient group with only 9 patients presenting normal aortic dimensions at diagnosis. Two thirds of the patients had dilation of the aortic root at sinuses of Valsalva, one fifth of the ascending aorta and in one eighth the descending aorta was involved. The mean age at diagnosis of aortic dilatation was 17,2 years (median 11,1 years). At the time of diagnosis, about 20% of the patients already suffered an aortic dissection, with a mean age of 22 years. Prevalence of aortic dissections was slightly higher in the TGFBR1 mutant patients than in the TGFBR2 mutant patients (26% versus 17%, p=0.3). Comparison of the clinical features of patients in our cohort with those of the patients with the identical mutations from the literature that were reported as MFS or FTAA, showed that virtually all patients had clinical features discriminating them from either MFS or FTAA, including hypertelorism, cleft palate/bifid uvula, arterial tortuosity, aneurysms beyond the aortic root, craniosynostosis, club foot, easy bruising, atrophic scarring, translucent skin and visceral rupture. Taken together, our data suggest that TGFBR1/2 mutations underlie a clinically distinct phenotype requiring specialized management and counseling.

344

Calcium Channel Blockers Exacerbate Aortic Disease and Cause Premature Lethality in Marfan Syndrome. *J. Doyle[1], J. Habashi[1], M. Lindsay[1], D. Bedja[1], H. Dietz[1,2].* 1) Inst Gen Med, Johns Hopkins, Baltimore, MD; 2) Howard Hughes Medical Institute (HHMI), Johns Hopkins, Baltimore, MD.

Aortic root dilatation and dissection is the leading cause of death in Marfan syndrome (MFS). We previously showed that excessive TGFβ-dependent activation of ERK signaling drives aneurysm progression in MFS. TGFβ neutralizing antibody, selective inhibitors of ERK activation, or the AT1R blocker losartan abrogate aneurysm growth in MFS mice by antagonizing TGFβ-dependent ERK activation. Pending the results of clinical trials of losartan, calcium channel blockers (CCBs, e.g. amlodipine) are currently used as second-line therapy for MFS in patients intolerant of beta-blockers, largely on theoretic grounds due to their blood pressure-lowering effect. To interrogate this practice, wild-type (WT) and MFS (C1039G/+) mice were treated with amlodipine or placebo from 2 months of age. Baseline echocardiography was performed prior to the onset of therapy at 2 months, and every 2 months thereafter. Placebo-treated C1039G/+ mice had larger aortic roots at 4 months ($p<0.001$) and more rapid growth of the aortic root from 2 to 4 months (p=0.04), compared to WT littermates. Amlodipine-treated C1039G/+ mice showed no improvement in either parameter when compared to placebo-treated counterparts (p=1.0 and p=0.95, respectively). There was no difference in the more distal ascending aortic (AscAo) diameter at 4 months (p=0.51) or growth from 2 to 4 months (p=0.87) between C1039G/+ and WT mice. In contrast, both AscAo size and growth were increased in amlodipine-treated WT (p=0.002 and p=0.025, respectively) and C1039G/+ ($p<0.0001$ and $p<0.0005$, respectively) mice, compared to genotype-matched placebo controls, with a greatly amplified growth effect in the Marfan model (p=0.007). Amlodipine-treated C1039G/+ mice uniquely showed early mortality, with 30% dead due to aortic dissection after 3 months of treatment. Accelerated AscAo growth and dissection correlated with increased ERK activation in the aortic wall. In a retrospective analysis, 2 of 6 MFS patients treated with CCBs showed aortic dissection in childhood, one of whom dissected at an aortic dimension that does not typically confer risk. Use of CCBs requires more study, but should be done with caution in MFS. In an era of expanding tension between concepts in evidence-based and personalized medicine (where the generation of evidence for small groups is inherently constrained), this study demonstrates how animal models can be used to fill the void.

345

_NT5E_ Mutations are Associated with Arterial Calcifications. *S.G. Ziegler[1], C. St. Hilaire[2], T. Markello[1,3], C. Groden[3], F. Gill[4], H. Carlson-Donohoe[1], A. Brusco[5], R. Nussbaum[6], R. Kleta[7], M. Boehm[2], W.A. Gahl[1,3].* 1) National Human Genome Research Institute, NIH, Bethesda, MD; 2) National Heart, Lung, and Blood Institute, NIH, Bethesda, MD; 3) Undiagnosed Diseases Program, NIH, Bethesda, MD; 4) NIH Clinical Center, Bethesda, MD; 5) A.O.U. San Giovanni Battista, S.C.D.U. Medical Genetics, Department of Genetics, Biology and Biochemistry, University of Turin, Italy; 6) Institute of Human Genetics, University of California, San Francisco, CA; 7) Departments of Medicine and Physiology, University College London, London, United Kingdom.

Medial arterial calcification of the lower extremities with periarticular calcification was first described in 1914. The familial nature of this condition was reported in 1954, leading to an OMIM citation (#211800). A total of seven cases have been published to date. Here we identified nine individuals in three families with the same extensive calcifications. Our proband in Family 1 was referred to NIH's Undiagnosed Diseases Program for peripheral vascular disease secondary to presumed tumoral calcinosis. Bone surveys revealed marked calcification of the lower extremity arteries and hand and foot joint capsules. All five siblings in one family, three siblings in another, and one patient in a third family had similar symptomatic calcifications. A cross-section of a femoral-popliteal bypass in one of the patients showed broken elastic fibers and localized medial calcification. A single nucleotide polymorphism array revealed a shared 22.4 MB region of homozygosity on chromosome 6 in all affected members of Family 1. Direct sequencing identified a homozygous nonsense mutation (c662C>A; pS221X) in *NT5E*, encoding CD73. This ecto-5-prime-nucleotidase converts AMP to adenosine extracellularly. Affected members of Family 2 had a homozygous missense mutation (c1073G>A; pC358Y) in *NT5E*; the proband of Family 3 was compound heterozygous for c662C>A and c1609dupA (pV537fsX7). All mutations result in nonfunctional CD73. Cultured fibroblasts of affected members of Family 1 exhibited markedly reduced expression of *NT5E* mRNA, CD73 protein, and enzyme activity. When cultured in calcific media, affected fibroblasts had decreased extracellular adenosine and increased alkaline phosphatase levels and accumulated calcium phosphate crystals. Transduction with a CD73 lentivirus normalized CD73 and alkaline phosphatase activity in patient cells. Adenosine treatment reduced alkaline phosphatase levels and calcification. The noncompetitive alkaline phosphatase inhibitor levamisole completely prevented calcification. Loss of CD73 function due to biallelic *NT5E* mutations most likely describes the previously reported cases and explains the pathogenesis underlying the vascular calcification seen in our three families. These findings reveal the critical role of adenosine in inhibiting ectopic calcification within specific vessels and joints. Further research into this pathway can give new insight into pathologic calcification processes and potential therapies.

346

Aortopathy Is a Common Manifestation of Fabry Disease. *R.E. Pyeritz[1], G. Griffin[1], B.A. Bernhardt[1], K. Weisiger[2], S. Segal[2], S. Packman[2], J.E. Bavaria[1], E. Rame[1,2].* 1) Department of Medicine, Univ Pennsylvania Sch Med, Philadelphia, PA; 2) Department of Pediatrics, Univ California San Francisco, San Francisco, CA.

Fabry disease affects many organs due to age-dependent accumulation of glycosphingholipids in lysosomes. Occlusive coronary artery disease, stroke, thickening of the ventricular walls and valves, and conduction defects are well known cardiovascular manifestations. We cared for 2 male patients with ascending aortic aneurysms requiring surgical repair, which prompted us to question whether these aneurysms represented unfortunate coincidences of two conditions or aortopathies due to Fabry disease. We reviewed the 18 patients followed at Penn for evidence of thoracic aortic involvement. All but 3 of the patients have received enzyme replacement therapy. All of the males (13) had an aortic root dimension at the sinuses of Valsalva at or above the upper limits of normal. None of the 5 females had aortic dilatation measured at the sinuses. In the 4 males who had CT or MR imaging of the chest, the ascending aorta was relatively or absolutely more dilated than the sinuses segment in all 4. Two patients underwent prophylactic aortic replacement: a 50 y-o old man at a diameter of 55 mm and a 58 y-o man at a diameter of 50 mm. Serial echocardiograms in the males showed a tendency toward progressive dilatation of the sinuses over time. There was no correlation between aortic dilatation and any other clinical feature of Fabry disease, including renal function, cardiac hypertrophy or cardiac valvular function. Routine histopathologic examination of excised aortic valve and aortic wall tissue showed 'myxoid degeneration' in the 2 patients who underwent aortic surgery. We conclude that an aortopathy occurs in Fabry disease that is more pronounced in the mid- to distal ascending aorta. Since this region of the aorta is not easily imaged on routine transthoracic echocardiography, we recommend that all adult males with Fabry disease undergo CT or MR examination of the thorax at least once to define aortic anatomy. Since heterozygotes are at risk for developing all of the typical features, albeit on average 20 years later than hemizygotes, screening by CT or MR of the thoracic aorta in affected women may well be warranted. Finally, since all patients with Fabry disease should have annual echocardiography, attention should be paid to the proximal aorta and any other region of the aorta that can be visualized. The pathophysiology of the aortopathy is unclear, but could be primarily a defect in vascular smooth muscle cells from accumulation of glycosphingolipids.

347

High Frequency of Huntington disease (HD) Intermediate Alleles (IA) on Predisposing Haplotypes for Repeat Instability in British Columbia's (B.C.) General Population. *A. Semaka, C. Doty, N. Tam, J. Collins, C. Kay, S. Warby, M. Hayden.* University of British Columbia, Centre Molecular Medicine & Therapeutics, Vancouver, B.C., Canada.

BACKGROUND: IAs (27-35 CAG) do not cause HD but due to possible repeat expansion beyond 36 CAG, they may cause HD in future generations. The frequency of IAs has largely been determined in clinical databases of HD families; only recently was the frequency of IA chromosomes found to be 3.0% in Portugal's general population. New mutation (NM) IAs, which have been shown to expand into HD range, are believe to be more susceptible to repeat instability compared to general population (GP) IAs, which have been identified on the non-HD side of a family. Detailed SNP analysis has identified two haplotypes that predispose chromosomes to instability. Both HD and IA chromosomes were significantly enriched for these haplotypes compared to controls. AIMS: The aims of this study were to establish the frequency of IAs in B.C.'s general population and determine the proportion of these IAs found on predisposing haplotypes for repeat instability. We also sought to examine whether the proportion of NM and GP IAs in the UBC-HD Database found on predisposing haplotypes differed. METHODS: CAG sizing was performed on 1600 anonymous DNA samples randomly ascertained from B.C.'s general population. As described in Warby et al., haplotypes were constructed using 22 tagging SNPs located throughout the HD gene. RESULTS: 5.8% (n=92/1594) of individuals in B.C.'s general population had an IA. These IA chromosomes were enriched for haplotypes that predispose to repeat instability (61%). This proportion did not differ from the proportion of IAs in the UBC-HD Database on predisposing haplotypes (72%; p=0.2280). In the UBC-HD Database, both IAs classified as NM (60%) and GP (76%) were found more frequently on haplotypes associated with instability. There was no difference in the proportion of IAs on predisposing haplotypes between these two categories (p=0.3217). CONCLUSION: IAs are common amongst individuals not associated with HD. This supports the recommendation that IAs be discussed with all clients during their pre-result genetic counselling as a possible predictive test result. These results also suggest that there may be no difference in the susceptibility of NM and GP IAs to repeat expansion given that they are both frequently found on haplotypes that predispose to instability. Risk assessment of repeat expansion provided during genetic counselling should reflect this finding.

348

Enhancing Power for Gene Discovery in Alzheimer's Disease: Intermediate Phenotypes and Functional Validation in *Drosophila*. *J.M. Shulman[1,2,3], P. Chipendo[1,2,3], C. Aubin[1,2,3], P. Kramer[4], J.A. Schneider[5,6], D.A. Bennett[6], M.B. Feany[2,7], P.L. De Jager[1,2,3].* 1) Department of Neurology, Brigham and Women's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 4) Layton Aging and Alzheimer's Disease Center, Oregon Health and Science University, Portland, OR; 5) Department of Pathology, Rush University Medical Center, Chicago, IL; 6) Rush Alzheimer's Disease Center, Department of Neurological Sciences, Rush University Medical Center, Chicago, IL; 7) Department of Pathology, Brigham and Women's Hospital, Boston, MA.

BACKGROUND: Gene discovery in Alzheimer's disease (AD) has largely relied on the case/control study design; however, this approach is potentially confounded by genetic heterogeneity of dementia in cases and subclinical disease in controls. We have implemented a complementary strategy using quantitative AD pathology as an outcome phenotype for genome-wide association (GWA) analysis. Candidate risk loci were subsequently tested for functional validation in a *Drosophila* disease model, based on transgenic expression of human Tau.

METHODS: We conducted a pilot GWA scan with genotypes at 334,575 single nucleotide polymorphisms (SNPs) and autopsy data from 227 subjects of the Religious Orders Study and Rush Memory and Aging Project. The outcome phenotype was a continuous measure of global AD pathology, based on counts of amyloid plaques and neurofibrillary tangles from five brain regions. In order to validate candidate susceptibility loci, we tested lines predicted to activate or disrupt homologous fly genes for enhancement or suppression of Tau retinal toxicity in vivo. Validated SNPs were genotyped in an additional 305 subjects for replication of associations.

RESULTS: 6 out of 15 associated genomic regions contained candidate susceptibility genes showing genetic interactions with Tau neurotoxicity in vivo. Among our best results, a SNP (*rs10845990*) in the *SLC2A14* gene, encoding a glucose transporter, showed evidence of replication for association with Alzheimer's pathology (p_{DISC}=6.9x10^{-5}, p_{REP}=0.03, p_{JOINT}=8.1x10^{-6}), and both gain- and loss-of-function of *glut1*, the *Drosophila* homolog, were associated with suppression and enhancement of Tau toxicity, respectively.

CONCLUSIONS: We have identified several novel candidate AD susceptibility loci, based on their association with neuropathologic burden in humans and their activity in a relevant animal model system. Our strategy of coupling GWA analysis with a functional validation pipeline is likely to be a powerful approach for gene discovery in AD and other complex genetic disorders.

349

Causative gene of amyotrophic lateral sclerosis; Optineurin. *H. Maruyama[1], H. Morino[1], H. Ito[2], Y. Izumi[3], M. Kamada[1], K. Hagiwara[4], K. Abe[5], M. Aoki[6], S. Matsuura[7], H. Kawakami[1].* 1) Epidemiology, Hiroshima University, Research Institute for Radiation Biology and Medicine, Hiroshima, Japan; 2) Neurology, Kansai Medical University, Moriguchi, Japan; 3) Clinical Neuroscience, University of Tokushima Graduate School, Tokushima, Japan; 4) Respiratory Medicine, Saitama Medical University, Saitama, Japan; 5) Neurology, Okayama University, Graduate School of Medicine, Dentistry, and Pharmaceutical Sciences, Okayama, Japan; 6) Neurology, Tohoku University School of Medicine, Sendai, Japan; 7) Genetics and Cell Biology, Hiroshima University, Research Institute for Radiation Biology and Medicine, Hiroshima, Japan.

Amyotrophic lateral sclerosis (ALS) is a middle-age onset and progressive disorder characterized by degeneration of motor neurons of the primary motor cortex, brainstem, and spinal cord. Most cases of ALS are sporadic, but about 10% are familial. The genes known to be causative of classical familial ALS (FALS) are superoxide dismutase 1 (*SOD1*), *ANG* encoding angiogenin, *TARDP* encoding TAR DNA-binding protein TDP-43, and *FUS/TLS* (fused in sarcoma/translated in liposarcoma). We analyzed these genes in 53 Japanese FALS families. *SOD1* mutations (17%) are most frequent. Mutations of *TARDP* is 1.9%, *FUS/TLS* is 5.7%, and Spinal and bulbar muscular atropy is 3.8%. Genes which is causing remaining 72% FALS families are unknown. To find new causative gene of ALS, we used homozygosity mapping. We analyzed 6 Japanese individuals from consanguineous marriages who had ALS. We performed a genome-wide scan of single nucleotide polymorphisms (SNPs), and selected for the run of homozygous SNPs (RHSs). We extracted RHSs of 6 individuals. A region in chromosome 10, which was an overlap among 4 subjects, was chosen as the primary candidate region, and sequenced. There are mutations in the gene encoding optineurin (*OPTN*), earlier reported to be a causative gene of primary open-angle glaucoma (POAG), in ALS patients. We found 3 types of mutation of *OPTN*: a homozygous deletion of exon 5, a homozygous Q398X nonsense mutation, and a heterozygous E478G missense mutation within its ubiquitin-binding domain. Cell transfection analysis showed that the nonsense and missense *OPTN* mutations abolished the inhibition of activation of NF-κB, and the E478G mutation revealed a cytoplasmic distribution different from that of the wild type or a POAG mutation. A case with the E478G mutation showed OPTN-immunoreactive cytoplasmic inclusions. Furthermore, TDP-43 or SOD1-positive inclusions of sporadic and *SOD1* ALS cases were also noticeably immunolabeled by anti-OPTN antibodies. Our findings strongly suggest that OPTN is involved in ALS pathogenesis. Our findings also indicate that NF-κB inhibitors could be used to treat ALS and that transgenic mice bearing various *OPTN* mutations will be relevant for the development of new drugs to treat ALS.

350

VCP Syndrome associated with Inclusion body myopathy, Paget disease of bone, Frontotemporal dementia, Parkinson's and ALS: Clinical features, Pathogenesis and Mouse models. *V. Kimonis[1], A. Nalbandian[1], M. Badadani[1], GD. Watts[2], E. Dec[1], J. Tanaja[1], B. Martin[3], C. Smith[3], V. Caiozzo[4].* 1) Div Gen, Dept Pediatrics, Univ CA, Irvine, Orange, CA; 2) School of Medicine, Health Policy and Practice, University of East Anglia, Norwich, Norfolk, UK; 3) Department of Neurology, University of Kentucky Medical School, Lexington, KY, USA; 4) Department of Orthopedic Surgery, University of California, Irvine, CA, USA.

Hereditary inclusion body myopathy associated with Paget's disease of bone and frontotemporal dementia is an increasingly recognized important disorder. VCP a member of the AAA-ATPase super-family, is at the crossroads of many cellular functions including ubiquitin proteasome-mediated degradation, p62 associated autophagy and degradation of phosphorylated IkB-a, an essential step in NF-kB activation. Our studies in this fascinating disorder extends from clinical, cell models, to a knock-in VCP R155H mouse model that replicates the human disease. Progressive proximal muscle weakness with rimmed vacuoles and inclusions in muscle is present in >80% of individuals. PDB seen in 50% affects the spine, hips, scapulae and skull caused by overactive osteoclasts. FTD in 30% at a mean age of 55 y. is characterized by aphasia, dyscalculia, and comprehension deficits. Patholy includes ubiquitin and TDP43 positive inclusions and characteristic nuclear and cytoplasmic paired helical filaments. Microarray analysis of muscle tissue from affected individuals has revealed new insights in the pathogenesis of the myopathy. Genotype-phenotype correlations reveal marked intrafamilial variations and varied phenotypes including Parkinsons, cardiomyopathy, ALS, myotonia, cataracts and gastrointestinal complications. VCP is essential for maturation of ubiquitin-containing autophagosomes. Myoblasts from patients identified large ubiquitin containing vacuoles that accumulate LC3-II, a marker for autophagy. The vacuoles fuse with lysosomes as indicated by LAMP-1 and LAMP-2-staining; these protein however are differentially N-glycosylated. Additionally, mutant myoblasts show decreased proliferation activity, and increased apoptosis. Our R155H VCP knock-in heterozygous mice are a good model for preclinical studies and demonstrate progressive muscle, bone and brain disease. Mice have progressive weakness, vacuolization of myofibrils, centrally located nuclei, and cytoplasmic accumulation of TDP-43 and ubiquitin in myofibrils and brain. Bone micro-CT morphometrics shows decreased trabecular pattern and increased cortical wall thickness, and histology reveals increased osteoclastogenesis. Homozygosity of the R155H mutation is lethal by 21 days; these mice reveal abnormal skeletal muscle architecture and large vacuoles in the cardiac muscle. VCP is thus at the crossroads of pathology involving progressive muscle, bone and brain disease.

351

Clinical and Molecular genetic study of ATR-X syndrome Patients in JAPAN and proposal of clinical diagnostic criteria of ATR-X syndrome. *T. Wada[1], H. Shimbo[1], T. Mitani[1], H. Osaka[1], K. Kurosawa[2], N. Okamoto[3].* 1) Division of Neurology, Kanagawa Children's Med Center, Yokohama, Kanagawa, Japan; 2) Division of Genetics, Kanagawa Children's Med Center, Yokohama, Kanagawa, Japan; 3) Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan.

X-linked a-thalassemia/ mental retardation (ATR-X) syndrome (OMIM301040) is among X-linked mental retardation syndromes, which is due to mutations of ATRX gene.More than 200 patients in the world, including more than 60 Japanese cases, have been diagnosed as ATR-X syndrome. ATR-X syndrome seems a rare disease, however, more patients should remain to be diagnosed. In 2010, we have established ATR-X syndrome Japan Network for patients and their families, and we have surveyed ATR-X syndrome patients in Japan. Molecular genetically, we have found ATRX gene mutations in 56 patients from 45 families in JAPAN, which include 35 missense mutatinons, one nonsense mutation, 2 splicing mutations, one nonsense mutations, 4 frame shift mutations, one exonic deletion, and one large insertion. As reported previously, most mutations reside in two functionally important regions, ADD domains and chromatin remodeling domain, but some mutations are out of these domains. A nucleotide, IVS34+4A>G, leading to skipping of exon 34. Interestingly the patient with this mutation showed no severe genital abnormalities, whereas previously reported patients with mutations around this C-terminal region showed severe genital abnormalities. Clinically, from the medical information of more than 60 Japanese ATR-X patients registered in our database, we propose a diagnostic criteria of ATR-X syndromes to assist the clinician in making a timely and accurate diagnosis. This criteria consists of four clinical features; consistent (>90%), frequent (>50%), associated (<50%), and other supporting features. The consisitent feature includes male patient, severe mental retardation, hypotonic facies, gastorointestinal abnormalities (gastroesophageal reflux, aerophagia, ileus, constipatition, etc), and characteristic behavior (autistic behavior, self-vomiting, avoiding eye contact, stereotype movement, etc). We hope that this criteria will facilitate further clinical study of individuals with proven ATR-X molecularly, and assist in the evaluation of those who appear to have clinical features of ATR-X with no ATRX mutations. We should evaluate validity of this criteria, and update the consensus for it in the near future.

352

Haploinsufficiency of *HDAC4* results in brachydactyly mental retardation syndrome, including type E brachydactyly, developmental delays, and behavioral problems. *S. Williams[1], M. Aldred[2], V. Der Kaloustian[3,4], F. Halal[5], G. Gowans[6], R. McLeod[7], S. Zondag[8], H. Toriello[8], E. Magenis[9], S. Elsea[1,10].* 1) Department of Human and Molecular Genetics, Virginia Commonwealth University School of Medicine, Richmond, VA; 2) Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH; 3) Dept. of Pediatrics, The Montreal Children's Hospital, McGill University, Montreal, Quebec, Canada; 4) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 5) Division of Medical Genetics, The Montreal Children's Hospital, McGill University, Montreal, Quebec, Canada; 6) Weisskopf Child Evaluation Center, Department of Pediatrics, University of Louisville, Louisville, KY; 7) Department of Medical Genetics, University of Calgary, Calgary, Alberta, Canada; 8) Genetic Services, Spectrum Health, Grand Rapids, MI; 9) Department of Genetics, Oregon Health and Science University, Portland, OR; 10) Dept. of Pediatrics, Virginia Commonwealth University School of Medicine, Richmond, VA.

Brachydactyly mental retardation syndrome (BDMR) is associated with a deletion involving chromosome 2q37. BDMR presents with a range of features, including intellectual disabilities, developmental delays, behavioral abnormalities, sleep disturbance, craniofacial and skeletal abnormalities, including brachydactyly type E, and autism spectrum disorder. To date, only large deletions of 2q37 have been reported, making the delineation of a critical region and subsequent identification of candidate genes difficult. We present clinical and molecular analysis of 6 individuals with overlapping deletions involving 2q37.3 which refine the critical region, reducing the candidate genes from >20 to a single known gene, *histone deacetylase 4* (*HDAC4*). In addition, driven by the distinct hand and foot anomalies and similar cognitive features, we identified other cases with clinical findings consistent with BDMR but without a 2q37 deletion. These subjects presented with intellectual disability, developmental delays, craniofacial abnormalities, obesity, sleep disturbance and were initially referred for a Smith-Magenis (SMS) phenotype. Sequencing of *HDAC4* in these two cases identified *de novo* mutations, including one intragenic deletion likely disrupting normal splicing and one intragenic insertion that results in a frame-shift and premature stop codon. HDAC4 is a histone deacetylase that regulates many genes important in bone, muscle, neurological, and cardiac development. Supporting these clinical and molecular findings are *Hdac4*$^{-/-}$ mice which have severe bone malformations due to premature ossification of developing bones caused by ectopic and early onset chondrocyte hypertrophy. Our data also show that mutation or deletion of *HDAC4* results in reduced *retinoic acid induced 1* expression (*RAI1*). *RAI1* is the gene that when haploinsufficienct results in Smith-Magenis syndrome, indicating shared pathways between these two phenotypically linked syndromes. Considering the known molecular function of HDAC4 and the mouse knockout phenotype, taken together with deletion or mutation of *HDAC4* in multiple subjects with BDMR, we conclude that haploinsufficiency of *HDAC4* results in brachydactyly mental retardation syndrome and propose that this pathway is integral to many syndromes with overlapping behavioral and physical phenotypes.

353

The CASK gene mutation in ten Japanese cases with severe mental retardation, microcephaly and disproportionate pontine and cerebellar hypoplasia (MIC-PCH). *S. Hayashi[1], O. Nobuhiko[2], S. Mizuno[3], Y. Chinen[4], J. Takanashi[5], Y. Makita[6], A. Hata[7], I. Imoto[1,8], J. Inazawa[1].* 1) Department of Molecular Cytogenetics, Medical Research Institute, Tokyo Medical and Dental University; 2) Department of Planning and Research, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan; 3) Department of Pediatrics, Central Hospital, Aichi Human Service Center, Aichi, Japan; 4) Department of Pediatrics, University of the Ryukyu School of Medicine, Okinawa, Japan; 5) Department of Pediatrics, Kameda Medical Center, Chiba, Japan; 6) Education Center, Asahikawa Medical College, Hokkaido, Japan; 7) Department of Public Health, Chiba University Graduate School of Medicine, Chiba, Japan; 8) Department of Human Genetics and Public Health, Graduate School of Medical Science, the University of Tokushima, Tokushima, Japan.

The CASK gene encoding a member of the MAGUK (membrane-associated guanylate kinase) proteins is highly expressed in the mammalian nervous system of both adults and fetuses, and to play several roles in neural development and synaptic function. Recently, CASK aberration caused by both mutation and genomic copy-number aberration is reported to cause severe mental retardation (MR), microcephaly and disproportionate pontine and cerebellar hypoplasia (MIC-PCH) in females. In order to investigate a correlation between CASK aberration and phenotypes, we recruited ten female patients with MR and MIC-PCH, and carried out mutation and copy-number analyses of CASK. The following aberrations involved in CASK were detected; nonsense mutations in three cases, 2bp-deletion in one case, mutations at exon-intron junction in two cases, heterozygous deletions encompassing CASK in two cases, tandem interstitial duplication in one case and inverted duplication in one case. The 2bp-deletion, the mutations at exon-intron junction and the tandem interstitial duplication caused a frame-shift producing truncated form of CASK. The inverted duplication also caused a truncated form. As a result, all of the CASK aberrations were confirmed to cause potentially haploinsufficiency of CASK. It may explain the reason of the similarity of the phenotypes of MR and MIC-PCH regardless of variety of the CASK aberration. The concordance between genotypic variety and phenotypic similarity would be clinically profitable to detect and investigate a potential patient with CASK aberration.

354

Sequence variants in ZIC2 in Brazilian patients within the phenotypic spectrum of holoprosencephaly. *L.A. Ribeiro[1], P. Hu[2], E. Roessler[2], A. Richieri-Costa[1], M. Muenke[2].* 1) Genetics, Hospital for Rehabilitation of Craniofacial Anomalies/USP, Bauru, Brazil; 2) National Human Genone Research Institute, National Institute of Health, Bethesda, MD, USA.

Holoprosencephaly (HPE) is a common developmental defect affecting both the forebrain and the face. It is characterized by the incomplete separation of the cerebral hemispheres into distinct right and left halves. Other derivatives of the forebrain, including the thalamus and pituitary, can also develop abnormally. The disease is genetically heterogeneous but additional environmental agents also contribute to the etiology of HPE. Four major types of holoprosencephaly have been well established through neuroradiologic studies: alobar, semilobar, lobar, and middle interhemispheric variant. Autosomal dominant, recessive, and X-linked families have been reported. Currently, four genes (SHH, ZIC2, SIX3, and TGIF) are routinely analyzed on a clinical basis in patients with HPE. ZIC2, located at chromosome 13q32, was first identified as an HPE candidate gene due to individuals with brain anomalies who were found to have deletions involving the long arm of chromosome 13. Subsequent analysis has identified nearly 90 different mutations in ZIC2 in patients with HPE. ZIC2 mutations are thought to be second most common identified cause of non-chromosomal non-syndromic HPE (after mutations in SHH). In recent estimates, at least 3% of probands with HPE have mutations in ZIC2, though a more accurate estimate is likely at least double that. Here approximately 90 probands with HPE were screened for ZIC2 gene by direct sequencing. Six sequence variants were found in eleven unrelated individuals: c.528C>T (1 individual), c.1239+18G>A (1 individual), c. 716_718dup (6 individuals), c.1215dupC (1 individual), c.1369_1383del (1 individual), and a complex indel likely resulting in an alanine expansion (1 individual). The patient who presented c.716_718dup also presented the c.653T>A mutation in the SHH gene. All HPE types were identified by neuroimaging performed by CT or MRI. Our findings reveal that the phenotypic HPE spectrum was variable between these patients raging from the HPE-Like to the more typical HPE phenotypes. This analysis contributes to new data on ZIC2 and enriches our understanding of HPE. Grants: Fapesp (06/60973-9) and CNPq (307595/2008-0).

355

Adaptor protein complex-4 (AP-4) deficiency causes a novel autosomal recessive cerebral palsy syndrome with microcephaly and intellectual disability. *A. Moreno-De-Luca[1], S.L. Helmers[2], H. Mao[3], T.G. Burns[4], A.M. Melton[4], K.R. Schmidt[1], P.M. Fernhoff[1], D.H. Ledbetter[1], C.L. Martin[1].* 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 2) Department of Neurology, Emory University School of Medicine, Atlanta, GA; 3) Department of Radiology, Emory University School of Medicine, Atlanta, GA; 4) Department of Neuropsychology, Children's Healthcare of Atlanta, Atlanta, GA.

Cerebral palsy (CP) is a heterogeneous group of developmental brain disorders resulting in motor and posture impairments often associated with cognitive, sensorial and behavioral disturbances. CP is the most common physical disability of childhood, with a prevalence of 2-3/1,000 births and an estimated 750,000 or more affected American children and adults. Although the etiology of CP has been attributed to a variety of factors, the specific mechanism underlying most cases remains unknown. Hypoxic-ischemic injury, long considered the most frequent causative factor, actually accounts for fewer than 10% of cases, whereas as many as 50% of idiopathic CP cases in children born at term are now thought to be caused by genetic factors. Here, we describe a novel autosomal recessive form of spastic tetraplegic CP in a consanguineous family with two affected individuals who have a homozygous deletion of 15q21.2 involving only two genes: *AP4E1* (ε subunit of the heterotetrameric AP-4 complex) and *SPPL2A* (signal peptide peptidase-like 2A). Both patients have profound intellectual disability, microcephaly, epilepsy, white matter loss and cerebellar atrophy. The *AP4E1* gene is a member of the AP-4 complex which is comprised of four subunits: AP4E1, AP4M1, AP4B1 and AP4S1. Evidence for *AP4E1* as the causative gene in our patients includes a previous report of a homozygous mutation in the AP4M1 subunit, identified in an inbred sibship with CP and a very similar clinical presentation (Verkerk et al. 2009), and a mouse knockout of AP4B1 that causes dysfunction of the entire AP-4 complex and characteristic neuroaxonal abnormalities (Matsuda et al. 2008). As demonstrated before for other AP-complexes (AP-2 and AP-3), we propose that disruption of any one of the four subunits of AP-4 results in loss of function of the complex, leading to a distinct "AP-4 deficiency syndrome" characterized by spastic tetraplegic CP, microcephaly, intellectual disability, cerebellar atrophy and white matter loss. Comprehensive mutation analysis, including resequencing and high-resolution deletion/duplication analysis of the genes encoding the four subunits of AP-4, is being pursued on individuals with similar phenotypes to determine the frequency of AP-4 deficiency syndrome. In conclusion, we describe a novel CP syndrome and add AP-4 complex deficiency to the growing body of evidence for multiple genetic etiologies of CP.

356

A genome-wide linkage and association study of serum substance P levels identifies potential candidate genes in the Genetics of Brain Structure Study. *V.P. Diego[1], D.C. Glahn[2,3], J.W. Kent Jr.[1], J.E. Curran[1], R.L. Olvera[4,5], T.D. Dyer[1], L.A. Almasy[1], E.K. Moses[1], H.H.H. Göring[1], M.A. Carless[1], J. Blangero[1].* 1) Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX 78227; 2) Department of Psychiatry, Yale University School of Medicine, New Haven, CT 06511; 3) Olin Neuropsychiatry Research Center, Institute of Living, Hartford Hospital, Hartford, CT 06106; 4) Division of Child and Adolescent Psychiatry, Department of Psychiatry, University of Texas Health Science Center at San Antonio, TX 78229; 5) Research Imaging Institute, University of Texas Health Science Center, San Antonio, TX 78229.

Substance P (SP) is a small neuropeptide transmitter comprised of 11 amino acids that is expressed in central and peripheral nervous systems. SP is classified as a neurokinin (NK), and its primary receptor is the NK-1 receptor (NK1R). Based on the observation that SP and NK1R are jointly expressed in neuroanatomical regions associated with depression and anxiety, SP has been hypothesized to play an important role in the etiology of these psychiatric disorders. This hypothesis is strongly supported by data on the effects of SP and/or NK1R in animal models of anxiety and stress, and by data from preclinical and clinical studies of SP and/or NK1R in humans with depression or anxiety disorders. While the data on the role of SP in the molecular etiology of affective disorders are quite compelling, the information regarding the genetic regulation of variation in SP levels are comparatively scant. To better understand the genetics of serum SP levels, which are highly correlated with SP levels in cerebrospinal fluid, we performed a genome-wide linkage and association study (GWLAS) of serum SP levels on 854 randomly-selected individuals from large Mexican-Americans pedigrees living in San Antonio who participated in the Genetics of Brain Structure and Function Study. Serum SP levels were transformed to normality by way of the inverse Gaussian transformation applied to residuals after accounting for age, sex, and their interactions as covariates. Using a maximum likelihood variance components model, we found that these normalized serum SP levels exhibited a heritability of 0.26 ($p = 6.8 \times 10\text{-}9$). We then performed a GWLAS on about 542,994 single nucleotide polymorphisms (SNPs) using a joint linkage and measured genotype association approach. Our best results, which all achieved genome-wide significance (where the adjusted significance threshold taking into account the non-independence of SNPs in linkage disequilibrium is about $1.3 \times 10\text{-}7$), were for two SNPs at the ABO blood group locus, namely rs657152 ($p = 1.4 \times 10\text{-}9$) and rs505922 ($p = 4.2 \times 10\text{-}9$) located at chromosome 9q34.2, two SNPs at the *PHOSPHO2* gene, namely rs4668156 ($p = 2.6 \times 10\text{-}8$) and rs7590594 ($p = 7.5 \times 10\text{-}8$) located at chromosome 2q31.1, and one SNP at the *PPIG* gene, namely rs6747350 ($p = 1.8 \times 10\text{-}7$) also located at chromosome 2q31.1. We are currently testing these associated SNPs to identify pleiotropic effects on our extensive collection of neuroimaging and cognitive phenotypes.

357

SHOC2 mutation analysis in Noonan-like syndrome and hematologic malignancies. *S. Komatsuzaki[1], Y. Aoki[1], T. Niihori[1], N. Okamoto[2], R.CM. Hennekam[3,4], S. Hopman[4], H. Ohashi[5], S. Mizuno[6], Y. Watanabe[7], H. Kamasaki[8], I. Kondo[9], N. Moriyama[10], K. Kurosawa[11], H. Kawame[12], M. Imaizumi[13], T. Rikiishi[14], S. Tsuchiya[14], S. Kure[1,14], Y. Matsubara[1].* 1) Department of Medical Genetics, Tohoku University School of Medicine, Sendai, Japan; 2) Osaka Medical Center and Research Institute for Maternal and Child Health, Izumi, Osaka, Japan; 3) Great Ormond Street Hospital, London, UK; 4) University of Amsterdam, Netherlands; 5) Saitama Children's Medical Center, Saitama, Japan; 6) Central Hospital, Aichi Human Service Center, Aichi, Japan; 7) Kurume University School of Medicine, Kurume, Japan; 8) Sapporo Medical University, Sapporo, Japan; 9) Oida Hospital, Kochi, Japan; 10) Hitachi Ltd. Mito General Hospital, Ibaraki, Japan; 11) Kanagawa Children's Medical Center, Yokohama, Japan; 12) Ochanomizu University, Tokyo, Japan; 13) Miyagi Children's Hospital, Sendai, Japan; 14) Department of Pediatrics, Tohoku University School of Medicine, Sendai, Japan.

Noonan syndrome is an autosomal dominant disease characterized by dysmorphic features, webbed neck, cardiac anomalies, short stature and cryptorchidism. Noonan syndrome has phenotypic overlap with Costello syndrome and cardio-facio-cutaneous (CFC) syndrome. Germline mutations of genes encoding molecules in the RAS/MAPK pathway cause Noonan syndrome and related disorders. Germline mutations in *PTPN11*, *KRAS*, *SOS1* and *RAF1* have been identified in 60-80% of Noonan syndrome patients. Recently, a gain-of-function of *SHOC2* p.S2G (c.4A>G) has been identified as a causative gene for Noonan-like syndrome with loose anagen hair. In order to understand the contribution of SHOC2 mutations to the clinical manifestations of Noonan syndrome and related disorders, we analyzed *SHOC2* in 92 patients with Noonan syndrome and related disorders who did not exhibit *PTPN11*, *KRAS*, *HRAS*, *BRAF*, *MAP2K1/2*, *SOS1* or *RAF1* mutations. We found the previously identified p.S2G mutation in eight of our patients. Parental samples were available in 3 families and the mutation was not identified in parents. We developed a rapid detection system to identify the p.S2G mutation using melting curve analysis, which will be a useful tool to screen for the apparently common mutation. All the patients with the p.S2G mutation showed short stature, sparse hair and atopic skin. Seven of the mutation-positive patients showed severe mental retardation and easily pluckable hair. No *SHOC2* mutations were identified in leukemia cells from 82 leukemia patients. These results suggest that clinical manifestations in *SHOC2* mutation-positive patients partially overlap with those in patients with typical Noonan or CFC syndrome and show that easily pluckable/loose anagen hair is distinctive in *SHOC2* mutation-positive patients.

358

Biting into Costello syndrome: Ras signaling regulates enamel deposition in humans and mice. *O. Klein[1], A. Goodwin[1], S. Oberoi[1], C. Charles[1], J. Groth[1], C. Fairley[1], X. Chen[3], J. Fagin[3], K. Rauen[2].* 1) Orofacial Sciences, Univ California, San Francisco, San Francisco, CA; 2) Pediatrics, Univ California, San Francisco, San Francisco, CA; 3) Human Oncology, Memorial-Sloan Kettering Cancer Center, New York, NY.

Ras/MAPK signaling is critical in animal development, and receptor-tyrosine kinase signaling, which activates Ras signaling, is known to play an important role in tooth development. Our previous work has shown that increasing Ras/MAPK signaling by inactivating Sprouty genes adversely affects tooth morphogenesis. Here, we directly examined the effects of activating Ras/MAPK signaling in both humans and mice. Costello Syndrome (CS) is caused by a heterozygous de novo germline mutation in HRAS that results in a constitutively active Ras protein. We examined a cohort of CS patients and identified a number of craniofacial and dental anomalies. We found that a large majority of patients presented with pronounced enamel hypoplasia. Micro computed tomography of exfoliated primary teeth from CS patients showed a significant decrease in enamel thickness compared to controls. We next examined the CS mouse model and found that the mice also had an enamel defect. Further inspection revealed disorganization of the ameloblasts in the mouse incisor. We are currently studying cell proliferation and polarity of the amelobasts in the mutant mice incisors and using an ameloblast-like cell line. Together, our studies point to a role for Ras signalling in regulation of cell polarity and deposition of mineralized matrices.

359

Applying Novel Genomic Tools Towards Understanding an Old Chromosomal Diagnosis: Using Genome-wide Expression and SNP Genotyping to Identify the True Cause of Pallister-Killian Syndrome. *M. Kaur[1], L. Conlin[2], S. Nancy[2], M. Deardorff[1], C. Fincher[1], A. Wilkens[1], Z. Zhang[3], I. Krantz[1].* 1) Dept of Genetics, Children's Hosp Philadelphia, Philadelphia, PA; 2) Dept of Pathology, Children's Hosp Philadelphia, Philadelphia, PA; 3) Dept of Bioinformatics, Children's Hosp Philadelphia, Philadelphia, PA.

Pallister-Killian Syndrome (PKS) (OMIM # 601803) is a rare sporadic multisystem developmental disorder characterized cytogenetically by tissue limited mosaic partial de novo tetrasomy of 12p. The spectrum of clinical manifestations in PKS is highly variable and includes craniofacial dysmorphia, clefts, ophthalmologic, audiologic, cardiac, musculoskeletal, diaphragmatic, gastrointestinal, genitourinary, and cutaneous anomalies in association with cognitive retardation and seizures. Growth parameters are often normal to elevated at birth with deceleration of growth postnatally. The prevalence of PKS has been estimated to be ~1/20,000 live births but is likely under-ascertained since tetrasomy 12p is often not present in the blood and requires fibroblast or other tissue sampling to identify. PKS can also be diagnosed prenatally by CVS, amniocentesis and cordocentesis. We follow a cohort of 46 children with PKS and have initiated genome-wide analyses to better understand the manner in which additional copies of chromosomes 12p result in the highly specific PKS phenotype. Through SNP genotyping (using the Illumina 610 array) we have been able to accurately quantify degrees of mosaicism (and loss of the isochromosome) from skin and blood, determine parent of origin and meiotic phase of non-dysjunction. We have also used these arrays to define a PKS critical region on chromosome 12p13.31 encompassing approximately 30 genes. Parallel to these studies we have performed genome-wide expression array analysis (using Affymetrix HG-U133 plus 2.0 GeneChip arrays) on 17 proband's and 9 control fibroblast cell lines. From these studies 354 genes were identified that were statistically significantly dysregulated (180 up-regulated and 174 down-regulated). Of the 180 up-regulated genes, 57 were located (not surprisingly) on chromosome 12p, fitting the pathology of PKS. The most statistically significantly dysregulated genes on 12p mapped to within the SNP-array defined critical region on 12p13.31. Amongst the top ranked dysregulated genes and gene clusters were families of homeobox genes (not on chromosome 12) and developmental genes. The level of mosiacism was confirmed by microarray and FISH analysis on cultured fibroblasts. We will present data linking candidate genes on 12p13.31 with clusters of important developmental genes elsewhere in the genome towards the creation of cause and effect molecular map for PKS.

360

Plexiform Neurofibromas Morbidity Outcomes in a Large Cohort of Pediatric Patients. *C. Prada, F. Rangwala, E. Sites, A. Gallas, A. Lovell, E. Schorry, H. Saal, R. Hopkin.* Dept Pediatrics Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

Plexiform neurofibromas (PNFs) are one of the most common and debilitating complications of neurofibromatosis type 1 (NF1). Large PNFs can compress vital organs and result in severe morbidity and even death. The objective of this study was to characterize PNFs morbidity outcomes and the presence of additional tumors in a large cohort of pediatric patients. A retrospective analysis of clinical information, tumor burden, surgical history and other tumors detected in children with PNFs who were seen at the NF1 Center at Cincinnati Children's between 1996 and 2005 was performed. A total of 154 children with NF1 and PNFs were identified. Clinical identification of PNFs had a bimodal age distribution with one peak in the early infancy (1-3 years) and a second peak during adolescence (11-13 years). Mediastinal tumors were diagnosed only during infancy. Average age at detection of PNFs was 5.5 years. Surgical interventions were performed in 62.3% of affected individuals. The most common morbidities related to PNFs were neurological (26.5%) and disfigurement (25%). PNFs of the head and neck represented 38% of the total tumors. Other common locations included extremities (22%) and paraspinal region (11%). Surgical resection was performed in 42% of the head and neck tumors; 7 individuals required tracheostomy placement. Tumor recurrence rate was 31%. The most significant improvement was relief of airway compression (77%), followed by improved physical appearance (61%) and relief of chronic pain (56%). Forty four percent had no improvement after surgical interventions. There were 42 new complaints after surgery, most of them neurological deficits. A subset of 57 (37%) children had other tumors. 10 (6.5%) children had three different tumor types. The most common tumors were optic pathway gliomas (24%), gliomas (9%), malignant peripheral nerve sheath tumor (2.6%), and neuroblastoma (2%). In summary, this study provides baseline clinical characteristics for evaluation of treatment outcomes. PNFs have a bimodal age distribution suggesting differential tumor growth during infancy and adolescence. Patients with PNFs were more likely to benefit from surgery if the indications were to relieve symptoms caused by airway compression or physical appearance. Surgical interventions were required in many cases and often resulted in a substantial burden for patients, emphasizing the need for development of early screen guidelines and effective non-surgical therapy for PNFs.

361

Submicroscopic chromosomal rearrangements in patients with a Prader-Willi syndrome-like phenotype. *K. Hosoki[1], T. Ohta[2], N. Niikawa[2], S. Saitoh[1].* 1) Department of Pediatrics, Hokkaido University Graduate School of Medicine, Sapporo, Japan; 2) The Research Institute of Personalized Health Sciences, Health Sciences University of Hokkaido, Tobetsu, Japan.

Prader-Willi syndrome (PWS) is characterized by neonatal hypotonia, mental retardation, and childhood hyperphagia resulting in obesity. A PWS-like phenotype has been reported to be associated with other chromosomal abnormalities, including maternal uniparental disomy 14 [upd(14)mat] and 1p36 deletions. In a series of molecular investigations of patients suspected to have PWS, we have collected 78 patients with a PWS-like phenotype for whom PWS was ruled out by a normal *SNURF-SNRPN* DNA methylation test. We first examined *MEG3* DNA methylation and detected 6 patients with the upd(14)mat syndrome. For the remaining patients with a PWS-like phenotype, we obtained parental consent in 64 patients to perform whole genome microarray analysis using Genome-Wide Human SNP Array 5.0 (Affymetrix). We detected 17 patients who had various copy number variation (CNV) rearrangements larger than 1 Mb (1p36delx2, 1q43del, 2q24del, 4q21del, 7p15del, 15q13del, 22q11.2del, Xq28dup, 1q21.1del, 1q21.1dup, 16q13.11dup, 1q22del, 1q42dup, 3q31del, 5q31.3delx2). While 1p36del has been reported to be associated with a PWS-like phenotype, many of these CNV (1q43del, 2q24del, 4q21del, 7p15del, 15q13del, 22q11.2del, and Xq28dup) have been identified in patients with distinct MR/MCA phenotypes. Three additional CNV (1q21.1del, 1q21.1dup and 16q13.11dup) have been reported to be associated with MR or an autism phenotype, but also to be present in normal populations. Indeed, in our study, 1q21.1dup was also identified in a phenotypically normal mother. Five CNV events, 1q22del, 1q42dup, 3q31del and 5q31.3delx2, have not been reported and these rearrangements arose *de novo*. Thus, these CNV rearrangements could represent novel microdeletion or microduplication syndromes demonstrating a PWS-like phenotype. Our results suggest that many submicroscopic chromosomal rearrangements may demonstrate a PWS-like phenotype, especially during infancy.

362

CNV and Expression Analysis of Familial 22q11.2 Deletions. *S.C. Saitta[1,2], T. Busse[1], R.G. Knowlton[1], D.M. McDonald-McGinn[1], E.H. Zackai[1,2], E. Goldmuntz[2,3], B.S. Emanuel[1,2], D.A. Driscoll[1,4].* 1) Division of Genetics, Childrens Hosp Philadelphia, Philadelphia, PA; 2) Dept of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA; 3) Division of Cardiology, Children's Hosp Philadelphia, Phila PA; 4) Dept of OB-GYN, Univ Penn School of Medicine.

Chromosome 22q11.2 microdeletions are associated with DiGeorge/Velocardiofacial syndrome. Deletions transmitted from affected parents or familial deletions, occur in ~10% of patients and often show discordant phenotypes within families. To better understand this variability, we used SNP arrays to study 72 individuals from 27 families where a parent and one or more children were affected. Deletion sizes were identical in families and the standard LCR-A to D deletion was the most prevalent, suggesting that familial deletions occur by the same mechanism as de novo deletions. We identified secondary CNVs in our cohort. In one family, a 3 Mb secondary deletion of chr5p14.3 was found in addition to a 22q11 deletion in an affected mother with epilepsy and emotional disorder, but not in her chr22 deleted daughter who does not share these features. Of the 4 genes in the 5p14 interval, 2 are pseudogenes, PRDM9 is a poorly characterized gene while CDH12 encodes N-cadherin 2, a cell adhesion molecule expressed in neuronal tissues. In addition, both mother and daughter show a chr6q26 deletion containing an exon from PARK2 and the bidirectional promoter shared with its neighbor, PACRG. These genes are involved in CNS development and PARK2 is associated with autosomal recessive Parkinson's disease. In another 22q11 deletion family, an affected father demonstrates an ~850kb amplification in Xp22.2 not present in his son, which includes several exons of the MID1 gene, associated with Opitz G/BBB syndrome. The duplication may disrupt the gene and affect the patient's phenotype. Our data suggests that CNVs throughout the genome can potentially modify the 22q11 deletion phenotype. We further performed Affymetrix exon arrays on a subset of familial deletion lymphoblast samples and found evidence for expression patterns that may correlate with specific phenotypes in our cohort. Overall, utilization of high-throughput techniques has enormous potential to uncover variation throughout the genome that may help explain discordant phenotypes in families and lead to improved outcome prediction and recurrence risk counseling.

363

22q11 Locus Physically Interacts with Candidate Disease-Modifying Genes in *trans*. *B.N. Steelman, M.J. Zeitz, W.B. Clayton, J.Q. Ling, A.R. Hoffman.* Department of Medicine, Stanford School of Medicine.

The 22q11 Deletion Syndrome is characterized by remarkable phenotypic variation, despite often identical genetic lesions. Genetic analyses have not uncovered the basis of this phenotypic variation. To search for candidate disease-modifying genes, we used a molecular assay called the Associated Chromosome Trap (ACT) to identify long-range chromatin interactions that involve the 22q11 locus. We found that this locus physically interacts with a hub of at least 12 genes that are distributed over 8 chromosomes. We propose that the 22q11 locus normally regulates gene expression in trans through physical chromatin interactions, and that haploinsuficiency of the 22q11 locus leads to dysregulation of disease-modifying loci.

Using ACT, we identified candidate genes that regulate the TGF-beta and BMP signaling pathways, which was intriguing because disruption of these pathways is known to recapitulate many aspects of the 22q11 Deletion Syndrome. Another candidate gene is involved in neuronal plasticity, which may provide an interesting link to the intellectual and psychiatric components of the deletion syndrome. Interestingly, the 22q11 locus physically interacts with components of both the HAT complex and the HDAC complex, which have antagonistic roles in histone acetylation.

We used 3D-FISH to characterize gene colocalization frequencies in fibroblasts and B-lymphocytes derived from syndromic patients and apparently healthy controls. We then used qPCR to correlate expression of the candidate genes with changes in gene colocalization. Finally, we found that some of the candidate genes are upregulated by TGF-beta, and that patient-derived fibroblasts have delayed responses to TGF-beta signaling.

Long-range chromatin interactions rely on both DNA sequence and epigenetic modifications. They therefore represent a unique opportunity to identify regulatory interactions between seemingly disparate DNA elements, even when the underlying mechanisms of those interactions are unknown. We propose that DNA sequence variation and epigenetic regulation can be captured by long-range chromatin interactions, and that long-range chromatin interactions could explain a substantial component of phenotypic variation.

364

Evaluation of alpha-synuclein aggregation in brain samples from patients carrying *GBA* mutations. *J. Choi[1], M. Cookson[2], G. Lopez[1], E. Goldin[1], O. Goker-Alpan[1], B. Stubblefield[1], E. Sidransky[1].* 1) Section on Molecular Neurogenetics, Medical Genetics Branch, NHGRI, National Institutes of Health, Bethesda, MD; 2) Cell Biology and Gene Expression Unit, Laboratory of Neurogenetics, NIA, National Institutes of Health, Bethesda, MD.

Recent findings demonstrate an increased frequency of mutations in glucocerebrosidase (*GBA*), the enzyme deficient in the lysosomal storage disorder, Gaucher disease (GD), among patients with synucleinopathies. Neuropathologic findings in some patients who developed both GD and parkinsonism revealed Lewy bodies and synuclein-positive inclusions in vulnerable brain regions. The association of *GBA* mutations and alpha-synuclein aggregation was examined by evaluating pathologic specimen. In this study, proteins were extracted from cerebral cortex from subjects carrying *GBA* mutations with and without a clinical history of parkinsonism. These include samples from patients with Lewy body dementia, Lewy body variant Alzheimer's disease, Parkinson disease and patients with different types of Gaucher disease. A total of 26 brain tissue samples were analyzed. 11 were from patients without Lewy body disorders (eight had *GBA* mutations and three were controls), and 15 had Lewy body disorders (nine carried *GBA* mutations). The samples were homogenized and fractionated into TBS-soluble, SDS-soluble and urea-soluble fractions. Most patients with synucleinopathies were shown to exhibit oligomeric forms of alpha-synuclein in the insoluble fraction, including the subjects with *GBA* mutations. However, patients with Gaucher disease and no clinical evidence of parkinsonism did not display oligomeric forms. The amount of the oligomeric alpha-synuclein correlated best with the degree of Lewy body pathology. These studies indicate that patients with syncleinopathies carrying *GBA* mutations show biochemical characteristics typical of Lewy body disorders, but these changes are not always seen with glucocerebrosidase deficiency.

365

Keratosis Follicularis Spinulosa Decalvans is caused by mutations in MBTPS2. *E. Aten[1], L. Brasz[1], D. Bornholdt[2], I.B. Hooijkaas[1], M.E. Porteous[3], V.P. Sybert[4], M.H. Vermeer[5], R.H.A.M Vossen[1], M.J.R. van der Wielen[1], E. Bakker[1], M.H. Breuning[1], K.H. Grzeschik[2], J.C. Oosterwijk[6], J.T. den Dunnen[1].* 1) Center of Human and Clinical Genetics, Leiden University Medical Center, Leiden, the Nederlands; 2) Department of Human Genetics, Philipps-Universität, Marburg, Germany; 3) Department of Clinical Genetics, West General Hospital, Edinburgh, United Kingdom; 4) Department of Medicine, University of Washington and Department of Dermatology, Group Health Permanente, USA; 5) Department of Dermatology, Leiden University Medical Center, the Netherlands; 6) Department of Clinical Genetics, University Medical Center Groningen, Groningen, the Netherlands.

Keratosis Follicularis Spinulosa Decalvans, (KFSD, OMIM 308800) is a rare genetic disorder characterized by development of hyperkeratotic follicular papules on the scalp followed by progressive alopecia of the scalp, eyelashes, and eyebrows with variable de-grees of inflammatory change. Associated eye findings include blepharitis, ectropion, photophobia in childhood and corneal dystrophy. KFSD closely resembles a group of disorders with abnormal keratinisation and sometimes eye symptoms, including Ichthyosis Follicularis with Atrichia and Photophobia syndrome (IFAP) and keratosis pilaris atrophicans. Due to both the genetic and clinical heterogeneity of these disordes, a defini-tive diagnosis of KFSD is often challenging. Towards identification of the causative gene we applied recently developed technologies and re-analysed a large Dutch KFSD family. 1M SNP arrays redefined the locus to a 2.9 Mb region at Xp22.12-Xp22.11 and excluded the presence of large deletions/duplications. Screening of all 14 genes in the candidate region using High Resolution Melting curve Analysis (HRMA) identified MBTPS2 as the candidate gene carrying a c.1523A>G (p.Asn508Ser) missense mutation. The variant was also identified in two unrelated X-linked KFSD families from the UK and USA and cosegregated with KFSD in all families. The c.1523A>G variant did not influence RNA splicing. In symptomatic female carriers, skewed X-inactivation of the normal allele matched with increased severity of symptoms. MBTPS2 encodes an integral membrane metalloprotease that is required for cleavage of sterol regulatory element-binding proteins (SREBPs). In vitro functional expression studies of the c.1523A>G mutation showed that sterol responsiveness was reduced by half. Other missense mutations in MBTPS2 have recently been identified in patients with IFAP syndrome, a disease that shows phenotypic overlap with KFSD. We postulate that both phenotypes are in the spectrum of one genetic disorder with a partially overlapping phenotype.

366

Mutations in the T (brachyury) gene cause a novel syndrome consisting of sacral agenesia, defective ossification of the vertebral bodies and persistent notochord. *A.V. Postma[1], M. Alders[2], M. Sylva[1], C.M. Bilardo[3], E. Pajkrt[3], M.M.A.M. Mannens[2], A.F.M. Moorman[1], R.J. Oostra[1], M.C. van Maarle[2].* 1) Heart Failure Research Center, Academic Medical Center, Amsterdam, Netherlands; 2) Department of Clinical Genetics, Academic Medical Center, Amsterdam, Netherlands; 3) Department of Obstetrics and Gynaecology, Academic Medical Center, Amsterdam, Netherlands.

The T gene (Brachyury gene) is the founding member of the T-box family of transcription factors. The protein encoded by this gene is an embryonic nuclear transcription factor that binds to specific DNA elements and effects transcription of genes required for mesoderm formation and differentiation. The protein is localized to notochord-derived cells and is vital for the formation and differentiation of the posterior mesoderm and axial development of all vertebrates. Recently, T gene duplications were linked to familial chordoma. We report here on three patients from two consanguineous families exhibiting sacral agenesia, delayed ossification of the vertebral bodies and a persistent notochord. The occurrence of three cases of a very rare disorder in two consanguineous families from the same ethnic background suggested homozygosity for a founder mutation. Therefore, we performed array CGH (negative) and homozygosity mapping and identified a 5.6Mb homozygous region on chromosome 6q27. This critical chromosomal region contains the T-gene. Sequencing of T in the affected individuals led to the identification of a shared homozygous mutation, p.H171R, present in the highly conserved T-box and absent from over 600 control alleles. This mutation causes a significantly decreased activation of promoters containing half or full T-sites. In line with this, qPCR analysis of several (target) genes, downstream of T, showed decreased activation (e.g. sox9 and fgfr3). Trafficking of the mutant protein to the nucleus is unaffected. In conclusion, we present evidence that a novel syndrome consisting of sacral agenesia, defective ossification of the vertebral bodies and persistent notochord is caused by a homozygous mutation in the T gene. The mutation leads to a moderate loss of function on DNA binding and disturbed expression of important downstream targets genes involved in proper posterior mesoderm and axial development, likely resulting in the unique phenotype.

367

Maximum likelihood estimation of recent ancestry (ERA) between pairs of individuals using high-density SNP-genotyping microarray data. *D.J. Witherspoon[1], C.D. Huff[1], Y. Zhang[1], W.S. Watkins[1], T.S. Simonson[1], T.M. Tuohy[2], D.W. Neklason[2], R.W. Burt[2], S.L. Guthery[3], S.R. Woodward[4], L.B. Jorde[1].* 1) Dept Human Genetics, Univ Utah, Salt Lake City, UT; 2) Huntsman Cancer Inst, Univ Utah, Salt Lake City, UT; 3) Dept Pediatrics, Univ of Utah School of Medicine, Salt Lake City, UT; 4) Sorenson Molecular Genealogy Foundation, Salt Lake City, UT.

Accurate and sensitive estimation of recent ancestry between individuals is important for ensuring correct pedigrees in genetic linkage analyses, for removing related individuals from analyses that require only unrelated individuals, and for identifying unknown relationships that can be used to construct, extend, and link pedigrees. The genetic evidence bearing on the relationship between two individuals comprises the number and sizes of chromosomal segments that they share identically-by-descent (IBD). We estimate these segments from high-density SNP-genotyping microarray data using GERMLINE 1.4.1 or Beagle 3.2. We then use maximum likelihood to identify the best model of recent ancestry for that pair, i.e, the number of generations back to the most recent common ancestor and the number of ancestors (one or two) shared at that point. Our likelihood calculation incorporates the expected distribution of shared segments between unrelated individuals, which we estimate from the shared segments observed in pairs among the 60 unrelated CEU HapMap samples. We test our method on 72 pairs of Utah individuals (24 full first, second and third cousin pairs) selected and genotyped using Affymetrix 6.0 SNP arrays. The precise relationship is retrieved for 65% of pairs, and 96% of our estimates are accurate to within one meiosis (the difference between second cousins and second cousins once removed.) We infer a single pair of third cousins to be genetically unrelated. This is consistent with the fact that a small proportion of third cousins will share no chromosomal segments that are IBD through their shared great-great-grandparents. The power of this method to detect relationships out to the third cousin level exceeds 95%. We also test this method on three pedigrees composed of seven individuals diagnosed with attenuated familial adenomatous polyposis resulting from the inheritance of a known mutation in the APC gene from common ancestors 3-5 generations ago. After adjusting the likelihood calculation for ascertainment on inheritance of a founder mutation, all estimates of recent ancestry for these individuals were accurate to within one meiosis. The methods described here are computationally efficient, make near-optimal use of the genetic signal of relatedness between individuals, and achieve a statistical power very close to the theoretical maximum. These methods are implemented in the freely-available software program ERA (estimation of recent ancestry.).

368

Fast detection of identity by descent in unrelated individuals. *B.L. Browning[1], S.R. Browning[2].* 1) Medical Genetics, University of Washington, Seattle, WA; 2) Biostatistics, University of Washington, Seattle, WA.

We recently developed a powerful method for detecting identity by descent (IBD) in unrelated individuals (Browning & Browning AJHG 2010) that accounts for linkage disequilibrium (LD) and has much lower false discovery rate than existing methods for dense marker sets. That method can be applied genome-wide to small numbers of individuals, but it is too computationally intensive to be applied to all pairs of individuals in large genome-wide association studies (GWAS). We have now developed a fast haplotype-cluster identity by state (HC-IBS) scoring algorithm that can be applied to large GWAS and that is highly predictive of IBD. This enables a two-step approach in which HC-IBS scoring identifies potentially IBD regions, which can then be followed up with exact IBD probability calculation using our existing method. Alternatively, one can use the HC-IBS scoring to test for population-based linkage in GWAS. We present results from applying our method to several data sets, including the WTCCC1 Bipolar, Type II Diabetes and Crohn's Disease studies, and the GENEVA Type II Diabetes study (Nurse's Health Study).

369

A Novel Haplotype-based Method Uncovers Association to Recent Variation. *A. Gusev[1], E.E. Kenny[1,2], J.K. Lowe[3,4], J. Salit[2], R. Saxena[4], S. Kathiresan[4,5], D.M. Altshuler[4,6,7], J.M. Friedman[2], J.L. Breslow[2], I. Pe'er[1].* 1) Computer Science, Columbia University, New York, NY; 2) Rockefeller University, New York, NY; 3) Department of Molecular Biology, Massachusetts General Hospital, Boston, MA; 4) Program in Medical and Population Genetics, The Broad Institute of Harvard and MIT, Cambridge, MA; 5) Cardiovascular Disease Prevention Center, Cardiology Division, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts; 6) Center for Human Genetic Research and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts; 7) Department of Genetics, Harvard Medical School, Boston, Massachusetts.

Rare variants affecting phenotype are a challenge for human genetics. While genomewide association studies have successfully detected multiple common causal variants, they are underpowered in identifying disease variants that are poorly represented on commercial SNP arrays, being too rare or population-specific. Recent multipoint and imputation-based methods boost the power of detecting and localizing the true causal variant, leveraging common haplotypes in a sequenced panel of reference samples. However, they are limited by the need to obtain a robust population-specific reference panel that is deep enough to observe a rare variant of interest. We set out to overcome these challenges and detect association to rare alleles using SNP array with long stretches of genomic sharing that are identical by descent (IBD). We have developed a novel algorithm and software tool, DASH, which builds upon pairwise IBD shared segments to infer clusters of individuals who are IBD at across a particular locus. Briefly, for each locus-window, DASH constructs a graph with nodes representing individuals and links based on pairwise IBD spanning that locus, and uses an iterative min-cut approach to identify densely connected components in this graph. As DASH slides the local window along the genome, links representing new shared segments are added and old ones expire; these changes cause the resultant connected components to grow and shrink, representing new haplotypes and their respective clusters of carriers. We code the corresponding haplotypes as genetic markers and use them for testing association. We have applied this method to simulated data and multiple GWAS datasets. In the classical WTCCC data we observe several novel associated loci (P<3e-9 each) to four diseases. In an isolated population from the island of Kosrae, Micronesia, where a dense reference panel is unavailable but the abundance of IBD due to founder effects enhances the method, we identified several variants associated beyond genome-wide significance that were not found through GWAS. In particular, an 800kb haplotype at the FXR1 region is associated with total cholesterol (P=3.90e-10, beta=0.41) and a 400kb haplotype strongly associated with levels of Hemoglobin A1c (P=4.3e-24, beta=0.71). We have replicated these associations in a large European cohort from the Diabetes Genetics Initiative and identified putative causal coding and structural changes in low-pass sequence from the haplotype carriers.

370

On the ranking of the disease susceptibility locus in family-based candidate gene studies: a simulation-based analysis. *C. Rakovski, L. Brown.* Mathematics and CS, Chapman University, Orange, CA.

The ranking of the p-value of the true causal single nucleotide polymorphism in the ordered list of individual SNP p-values is an important factor for achieving success in the ultimate objective of association studies - identifying deleterious genetic variants. Thus, we undertake a study to assess the implications of complex, multimarker correlation structure, sample size and disease models on the ranking of the causal SNP. We carry out an extensive family-based candidate gene simulation study to analyze the position of the disease susceptibility locus in the complete list of individual SNP p-values ordered according to their statistical significance. We simulate data based on the haplotype distributions of ten randomly selected genes extracted from the HapMap database, various sample sizes (600,1000 and 2000) that current association studies employ, and disease models that mimic the characteristics of complex human disorders. Based on appropriate linear model analysis, we conclude that the average ranking of the causal SNP for all studied sample sizes dramatically distant from the most significant and intuitively appropriate top position. This result is even more pronounced for genes with high average correlation and large number of common SNPs. Moreover, the gain of the DSL ranking when comparing sample sizes 600 to 1000 and 1000 to 2000, averaged over disease models, causal SNPs and genes, was approximately 1.3. These outcomes both reveal the importance of the sample size and quantify the magnitude required to unequivocally determine the identity of the DSL in family-based candidate gene studies. Our results show the overwhelming importance of large sample sizes in the localization of deleterious SNPs even under simple disease models. These conclusions possess pronounced importance for the design and result interpretation of candidate gene, next generation high-density genome-wide association studies, as well as for the construction and implementation of association tests based on the distribution of the most significant (minimum p-value) test statistics.

371

Power of equal intercepts differential effect regression models to infer gene-environment interactions in genome-wide association studies. *G. Guffanti[1], T. Bakken[2], S.G. Potkin[1], F. Macciardi[1,3], N. Schork[2]*. 1) Dept. of Psychiatry and Human Behavior, University of California Irvine, Irvine, CA; 2) The Scripps Translational Science Institute, Scripps Health and The Scripps Research Institute, La Jolla, CA; 3) Dept. of Sciences and Biomedical Technologies, University of Milan, Milan, Italy.

The identification of genetic variants that contribute differentially to quantitative phenotypes in different genetic or environmental backgrounds may improve our understanding of the biology of complex diseases. We have considered a simple differential effects regression model that assumes equal intercepts for detecting specific types of gene-environment interactions on quantitative traits in the context of genome-wide association studies. We describe the results of simulation studies for different settings that result in both accurate p-values and increased power in comparison with traditional general linear model that assumes interaction effects. We consider modeling the modifying influence of a particular factor (such as diagnosis, environmental exposure, gender, etc.) on the effect of genotype at a locus on a quantitative phenotype. We do this by limiting interest to interactions between the genetic factors and modifying factor in a regression model that assumes equal intercepts but different slopes. Standard F-tests can be constructed that test the equal slopes hypothesis. We explored the performance of the proposed test statistic in a wide variety of scenarios involving different effect sizes and sample size via simulation studies. We also considered theoretical details of the proposed model. For each hypothetical interaction setting, we simulated 1,000 data sets, each consisting of a causal SNP for 200 or 400 subjects, with differential diagnoses/environmental exposures across a wide range of effect size and allele frequencies from 0.05 up to 0.4. We then compared the performance of the proposed equal intercepts model with a traditional linear model. Both methods preserve a nominal type I error rate (0.05) assuming different phenotypic error distributions. Across a range of effect sizes, the proposed method is more powerful than the traditional method to detect interactions. For example, for moderate effect sizes, when the allelic frequency = 0.2, power is 26.7% using traditional statistics compared to 49.9% using our approach. As the allele frequency and sample size increase, both tests gain power, but the difference in power remains constant in favor of the proposed model. We show that for quantitative traits the equal intercepts differential effect regression model is substantially more powerful for detecting certain types of interactions than traditional approaches in a wide variety of settings.

372

Knowledge-driven multi-locus analysis reveals gene-gene interactions influencing HDL cholesterol level in two independent biobanks. *S.D. Turner[1], R.L. Berg[2], D.C. Crawford[1], J.C. Denny[3], J.G. Linneman[2], C.A. McCarty[2], P.L. Peissig[2], L.V. Rasmussen[2], D.M. Roden[4], R.A. Wilke[5], M.D. Ritchie[1]*. 1) Center for Human Genetics Research, Department of Molecular Physiology & Biophysics, Vanderbilt University Medical Center; 2) Biomedical Informatics Research Center, Marshfield Clinic Research Foundation; 3) Departments of Biomedical Informatics and Medicine, Vanderbilt University Medical Center; 4) Departments of Medicine and Pharmacology, Vanderbilt University Medical Center; 5) Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University.

Cardiovascular disease (CVD) is the #1 cause of morbidity and mortality in industrialized nations. High density lipoprotein (HDL) particles facilitate reverse cholesterol transport and may attenuate the development of CVD. HDL level is highly heritable (70%), yet much of this heritability is unexplained by even the most highly powered genome-wide association studies (GWAS). Some of this unexplained heritability may lie in largely unexplored gene-gene interactions. Here we present a knowledge-driven multi-locus analysis of plasma HDL cholesterol concentration as a quantitative trait using 3740 European Americans from a GWAS (using the Illumina 660-Quad platform) in the Marshfield Personalized Medicine Project (PMRP). By incorporating information from biological pathways, protein families, and other data sources into our analysis, we have identified several replicating gene-gene interaction models that significantly impact HDL level in the general population. Utilizing this information, we reduced the number of interaction tests from 136 billion to 22768, a 6 million fold reduction in the number of tests. Using 3740 individuals from the PMRP we identified 194 models with a significant interaction term p-value ($p<0.01$) and a significant overall model fit ($p<0.05$). Of these, 7 models replicated in 980 European Americans in the Vanderbilt Genome Electronic Records (VGER) project with a significant interaction term p-value ($p<0.05$) and a significant model fit p-value ($p<0.10$). Both the PMRP and VGER are part of the electronic Medical Records and GEnomics (eMERGE) network - a consortium formed to develop, disseminate, and apply approaches to research that combine DNA biorepositories with electronic medical record systems for large-scale, high-throughput genetic research. Several of these replicating models included gene-gene interaction models between ABCA1 and LPL - two genes known to be involved in HDL homeostasis. Other models included interactions between C7 and C8A (complement components), HK1 and HK2 (hexokinase genes), and NMNAT2 and NMNAT3 (nucleotidyltransferases). Searching for gene-gene interactions which influence a complex trait is challenging both statistically and computationally. The methodology and results presented here illustrate the utility of knowledge-based multi-locus analysis for the discovery of novel gene-gene interactions, and demonstrate several possible unexplored mechanisms involved in HDL cholesterol homeostasis.

373

Bayesian multivariate phenotype modeling for genome-wide association studies. *J. Marchini[1], V. Iotchkova[1], T. Spector[2], W. Ouwehand[3], N. Soranzo[3]*. 1) Dept Statistics, Oxford Univ, Oxford, United Kingdom; 2) Department of Twin Research & Genetic Epidemiology, King's College London, UK; 3) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.

The majority of genome-wide association studies have been carried out using a single binary or single quantitative trait as the phenotype of interest. For many traits several potential phenotypes may be available so it is natural to ask the question of how best to test for association in the presence of multiple phenotypes. We consider the case in which a number of continuous phenotypes are measured on each individual. In our approach we consider models in which the set of phenotypes are partitioned into an associated subset and an un-associated subset. We then average over all of the possible bi-partitions of the phenotypes to calculate a Bayes factor for each SNP or set of SNPs in a region or gene. If an association is uncovered the model fit of each bi-partition can be examined to infer which subset of phenotypes are driving the association. Additional properties of our method are that we allow for correlations between the residuals of both the associated and un-associated phenotypes and allow for mutiple cohorts to be analysed together using a mixed effects model. We have shown using simulated data that this approach leads to an increase in power to detect effects, over and above using single phenotype analysis, and is able to accurately uncover the true set of associated phenotypes. We are applying the approach to a genome-wide association of eight hematological parameters collected on the TwinsUK and UKBS Common Control collection.

374

The PCPD method: reducing dimensionality for diseases with rich clinical data. *G. Nelson[1], S. Hendrickson[2].* 1) BRP/LGD-NCI, SAIC Frederick, Frederick, MD; 2) LGD, NCI-Frederick, Frederick, MD.

Complex epidemiological phenomena such as response to HIV are measured by multiple clinical variables collected on study individuals. While this gives more information per subject for association analysis, the interpretation is difficult: since these variables in general are non-independent, when a genetic factor shows significant association with several disease outcomes is this confirmatory or does it simply reflect the correlation of these measures? In addition, this non-independence complicates the problem of correcting for multiple comparisons. Traditionally, biometric measurements on individuals are transformed to independent variables by the principal components method, but this approach is not directly applicable to many of the common epidemiological statistics, in particular because of the problem of censored data. We present a general method for decorrelating association statistics using a principal components transformation. To determine the intrinsic correlation of these measures for different outcomes, we generate a population of random associations by permutation of the genetic association in question, and for the set of association statistics calculated from these, compute the Principle Components over the Permuted Data (PCPD). Applying this principle component transformation to the actual association data yields a set of uncorrelated association measures. The independence of the transformed variables allows calculating a single measure of significance, replacing the multidimensional variation of the association data with a single measure of distance of displacement from the center of the cloud of random associations generated by permutation. We test the method by considering the associations of five clearly established genetic factors affecting AIDS, combining six related measures of AIDS progression, using published AIDS association data. To simulate the case of real associations that are marginally significant due to limited data, we calculate the untransformed and PCPD-transformed association variables on 40 randomly drawn subsets of 1400 individuals (seroconverters and seroprevalents) from six AIDS natural history cohorts. The PCPD combined p value is in all cases less than the smallest Bonferroni-transformed smallest p value for the six original statistics, and in four of five cases has smaller variance.

375

A Flexible Likelihood Framework for Dissecting Gene Pleiotropy Combining Non-randomly Ascertained Samples: Application to Sequence Data. *D.J. Liu[1,2], S.M. Leal[1,2].* 1) Dept Statistics, Rice Univ, Houston, TX; 2) Department of Molecular and Human Genetics, Baylor Collage of Medicine, Houston, TX.

There is solid evidence supporting the existence of gene pleiotropy. The study of pleiotropy by carrying-out phenome mapping using next generation sequence data will provide insight into underlying genetic pathways and lead to more powerful detection of causal gene loci. Due to the high cost of sequencing, most studies use carefully selected samples to improve power, e.g. collecting “super” controls or cases with family history or extreme trait values. Since many clinically relevant secondary traits are often measured besides the main phenotype, it is greatly beneficial to combine samples from different studies for phenome mapping. The use of selected samples poses great challenges for analyses. Even for case-control studies of the same main phenotype, sample ascertainment can be complex and different. The selection criterion can also involve multiple phenotypes or sub-phenotypes, e.g. study of type 2 diabetes in non-obese individuals. Due to phenotypic correlations, the analyses of gene pleiotropy can be seriously biased if sample ascertainment is not properly adjusted. Existing methods for analyzing secondary phenotypes are limited in their ability to jointly model phenotypic correlations and complicated ascertainment schemes. To address these problems, we propose a modified liability-threshold likelihood framework (MoLTeF) for mapping genes with pleiotropic effects using selectively ascertained samples. Statistically efficient inferences can be made using either prospective or retrospective likelihoods. MoLTeF is flexible for accommodating a broad spectrum of ascertainment schemes and integrates many existing gene based tests for common and rare variants. In addition, genetic parameters of interest, e.g locus heritability can be efficiently estimated. Extensive simulations under a rigorous population genetic model are carried-out with phenotypic parameters estimated from complex traits. It is shown that the power for mapping secondary phenotypes can be greatly improved when multiple cohorts are combined. In the presence of pleiotropy, if a selected sample is used where the gene locus is associated with both main and secondary phenotypes, the power will be further increased. The magnitude of the increase is jointly affected by the sizes and directions of the main gene locus effects and phenotypic residual correlations. In conclusion, MoLTeF will play an important role in dissecting human phenomes in sequencing based disease studies.

376

Disease prediction from GWA data - utilising prior information. *C. Hoggart, H. Eleftherohorinou.* Epidemiology & Pub Hlth, Imperial Col, London, United Kingdom.

Genome-wide association studies have identified many SNPs associated with common diseases, however, only a small proportion of the predicted genetic contribution has so far been identified. The HyperLasso algorithm (Hoggart et al, 2008) implements a shrinkage regression method capable of analysing up to a million SNPs simultaneously. In Hoggart et al (2008) we show that, when applied to sequence level data, the method greatly improves localisation in comparison to a single a SNP analysis. The HyperLasso has been successfully applied to the prediction of three auto-immune diseases using genome-wide data and in the process identify variants which do not reach genome-wide significance but do have out of sample predictive power, Eleftherohorinou et al (2009). The success of this method relies on the pre-selection of SNPs which lie in known immune response pathways which thus alters the prior probability of their causal effect on auto-immune diseases. This is highlighted by the relatively poor predictive performance of the algorithm when applied to genome-wide data without pre-selection. We extend the HyperLasso software to enable the Bayes Factor required for each SNP to enter the model to be controlled. Defining the required Bayes factor for a SNP as the prior odds against its association with the phenotype provides a coherent approach to weight SNPs genome-wide. We demonstrate the predictive performance of this method in several real genome-wide data sets utilising prior information from previous studies and the genomic annotation set in which the SNPs reside, Hindroff et al (2009).

377

Genetics of infectious disease susceptibility and morphological traits in a participant driven cohort. *D.A. Hinds, J.M. Macpherson, J. Tung, B. Naughton, A. Kiefer, C. Do, A. Wojcicki, J. Mountain, N. Eriksson.* 23andMe, Inc., Mountain View, CA.

We have developed a platform and cohort for genetics research combining self-reported phenotype information from web based surveys with genome-wide SNP genotype data. The platform enables us to pose new research questions and quickly collect responses to new surveys, with sample sizes for common traits that compare well with conventional genetic studies designed to query just a single phenotype. We have shown that the platform can successfully replicate many conventional GWAS findings, and can also generate novel association results.

Here we describe recent findings generated using this research platform, addressing the genetics of common infectious disease susceptibility as well as several common morphological traits. We report novel genome-wide significant or suggestive associations with several types of childhood and adult infections, as well as pigmentation traits and hair morphology. We find a strong association between reported mumps infection and rs485186, a synonmous variant in the FUT2 gene (P = 2e-12), as well as suggestive associations with cold sores and with shingles. We also find associations for hair color with rs1668619 near the EDNRB gene (P = 1e-8) and for hair curl with rs17558560, a non-synonymous variant in a keratin gene, KRT27 (P = 1e-9).

378

Expanding the psoriasis transcriptome: Altered expression of known and novel microRNAs in psoriatic skin. *C.E. Joyce[1], X. Zhou[2], J. Xia[2], C. Ryan[3], A. Menter[3], W. Zhang[2], A.M. Bowcock[1].* 1) Dept Gen, Washington Univ, Saint Louis, MO; 2) Dept Comp Sci, Washington Univ, Saint Louis, MO; 3) Dept Dermatology, Baylor Research Institute, Dallas, TX.

Psoriasis (PS) is a complex inflammatory skin disease where genetic risk factors and environmental triggers contribute to the “rewiring” of the circuitry of the transcriptome of the skin. PS affects as many as 7.5 million Americans (2-3 percent of the population) and between 10 and 30 percent of PS patients develop psoriatic arthritis. Complex interactions between immune- and keratinocyte-derived genes mediate disease pathogenesis. Microarray profiling by our lab and others has revealed extensive networks of differentially expressed genes in psoriatic skin involving both the epidermis and the immune system. Despite the extensive profiling of protein coding genes, little is known about the role of microRNAs (miRNAs) in PS. We used NexGen sequencing to generate small RNA libraries from normal, uninvolved psoriatic, and involved psoriatic skin (n=20, 20, 20). From the deep sequencing data we have identified 25 upregulated and 11 downregulated miRNAs in involved versus normal skin (FDR P<0.05), 17 upegulated and 5 downregulated miRNAs in involved versus uninvolved skin, and one upregulated and three downregulated miRNAs in uninvolved versus normal skin. We validated the differential expression of seven of these miRNAs with qRT-PCR (miR-21, 31, 101, 132, 369-3p, 378, and 1299). Furthermore, we have computationally predicted ~100 putative novel miRNAs, eight of which have been experimentally validated by qRT-PCR. Several of these novel miRNAs are also differentially expressed in psoriatic skin. Of particular note are two novel miRNAs that map to introns in the CD36 and ZC3HAV1 genes. Both of these protein coding genes are upregulated in psoriatic skin, and we have confirmed that their predicted intronic miRNAs are upregulated as well. This work represents the first NexGen sequencing profile of miRNAs in healthy and diseased human skin biopsies and has expanded our understanding of the psoriasis transcriptome.

379

Association of variants in the *TNFAIP3* region with systemic lupus erythematosus in a multi-ethnic study. *I. Adrianto[1], C.J. Lessard[1,2], K.M. Kaufman[1,2,3], J.M. Anaya[4], M.E. Alarcón-Riquelme[1,5], S.C. Bae[6], S.A. Boackle[7], E.E. Brown[8], L.A. Criswell[9], J.C. Edberg[8], B.I. Freedman[10], P.K. Gregersen[11], G.S. Gilkeson[12], C.O. Jacob[13], J.A. James[1,2], R. Kimberly[8], J. Martin[14], J.T. Merrill[1,2], T.B. Niewold[15], B.A. Pons-Estel[16], R.H. Scofield[1,2], A.M. Stevens[17], B.P. Tsao[18], T.J. Vyse[19], C.D. Langefeld[20], M.B. Humphrey[2], J.B. Harley[2,3], K.L. Moser[1,2], C. Gray-McGuire[1,21], P.M. Gaffney[1,21].* 1) Oklahoma Medical Research Foundation, Oklahoma City, OK; 2) The University of Oklahoma Health Sciences Center, Oklahoma City, OK; 3) Oklahoma City VA Medical Center; 4) Universidad del Rosario, Colombia; 5) Center of Genomics and Oncological Research (GENyO), Granada, Spain; on behalf of the BIOLUPUS and GENLES Networks; 6) Hanyang University, Seoul, Republic of Korea; 7) University of Colorado Denver; 8) University of Alabama at Birmingham; 9) Rosalind Russell Medical Research Center for Arthritis, University of California, San Francisco; 10) Wake Forest University Baptist Medical Center; 11) The Feinstein Institute for Medical Research, North Shore Long Island Jewish Health System; 12) Medical University of South Carolina; 13) University of Southern California Keck School of Medicine; 14) Instituto de Parasitologia y Biomedicina López-Neyra, CSIC, Granada, Spain; 15) University of Chicago; 16) Sanatorio Parque, Rosario, Argentina; 17) Seattle Children’s Hospital; 18) University of California, Los Angeles; 19) Imperial College London; 20) Wake Forest University Health Sciences; 21) Co-senior authors.

Systemic lupus erythematosus (SLE) is a complex autoimmune disease with varied and potentially severe clinical manifestations affecting multiple organs. Prevalence of this disease varies between genders and among age-groups and ethnicities. SLE affects women nine times greater than men and is more common in non-Caucasians than in Caucasians. Recent studies indicate that genetic variants in the region of tumor necrosis factor alpha-induced protein 3 (*TNFAIP3*) are associated with SLE in subjects of European and Asian ancestry. *TNFAIP3* encodes a zinc-finger protein called A20, a critical regulator of inflammatory signaling pathways. To further characterize and localize the effect of *TNFAIP3*, we genotyped and imputed single-nucleotide polymorphisms (SNPs) within and flanking *TNFAIP3* in multiple populations: European, African-American, Asian including Korean, Hispanic enriched for Amerindian-European admixture, and Gullah populations. Using a custom designed SNP panel for the Illumina iSelect system, we genotyped 127 SNPs in and around *TNFAIP3* and 343 ancestry-informative markers (AIMs) in a total of 8922 SLE unrelated cases and 8077 controls. Then, using HapMap Phase III and 1000 Genomes Project data we imputed a minimum of 274 additional SNPs for each of the populations (the number varied based on linkage disequilibrium structure). We assessed single marker association to SLE using logistic regression with sex, global and local ancestry adjustments. Association analysis identified risk haplotypes in Europeans and Asians likely to harbor a causal variant. Further haplotype comparisons across populations revealed a variant that likely explains the majority of the genetic association between SLE and *TNFAIP3* identified in our study populations with $P=1.96 \times 10^{-8}$, OR=1.69, 95% CI=1.41-2.04 in Europeans, $P=3.30 \times 10^{-9}$, OR= 2.26, 95% CI=1.73-2.97 in Asians, and $P=1.24 \times 10^{-9}$, OR=2.53, 95% CI=1.88-3.41 in Koreans alone. No significant association was found in other populations. These results support genetic association with SLE in the region of *TNFAIP3*, unique to Europeans and Asians and further demonstrate the complexity of identifying associations across different populations. Sequencing and functional studies are necessary to validate this variant and determine the contribution of *TNFAIP3* to SLE.

380

The Importance of the "Interferome" in SLE: Six New Susceptibility Genes for Lupus. *D.S. Cunninghame Graham[1], D.L. Morris[1], V. Gateva[2], T.W. Behrens[2], R.R. Graham[2], T.J. Vyse[1].* 1) Dept Rheumatology, Imperial Col, Hammersmith, London, United Kingdom; 2) Immunology Biomarkers Group, Genentech Inc, San Francisco, United States.

Background: The "*interferome*" is a complex network of genes regulated by type I, II and II interferons (IFNs). These genes are critical immune regulators which may play a role in autoimmune pathogenesis. Up-regulation of the IFN signalling pathway is correlated with increased disease severity for SLE. However, the genetic basis of SLE was largely intractable until the publication of several moderate-sized genome-wide (GW) association studies - which identified the loci with the strongest effect (OR>1.3) (~8%); of the genetic contribution to SLE). However, the published studies lack sufficient power to discover susceptibility genes of moderate risk (OR 1.1-1.3). **Aim:** This replication study, in an independent cohort of 926 UK SLE cases and 5392 out-of-study WTCCC2 controls, sought to identify novel susceptibility genes for lupus. **Materials and methods:** Using published data from a US-Swedish GWAS/replication cohort comprising 3273 SLE cases and 12188 controls, we tested 30 candidate SNPs which had shown a moderate risk ($5x10^{-3}<P>5x10^{-8}$), using a custom Illumina chip, on the BeadXpress platform. Case-control association analyses were conducted using PLINK. Total *P* values were calculated using Fisher's test by combining the UK results with the published data. **Results:** Our combined analysis identified six new lupus susceptibility genes within the interferome - each of which crosses a consensus threshold for GW significance ($P>5x10^{-8}$). *IRF8* is a transcription factor regulating IFN-responsive genes, including *NCF2* itself, which is part of the respiratory burst in neutrophils. Both *IFIH1*, an innate immune system receptor for dsDNA and *IL12RB2*, a subunit of the interleukin 12 receptor complex are upregulated by type I IFNs. *TAOK3* is a member of the TAO kinase family - which inhibits p38MAPK, a protein essential for IFN-dependent transcriptional regulation. *TYK2* is a tyrosine kinase which phosphorylates the receptor subunits of cytokine receptors, including IL-12 and type I IFNs. **Conclusion:** These six novel risk loci for SLE provide more evidence of a role for the "*interferome*" in lupus pathogenesis. Detailed understanding of the functional consequences for each gene requires the identification of causal allele(s). To address this, we are undertaking a trans-ethnic fine-mapping study.

381

Undiscovered common SNPs explain most of the missing heritability of the complex disease rheumatoid arthritis. *E.A. Stahl[1,2], R. Chen[1,2], B. Hickey[1,2], F.A.S. Kurreeman[1,2], S. Raychaudhuri[1-3], P.I.W. de Bakker[2,4], P.K. Gregersen[5], L. Klareskog[6], K.A. Siminovitch[7], J. Worthington[8], R.M. Plenge[1,2].* 1) Division of Rheumatology Immunology and Allergy, Brigham and Women's Hospital, Boston, MA; 2) Medical and Population Genetics, Broad Institute, Cambridge, MA; 3) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 4) Division of Genetics, Brigham and Women's Hospital, Boston, MA; 5) The Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, NY; 6) Rheumatology Unit, Department of Medicine, Karolinska Institutet at Karolinska University Hospital Solna, Stockholm, Sweden; 7) Dept of Medicine, University of Toronto, Mount Sinai Hospital and University Health Network, Toronto, Ontario, Canada; 8) Arthritis Research Campaign (arc)-Epidemiology Unit, Stopford Building, The University of Manchester, Manchester, United Kingdom.

Purpose: In the complex autoimmune disease rheumatoid arthritis (RA), genome-wide association studies of tens of thousands of individuals have discovered >30 risk loci; however, these risk loci together explain less than 20% of disease variance (whereas >50% is thought to be genetic). We hypothesized that, after removing all known RA risk loci, "polygenic risk scores" would capture the combined effects of many more common SNPs weakly associated with RA risk.

Methods: Based on GWAS in a "discovery" set of 4014 seropositive RA cases and 12,046 controls of European ancestry, we developed polygenic risk scores of additive, log-odds weighted risk allele counts at independent SNPs achieving different threshold P_{GWAS} values. We calculated and tested polygenic risk scores in an independent validation set of 1521 seropositive RA cases and 10,557 controls of European ancestry. To interpret these results, we simulated the polygenic risk score analysis under a polygenic model with disease-associated and null independent marker SNPs, and used acceptance/regression approximate Bayesian computation to sample from the posterior distribution of model parameters given our observed allele frequency-stratified polygenic risk score results.

Results: Polygenic risk scores for four discovery set P_{GWAS} thresholds, P_{GWAS} < 0.001, 0.01, 0.05 and 0.1, incorporating 505, 3704, 13,850 and 23,339 SNPs, respectively, showed highly significant associations with RA disease status in our independent validation dataset: logistic regression *P* = 7e-4, 9e-6, 3e-10, and 2e-9, respectively. Bayesian inference on the results observed for RA lead us to estimate that an additional 21% of RA disease variance (95% CI 11-29%) is explained by over 4000 independent risk alleles (95% CI 2076-7305) that currently remain undiscovered. These SNPs are mostly common in the general population (MAF > 0.34).

Conclusions: These findings have implications for the genetic architecture of complex disease, suggesting that 80% of the heritability of rheumatoid arthritis is explained by common tag SNPs, and will inform future disease gene discovery efforts as well as risk prediction especially utilizing whole-genome data.

382

Meta-analysis of GWAS in CD and RA reveal new shared autoimmune loci. *A. Zhernakova[1,2], E.A. Stahl[3,4], G. Trynka[5], S. Raychaudhuri[4,6], E.A.M. Festen[4], F.A.S. Kurreeman[3,1], L. Franke[4,7], R.S.N. Fehrmann[4], P. Dubois[7], P. Gregersen[8], J. Worthington[9], K. Siminovitch[10], L. Klareskog[11], P. Saavalainen[12], D. Barisani[13], M.T. Bardella[14,15], R. McManus[16], A.W. Ryan[16], G. Turner[7], E.F. Remmers[17], L. Greco[18], E. Grandone[19], B. Mora[20], A. Rybak[21], B. Cukrowska[22], Y. Li[23], T. Huizinga[1], D.A. van Heel[7], C. Wijmenga[5], R. Plenge[3].* 1) Dept Rheumatology, UMC Leiden, Leiden, PO Box 9600, Netherlands; 2) Complex Genetics Section, Dept of Medical Genetics, University Medical Centre Utrecht, PO Box 85060, 3508 AB, the Netherlands; 3) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Boston, Massachusetts, 02115, USA; 4) Broad Institute, Cambridge, Massachusetts, 02142 USA; 5) Genetics Dept, University Medical Centre Groningen and University of Groningen, PO Box 30.001, 9700 RB Groningen, the Netherlands; 6) Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, 02114, USA; 7) Blizard Institute of Cell and Molecular Science, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London E1 2AT, UK; 8) The Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, New York 11030, USA; 9) Arthritis Research Campaign (arc)-Epidemiology Unit, Stopford Building, The University of Manchester, Manchester M13 9PT, United Kingdom; 10) Dept of Medicine, University of Toronto, Mount Sinai Hospital and University Health Network, Toronto, Ontario M5G 1X5, Canada; 11) Rheumatology Unit, Department of Medicine, Karolinska Institutet at Karolinska University Hospital Solna, Stockholm 171 76, Sweden; 12) Department of Medical Genetics & Research Program for Molecular Medicine, Biomedicum Helsinki, University of Helsinki, Helsinki, Finland; 13) Department of Experimental Medicine, Faculty of Medicine University of Milano-Bicocca, Monza, Italy; 14) Fondazione IRCCS Ospedale Maggiore Policlinico, Mangiagalli e Regina Elena, Milan, Italy; 15) Department of Medical Sciences, University of Milan, Italy; 16) Department of Clinical Medicine, Trinity Centre for Health Sciences, Trinity College Dublin, St James's Hospital, Dublin, Ireland; 17) Genetics and Genomics Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, US National Institutes of Health, Bethesda, Maryland 20892, USA; 18) European Laboratory for Food Induced Disease, University of Naples Federico II, Naples, Italy; 19) Unita' di Aterosclerosi e Trombosi, I.R.C.C.S Casa Sollievo della Sofferenza , S. Giovanni Rotondo, Foggia, Italy; 20) Department of Experimental Medicine, Sapienza University of Rome, Rome, Italy; 21) Department of Gastroenterology, Hepatology and Immunology, Children's Memorial Health Institute, Warsaw, Poland; 22) Department of Pathology, Children's Memorial Health Institute, Warsaw, Poland; 23) Celera, Alameda, California 94502, USA.

Purpose: Recent genome-wide association (GWA) studies in autoimmune diseases have discovered dozens of susceptibility loci, some of which show a shared association to two or more immune-related diseases. Family data indicate shared genetic factors for two autoimmune diseases, rheumatoid arthritis (RA) and celiac disease (CD). No study has systematically compared known risk loci and GWAS data for RA and CD. We hypothesized that combining GWAS data from the two diseases could discover new shared RA-CD risk loci. Methods: We performed an unbiased search for shared CD-RA genes by combining the GWAS datasets of CD (4533 cases, 10750 controls) and RA (5,539 autoantibody positive RA cases and 20,169 controls) in a meta-analysis. The non-directional inverse variance method was used to pick up association with opposite allelic effects. We attempted replication of SNPs with P<10-5 in an additional 2169 CD cases and 2255 controls, and 2845 autoantibody positive RA cases and 4944 controls. We conducted an association study in the replication and combined cohorts with inverse variance meta-analysis, where we analyzed RA-only samples, CD-only samples, and RA+CD samples. We investigated the functional relevance of the associated loci by analyzing their effect on gene expression in an expression quantitative trait meta-analysis of 1,469 whole blood samples. Results: Outside the HLA locus, we observed association to both diseases with 102 SNPs in 34 loci at P<1x10-5. From 34 loci, 9 were already established in both CD and RA. Twenty-one of 25 SNPs could be designed and/or passed quality control filters in our replication samples. One SNP (rs10892279) was consistently associated with risk in both CD and RA (P<5x10-4 in both replication cohorts), and achieved P=5.8x10-9 in all RA samples, and 4.4x10-13 in combined CD_RA analysis. Another 6 loci showed suggestive association to both diseases with P<5x10-8 in a combined analysis of all CD and RA samples, and p<0.001 in each disease. From 16 CD-RA loci (9 previously known and 7 identified in this study), in 8 (50%) the risk variant was correlated with cis gene expression. Conclusions: Our results substantially improve our understanding of shared genetic basis of CD and RA. Meta-analysis of GWAS datasets in related diseases allows identification of shared genetic variants with moderate effect size. The co-morbidity of CD and RA in families and individuals can be explained by sharing of genetic risk factors.

383

Deep resequencing of GWAS loci identifies rare variants associated with Crohn's Disease. *M.A. Rivas[1,2], C. Stevens[1], C. Zhang[3], Y. Sharma[3], J. Kang[3], P. Schumm[4], A. Taylor[1], N. Burtt[1], A. Kirby[1,2], S. Gabriel[1], D. Altshuler[1,2,5], R. Xavier[1,2], J. Rioux[6], J. Cho[3], M.J. Daly[1,2,5], NIDDK IBD Genetics Consortium.* 1) Medical & Population Genetics, Broad Institute, Cambridge, MA, USA; 2) 2CHGR, Massachusetts General Hospital, Boston, MA, USA; 3) Department of Genetics, Yale University, New Haven, Connecticut, USA; 4) Department of Health Studies, University of Chicago, Chicago, Illinois, USA; 5) Harvard Medical School, Boston, MA,USA; 6) Université de Montréal and the Montreal Heart Institute, Research Center, Montréal, Québec, Canada.

Genome-wide association studies of complex diseases have identified over a hundred disease-susceptibility loci. This is a result of systematic, well-powered, genome-wide surveys exploring the relationships between common sequence variation and disease predisposition. The next challenge in complex trait genetics is in understanding the role that rare variation has in phenotypic outcome. In this study, we sequence coding exons of 56 genes in regions associated to Crohns Disease. Follow up genotyping of rare variants discovered in sequencing of case patients and controls demonstrates that NOD2 rare variants Met863Val (p-value = 0.001596, 21 copies in cases 0 copies in controls) and Asn852Ser (p-value = 0.03756, 22 copies in cases 4 copies in controls) are associated with Crohn's Disease. Whereas rare variants in genes PRDM1 (Ser220Asn) , JAK2 (Leu393Val) , IRF8 (Gly191Gly) , C1orf106 (Tyr333Phe) demonstrate suggestive association and replicated in multiple independent cohorts. Our results demonstrate that rare variants play a role in susceptibility to Crohns disease and is the first exhaustive screening of rare variants in loci with common variants associated to a complex disease.

384

A Genetically Defined Mendelian Presentation of Asthma, Food Allergy, Eosinophilic Esophagitis, and Inflammatory Bowel Disease. *A.L. Guerrerio[1], P.A. Guerrerio[1], L. Myers[1], M.K. Halushka[1], G. Dhillon[1], R. Anders[1], M. Oliva-Hemker[1], R.A. Wood[1], H.C. Dietz[1,2].* 1) Johns Hopkins University School of Medicine, Baltimore, MD; 2) HHMI.

Disorders of immunologic tolerance are common, with frequent manifestations in the respiratory and gastrointestinal system. Conditions such as asthma, food allergy, eosinophilic esophagitis (EE) and inflammatory bowel disease (IBD) present as complex traits, frustrating efforts to elucidate pathogenesis and develop mechanism-targeted therapies. TGFβ is a regulatory cytokine with essential roles in T cell development and immune tolerance. Loeys-Dietz syndrome (LDS) is a recently described autosomal dominant aortic aneurysm syndrome caused by mutations in either of the two genes encoding subunits of the TGFβ receptor (TGFBR1 or TGFBR2). Food allergy was reported in 26/58 (45%; ~10X population prevalence (PP)) of LDS patients with known TGFBR mutations. 19/42 (45%) were sensitized to at least one of seven common food allergens and 27/42 (64%) to at least one of seven aeroallergens. Asthma was diagnosed in 26/58 (45%; ~6X PP), allergic rhinitis in 28/58 (48%), and eczema in 22/58 (38%). 19/58 (33%; ~500X PP) reported gastrointestinal (GI) complaints consistent with EE and 6/10 undergoing GI biopsy had overt histologic evidence of EE. LDS patients had elevated total IgE levels (p<0.02), peripheral eosinophil counts (p<0.02), and Th2 cytokines (IL-5, IL-13) and MCP-1 in their serum. LDS patients with food allergy had elevated serum levels of IL-1β, GM-CSF, and a trend toward higher IL-17 compared to nonallergic controls. A significant subset of these patients demonstrated eosinophilic colitis and progressed to develop IBD (5%; ~1500X PP) that was confirmed by histologic changes on biopsies. LDS patients showed an increase in CD4+ cells that express low levels of Foxp3+ and lack regulatory activity (p<0.02) and increased naïve resting T regulatory (Treg) cells (p<0.03). Thymic biopsies obtained at the time of aortic surgery showed markedly elevated TGFβ signaling compared to controls, as evidenced by nuclear accumulation of phosphorylated Smad2. The monogenic nature of LDS allows conclusion that altered TGFβ signaling is sufficient to predispose to asthma, food allergy, EE, and IBD in people, and that this pathway is therefore an attractive therapeutic target. These data will inform efforts to elucidate pathogenesis in nonsyndromic presentations including interpretation and prioritization of the output of genome-wide association studies. The therapeutic utility of TGFβ antagonists is being tested in LDS mice that recapitulate these common human phenotypes.

385

A multi-center meta-analysis of pediatric asthma. *P. Sleiman[1], K. Bonnelykke[2], S.F.A Grant[1], C. Kim[1], R. Chiavacci[1], E.A Nohr[3], D. Strachan[4], E. Widen[5], L. Palmer[6], A. Kustovic[7], J. Heinrich[8], V. Jaddoe[9], H. Bisgaard[2], H. Hakonarson[1].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) COPSAC, University of Copenhagen; 3) Department of Epidemiology, Institute of Public Health, University of Aarhus; 4) St George's, University of London; 5) Dept Public Health, University of Helsinki; 6) Centre for Medical Research, University of Western Australia; 7) Respiratory Research Group, University of Manchester; 8) Institute of Epidemiology, Helmholtz Zentrum München; 9) Dept of Epidemiology, Erasmus MC, University Medical Center Rotterdam.

Asthma is the most common chronic condition in children resulting in approximately 500,000 hospitalizations in the US each year; yet the previously reported predisposition loci only explain a fraction of the heritability of the disease. We have assembled an international consortium, the Eagle consortium, that includes 10 European and North American centers with a total of over 11,000 samples to further investigate the genetic basis of pediatric asthma through a meta-analysis of all the cohorts. To satisfy the inclusion criteria for the meta-analysis, cases and controls had to be of Caucasian ancestry the cases were defined as having doctor diagnosed asthma before the age of 7. Each contributing center performed a case control genome-wide analysis following imputation of their dataset to the 2.5M SNPs available in HapMap. The summary stats were subsequently meta-analyzed using the software package, metal. In addition to the previously reported predisposition loci, including DENND1B and ORMDL3, we have identified an additional locus on chromosome 2p22.3 that surpassed genome-wide significance. The associated SNPs map to an LD block containing the RASGRP3 gene which has been previously associated with susceptibility to systemic lupus erythematosus in Han Chinese. A further 6 loci showed consistent association with asthma in the meta-analysis but did not surpass the statistical threshold for genome wide significance. We are currently pursuing replication of all associated loci in two independent cohorts of asthmatics of European ancestry. The results of the meta-analysis and replications will be presented in full.

386

Sequencing study of candidate genes under weak purifying selection implicates rare variants in asthma susceptibility. *D. Torgerson[1,7], D. Capurso[1], K. Barnes[2], D. Meyers[3], B. Raby[4], R. Mathias[2], P. Graves[5], R. Strunk[6], T. Beaty[2], E. Bleecker[3], S. Weiss[4], F. Martinez[5], D. Nicolae[1], C. Ober[1].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Division of Asthma and Clinical Immunology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 3) Genomics Center, Wake Forest University, Winston-Salem, NC; 4) Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 5) Arizona Respiratory Center and Bios Institute, University of Arizona, Tucson, AZ; 6) Washington University, St. Louis, MA; 7) Current Address: Department of Medicine, University of California San Francisco, San Francisco, CA.

Despite the success of many genome-wide association studies (GWAS), only a small fraction of the heritability of common diseases are accounted for by the risk alleles identified, potentially due to the limited inclusion of rare variants on current genotyping platforms. To investigate the contributions of rare variants to asthma susceptibility, we sequenced the coding exons of 9 genes previously implicated in asthma susceptibility and showed signatures of weak purifying selection. By sequencing these genes in 513 asthma cases and 515 non-asthmatic controls, including both European and African Americans, we identified 1,253 variants, 74% of which were novel and predominantly rare (minor allele frequency <0.05 in either cases or controls). In the European Americans, individual rare variants in 5 genes were significantly associated with asthma at $p<4x10^{-5}$ (*AGT*, *CFTR*, *DPP10*, *IKBKAP*, and *IL12RB1*), many of which were intronic variants in the *CFTR* gene (presented elsewhere). For example, a novel intronic variant in *CFTR* and a synonymous variant in *IL12RB1* were significantly enriched in the cases (6.3% vs. 0%, $p=5.9x10^{-9}$ and 6.2% vs. 0%, $p=4.3x10^{-9}$, respectively). No individual variants were significantly associated with asthma in the African Americans following Bonferroni correction, even when considering patterns of local European admixture. However, there were significantly more carriers of rare nonsynonymous mutations in the *DPP10* gene, and significantly more carriers of rare mutations that were modestly associated with asthma ($p<0.05$) in 5 additional genes (*AGT*, *CFTR*, *CHIA*, *IL12RB1*, and *TGFB1*) in African American cases compared to controls ($p<0.05$). Overall, resequencing candidate genes showing evidence of weak purifying selection implicated rare variants in asthma susceptibility in 7 of the 9 genes studied. Our results suggest that multiple rare variants may increase the risk of asthma susceptibility in both European and African Americans, and that rare variation may explain some of the missing heritability of asthma. Notably, we find that individual rare variants were predominantly implicated in asthma susceptibility among European Americans, whereas multiple rare variants in the same genes were predominantly implicated in asthma susceptibility among African Americans. Lastly, most of the rare variants identified by sequencing are not captured by SNPs on genotyping platforms, and would not be (and have not been) identified in a GWAS.

387

Patients' Attitudes and Preferences About Data Sharing, Privacy, and Participating in Genetic Research. *M.E. Smith[1], E.J. Gordon[2], K.A. Cameron[3], M.A. Graves[1], A.L. Rosenfeld[1], W.A. Wolf[1].* 1) Ctr Gen Med, Northwestern Univ, Chicago, IL; 2) Institute for Healthcare Studies, Dept of Surgery, Northwestern Univ, Chicago, IL; 3) Institute for Healthcare Studies, Northwestern Univ, Chicago, IL.

Federal policy requires that phenotypic and genetic data from NIH-funded genome-wide association studies (GWAS) be deposited into dbGaP, a government controlled, limited access data repository that facilitates data sharing. Novel ethical questions have begun to arise about this new data sharing policy. Little information exists regarding patients' attitudes and preferences about their research data being deposited into dbGaP and if it affects their willingness to participate in large-scale genetic research. We conducted semi-structured interviews with 55 outpatients from a Chicago medical center from January-April 2010. Interviews used open- and closed-ended and Likert questions to assess attitudes about data sharing, privacy, governance, and willingness to participate in genetic research. Qualitative data were analyzed using ethnographic content analysis, with high inter-rater reliability (0.81-0.92). After learning about the NIH data sharing policy, the vast majority (98%) of patients felt it very or extremely important for investigators to share research findings with other investigators. However, the majority (71%) stated that it would be very or extremely important that they not be re-identified if their information were kept in the NIH database. Most participants (97%) understood the need for data sharing but desired more information about the research process and governance of the NIH database. Most (73%) reported that data sharing would not affect their willingness to participate in genetic research. At the outset of the interview, 69% were very or extremely likely to participate in genetic research; when asked again at the end of the interview, preferences remained stable (69%). Although most (69%) respondents were confident that NIH would protect data in the database, approximately one third (30%) doubted that the NIH could effectively maintain privacy. Qualitative data revealed that most participants had little understanding of what genetic research involves and expressed doubts about researchers' appropriate management of shared dbGaP data and about the NIH's governance and security of dbGaP. Thus, although there is significant support for the idea of data sharing among research participants, our findings suggest an urgent need to educate the public and potential participants about genetic research, the role of NIH in research, and the processes and oversight of data sharing for GWAS.

388

Returning individual research results from large cohort genetic research: what do people really want? *J. Murphy Bollinger, D. Kaufman, J. Scott.* Gen & Pub Policy Ctr, Johns Hospkins Univ, Washington, DC.

Objective: Large, prospective cohort studies are important research tools. Whether and to what extent individual research results (IRRs) should be returned to participants has been debated recently in the research and bioethical literature. To more fully explore previous findings about the public's strong desire for IRRs when considering participating in such studies, we conducted 10 focus groups with 89 people across the U.S. **Methods**: Between October 2009 and January 2010, 8 focus groups were conducted in-person in DC, Philadelphia, and Denver, and 2 groups online. Participants represented a range of demographics including race, gender, age, ethnicity, socioeconomic status, and social networking behavior. After describing a potential U.S. based cohort study, participants were asked about receiving IRRs depending on (1) the availability of a treatment; (2) the magnitude of risk; (3) whether the risk would change over time as more was learned; (4) whether the significance of the finding was known; (5) whether the phenotype was a disease or trait; and (6) whether the risk factor was environmental. Attitudes about the challenges of returning IRRs, trade-offs in study design, and willingness to pay for IRRs were explored. Focus group transcripts were analyzed using the NVIVO 8.0 software. **Results**: While some felt IRR should not be returned and would participate without receiving IRRs, majorities in all groups expressed a strong desire for their IRRs. IRRs were viewed as compensation for participating, as an obligation of the study, or as an essential element of study design. Generally, most wanted IRRs regardless of the availability of treatment or the level of risk. Desire for IRRs varied with the medical relevance, the severity of the phenotype, and the certainty of the information being returned. To obtain IRRs, participants would consider reducing the size and scope of the study, limiting the IRRs returned, or paying for some results. **Conclusion**: For many, the desire for IRRs was strong. While some wanted all IRRs, as participants considered different types of results, limitations of the information, and the challenges of returning IRRs, some changed their opinions about the type of results important to them. While IRRs with clinical utility were highly valued, many examples of personal utility were identified.

389

Citizens' values regarding the use of stored newborn screening samples: A public engagement study. *Y. Bombard[1], F.A. Miller[1], R.Z. Hayeems[1], J.C. Carroll[2], D. Avard[3], J. Allanson[4], R. Axler[1], J. Bytautas[1], P. Chakraborty[4], Y. Giguere[5], J. Little[6], B.J. Wilson[6].* 1) Department of Health Policy, Management and Evaluation, University of Toronto, Canada; 2) Department of Family and Community Medicine, Mount Sinai Hospital, University of Toronto, Canada; 3) Centre for Genomics and Policy, Department of Human Genetics, McGill University, Montreal, Canada; 4) Department of Genetics, Children's Hospital of Eastern Ontario, and Department of Pediatrics, University of Ottawa, Canada; 5) Department of Medical Biology, Centre Hospitalier Universitaire de Quebec (CHUQ), University of Laval, Canada; 6) Department of Epidemiology and Community Medicine, University of Ottawa, Canada.

Background: Newborn screening (NBS) programs may store samples for quality control, confirmatory diagnosis, research and forensic investigations. Recent public controversies and lawsuits over the scope of storage and role of parental consent underscore the need to engage the public on these issues. Objectives: We explored public values regarding storage of NBS samples for various purposes and the role of parental consent for anonymous research conducted with NBS samples. Methods: Using a staged approach, 8 focus groups were conducted in Toronto and Montreal (n=60); participants were recruited through community advertisements. A Citizens' Panel was subsequently held with a subset of Toronto participants (n=16) and was asked to make group recommendations about aspects of storage and consent. Both approaches included: an educational component, deliberative discussion and pre- and post-questionnaires assessing knowledge and values of NBS and related issues. Data were analyzed with descriptive statistics and qualitative content analysis. Results: Compared with baseline, participants' knowledge about NBS and storage significantly improved across focus groups and the Citizens' Panel ($p<0.0001$), supporting a positive impact of the educational intervention and discussion. The majority (>90%) supported storing NBS samples for quality control, diagnosis and anonymized research, but participants were ambivalent about their use for forensic investigations (54% support) or unspecified purposes (50%). Most preferred that parents be strongly encouraged to have their baby's sample stored (77%) and be able to choose about storage without pressure (77%). During discussions, some participants felt comfortable routinizing storage, evoking a sense of trust in the system; others worried that samples might be used inappropriately, and thus favoured an active consent process. All members of the Citizens' Panel endorsed the importance of informing parents about NBS storage prenatally; 14 favoured routinizing storage of anonymized NBS samples for research (1 abstained, 1 left early). Conclusion: Our staged approach to public engagement helped to elicit values and develop group recommendations. Citizens value research and believe they need to be informed about storage. While some endorse a need for explicit consent, a majority is willing to permit a more routinized process in light of feasibility constraints.

390

Direct from consumers: A survey of 1,048 customers of three direct-to-consumer personal genomic testing companies about motivations, attitudes, and responses to testing. *D. Kaufman, J. Murphy Bollinger, S. Devaney, J. Scott.* Genetics & Public Policy Center, Johns Hopkins University, Washington, DC.

Purpose: Several companies sell personal genomic tests directly to US consumers (DTC). Speculation about the limitations, harms, and benefits of these DTC tests abound. To collect empirical data about customers' motivations and experiences with personal genomic DTC tests, an online survey of customers of 23andMe, deCODEme, and Navigenics was conducted. **Methods:** Random samples of US customers who received DTC results between June 2009 and March 2010 were invited to participate. Survey topic included reasons for testing, attitudes about test results, health behaviors, and demographics. Respondents were asked to interpret two genetic test results shown using the DTC companies' graphics and text. **Results:** Between January and May of 2010, 3,167 DTC customers were invited, 1,163 (37%) responded, and 1,048 participated. Most (87%) were White non-Hispanics; 24% had a BA degree and 54% had post-graduate education; 67% had incomes over $85,000. The top 3 reasons for using DTC genetic testing were satisfying curiosity (94% very or somewhat important), learning about elevated risks of diseases (91%), and learning about ancestry (90%, among those who got ancestry and health data); 77% said testing was important to improve their health. 1 in 3 were interested because a 1° relative was affected with a tested condition and 42% were interested in one or more specific conditions. A total of 58% learned something new to improve their health, while 9% felt they could not change their health risks. Only 29% had shared results with a healthcare provider or had a follow-up test; 18% intended to do so. As a result of testing, 34% said they were being more careful about their diet, 14% were exercising more, and 15% changed medication or supplement regimens. Although 88% agreed their risk report was easy to understand, 38% said the conclusions were too vague. Between 4% and 7% misinterpreted examples of company's risk results, either answering incorrectly or saying they did not know. While 66% felt DTC personal genomic tests should be available without government oversight, 84% and 73% said it was important for an NGO like Consumer Reports, or the FTC, respectively, to monitor companies' claims for scientific accuracy. **Conclusion:** Early adopters of personal genomic DTC tests, who tended to be highly educated, indicated satisfaction with their DTC services. Long term follow-up of DTC test users is needed to evaluate the impact of DTC testing on healthcare usage and behavior.

391

Implementation of Direct-to-Patient (DTP) testing for Fabry disease in a large population of adults with unexplained left-ventricular hypertrophy. *C. Wells[1], K. Goss[2], K. Sims[2], M. Browning[2], M. Taylor[1].* 1) University of Colorado Denver, Aurora, CO; 2) Harvard Medical School, Boston, MA.

Background: Technological advantages have been capitalized upon by various commercial companies promoting direct-to-consumer genetic testing. The genetics community has voiced concerns related to the utility, appropriateness, and ethical basis for these practices. A lack of oversight and involvement by geneticists and genetic counselors has been cited as a significant deficiency in how results are reported and 'explained' to consumers. **Methods:** We implemented a direct-to-patient (DTP) study targeting adult patients with unexplained left-ventricular hypertrophy (LVH), hypothesizing that a modest subset of LVH was due to Fabry disease. We adopted a 'simple' study design that involved: 1) screening an LVH database, 2) contacting LVH patients, 3) enrolling them by mail, including obtaining both a dried-blood spot (DBS) and saliva sample, 4) testing DBS alpha-galactosidase activity, and 5) reporting results. **Results:** 2000 patients who had LVH in our echocardiogram database received letters inviting them to participate in a DBS study for Fabry: 20% agreed to participate, 24% indicated no interest in the study, 23% of letters were returned with no current address, 33% of letters were never responded to (even after multiple mailings). Of the 20% who initially agreed to participate, 31% successfully returned the consent forms and appropriate samples and 4.1% of these were diagnosed with Fabry disease. Multiple questions and misunderstandings were documented from DTP subjects despite clear instructions written at a 3rd grade level with "stand alone" photographic instructions, in addition to a direct phone line to a study coordinator/genetic counselor. **Conclusions:** Interest, uptake, and understanding of this DTP genetic testing protocol were highly variable. Skilled personnel (geneticists, genetic counselors, laboratory experts) are invaluable to tackle the complexities of a 'genetic testing by mail' approach. These data show that even with specific contact from clinicians to individuals with overt phenotypes (LVH) who may benefit from an accurate diagnosis and treatment, **multiple hidden counfounders persist.** The lessons learned from this project should guide the development of additional DTP projects and should temper enthusiasm for efforts to offer testing, commercial or otherwise, where involvement by genetics professionals is limited or absent.

392

"It's not like judgment day": Public understanding of and reactions to personalized genomic risk information. *B.A Bernhardt[1], S. Gollust[2], G. Griffin[1], R.E. Pyeritz[1], L. Wawak[3], C. Zayac[1], E. Gordon[3].* 1) Dept Med Gen, Univ Pennsylvania, Philadelphia, PA; 2) University of Pennsylvania, Philadelphia, PA; 3) Coriell Institute, Camden, NJ.

Little is known about how people interpret and react to personalized genomic risk information for common complex conditions. To address this gap, we interviewed people who received personalized risk results for 7 disorders through the Coriell Personalized Medicine Collaborative (CPMC), a research study examining the clinical utility of personalized genomic risk information. The interview included questions about understanding of results and actions taken based on results. Of the 42 interview transcripts analyzed to date, we found that most people appeared to have a good understanding of their results, although some had difficulty interpreting relative risk values. Participants tended to interpret their risks based on genetic results combined with their own family and medical history. Although some participants were surprised to learn about an increased genetic risk for a disorder that was not in their family, no participant reported being overly concerned about their risks. About 1/4 of the participants acted on results by changing behavior, but most reported only slightly modifying healthy behaviors they were already engaged in (exercising, using sunscreen, lowering dietary fat, etc.), and only a few participants initiated a new risk-reduction behavior (quitting smoking, beginning an exercise program). Those who reported no behavior change either felt they were already doing everything they could to reduce their risk, or did not feel their level of risk was high enough to warrant changes. About 1/3 of participants had shared results with a physician and most others planned to share results in the future. Results were shared based on a belief that the doctor would be interested and would provide recommendations to reduce risk. Almost half the participants who shared results believe their doctor had a good understanding of the results, but about 1/4 indicated that their doctor did not understand the results, or did not know what to do with them. Recommendations made by doctors (blood test for lipid levels, start vitamins, lose weight, etc.) follow standard population recommendations. We conclude that many of the concerns about providing personalized genomic risk information (poor understanding of results, undue worry, false reassurance, unnecessary test-ordering) are not supported by our findings. There is a need, however, for comparative effectiveness research to determine the value added of personalized genomic risk results.

393

Reactions of Smokers to an Information Pamphlet About Genetic Testing for a Common Lung Cancer-Associated Gene Variant. *S.C. Sanderson[1], J. Shepperd[2], C.M. McBride[3], S. Docherty[4], S.C. O'Neill[5], I.M. Lipkus[4].* 1) Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY; 2) Department of Psychology, University of Florida, Gainsville, FL; 3) Social and Behavioral Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 4) School of Nursing, Duke University, Durham, NC; 5) Cancer Control Program, Georgetown University, Washington, DC.

Most common gene variants are only weakly associated with complex diseases such as lung cancer. Little is known about how smokers respond to learning about genetic testing for such variants. Our aim was to assess smokers' reactions to an information pamphlet describing genetic testing for a common lung cancer-associated gene variant (GSTM1-null) . We assessed whether smokers (a) understood that the gene variant slightly influenced lung cancer risk, (b) felt the information to be useful in decision-making about genetic testing for lung cancer risk, and (c) were interested in being tested for the variant. Participants were 131 students who smoked cigarettes. They read and than answered questions about a 20-page information pamphlet, "Genetic Testing for Lung Cancer Risk", which included information that people missing the GSTM1 gene (GSTM1-null) may have a 20% higher lung cancer risk than those who are GSTM1-present, and that lifetime risks are estimated to be 11% and 9% respectively. Most (78%) interpreted the GSTM1-present result to indicate that lung cancer risk is "slightly lower than average", whilst 18% interpreted it as meaning "average" risk, and 2% as "much lower than average". Similarly, most (89%) interpreted the GSTM1-null result as "slightly higher than average"; 3% interpreted it as "average" risk, 6% as "much higher than average". When asked whether the information would help people make a decision about genetic testing for lung cancer risk, 68% said "yes", 29% "somewhat", 3% "no". When asked to rate how understandable the information pamphlet was overall, the mean score was 6.60+0.71 (where 1=not at all to 7=completely understandable). The mean scores for importance and interest in getting tested for GSTM1 were 5.79+1.17 (where 1=unimportant to 7=important) and 5.21+1.66 (where 1=not at all to 7=extremely interested). The results suggest that smokers may perceive genetic information as important even when they understand a gene variant is only slightly related to disease risk, and express interest in genetic testing using a marker not highly predictive of disease. Further research is needed on how best to convey information to the public about small effects on risk conferred by common genetic variation, and to explore why people are interested in receiving personal genetic information about genetic variants of low penetrance.

394

Assessing the Accuracy of Curation-Based Risk Prediction using Self-Reported Phenotypes. *C.B. Do, M. Macpherson, N. Eriksson, J. Tung, B. Naughton, A. Hsu, A. Kiefer, M. Miyazawa, M. Polcari, J. Mountain, D. Hinds, A. Wojcicki.* 23andMe, Inc, Mountain View, CA.

Over the last several years, a number of direct-to-consumer (DTC) genetic testing companies have begun to offer personal genotyping services on the premise that single nucleotide polymorphisms (SNPs) can be used to predict complex disease risk. Typically, these methods work by combining estimates of individual SNP odds ratios and disease incidence data curated from the literature. While a number of groups have proposed techniques for theoretically estimating the predictive accuracy possible with collections of SNPs, to date, little is known about the empirical accuracy of these methods in practice. In this work, we perform empirical benchmarking of the predictive accuracy of a curation-based risk prediction algorithm for a diverse collection of 22 self-reported medical phenotypes using a population cohort of over 30,000 unrelated European individuals.

We assess each disease in terms of discrimination (i.e., the ability of the algorithm to correctly order individuals by increasing disease risk) and calibration (i.e., the ability of the algorithm to assign accurate probabilities of developing a disease within each risk group). We show that SNPs are mildly predictive for a wide variety of conditions, ranging from common cancers (e.g., breast and prostate) to autoimmune disorders (e.g., psoriasis and Crohn's disease), with typical area-under-curve (AUC) statistics around 0.60. In a comparison of our empirical accuracies with theoretical estimates derived from known SNP effect sizes and frequency estimates in Europeans, we find that the latter consistently exhibit upward bias, resulting from a combination of factors including case/control set sizes, rigor in phenotype definition, self-report phenotyping error, and the winner's curse. These data thus provide one of the first empirical assessments of curation-based risk prediction in a DTC setting, and demonstrate the value of self-reported phenotype data as a resource for performance evaluation.

395

Family history education to improve genetic risk assessment for cancer. *M. Scheuner[1,2], N. Smith[1], E. Schalles[1].* 1) VA Greater Los Angeles, Los Angeles, CA; 2) David Geffen School of Medicine at UCLA.

Introduction: We have developed and implemented, and are now evaluating an educational program designed to improve familial risk assessment for hereditary breast-ovarian cancer and Lynch syndrome, with the goal to increase family history documentation and appropriate referrals for genetic consultation. Methods: Our educational program is targeted to primary care clinicians staffing the Women's Clinics at the VA Greater Los Angeles Health Care System. We are using a pre/post design that will assess changes in clinician knowledge, attitudes and behaviors relating to familial risk assessment, genetic testing and management of hereditary cancer, and cancer family history documentation and referral for genetics consultation and testing. Our educational interventions are informational (Website, CME-approved lecture series), clinical (patient information sheets, family history tools in the electronic health record (EHR)) and behavioral (practice-feedback reports regarding use of EHR tools). We are using multiple modes of both qualitative and quantitative evaluation methods. Results: Our educational program was implemented in March 2010. At baseline, 8 primary care clinicians completed our knowledge and attitudes survey. The mean percent correct score was 55% (range, 26-77%). The highest knowledge scores related to topics of basic genetics concepts, terminology, and ethical issues. Genetic testing and recognition of hereditary cancer syndromes received the lowest scores. Topics rated most relevant to primary care practice included management of hereditary cancer and ethical issues. Topics with least relevance were genetic testing and basic genetics concepts and terminology. Lecture evaluations show high scores regarding usefulness and ability to understand content. Clinicians are using the EHR tools (260 family histories completed in first 2 months), and as a result, high-risk patients are being referred for consultation. Discussion: Topics relating to cancer family history risk assessment and management of high-risk patients are relevant to primary care clinicians at the VA in Greater Los Angeles, yet there is a substantial need for education. In a short period, our educational interventions have improved cancer family history documentation and genetics referral. We will continue our evaluation over the next 12 months. If successful, our educational program may serve as a model program for risk assessment of hereditary cancer and other conditions.

396

Nationwide implementation and adoption of pharmacogenetic testing through a novel pharmacy-based approach. *E. Stanek, R. Verbrugge, J. Barlow, R. Aubert, M. Khalid, F. Frueh, R. Epstein.* Medco Health Services, Inc., and Medco Research Institute, Franklin Lakes, NJ.

In a late-2008 survey of over 10,000 US physicians, we found that while 13% had already engaged in pharmacogenetic testing, 67% hadn't yet done so because of a need for more information about genomic markers and testing. While the potential for broader adoption of pharmacogenetic testing exists, little is known about the effectiveness of systems and processes designed to do so. We have utilized the unique position and capabilities of a nationwide pharmacy benefit manager to expedite the adoption of pharmacogenetic testing in routine clinical practice. We devised a pharmacy-based process, with implementation support from a national laboratory partner, for identifying patients who may be appropriate for pharmacogenetic testing, securing written physician and patient consent, obtaining DNA samples through home-based self-collection, and communicating test results with interpretation to the physician. Physician and patient consent for testing was assessed in two ongoing clinical pharmacogenetic testing programs initiated in mid-2008 (CYP2C9/VKORC1 genotyping for warfarin; CYP2D6 genotyping for tamoxifen). Physicians deemed testing as potentially appropriate in 40,259 patients, including 32,192 warfarin patients and 8,067 tamoxifen patients. Among these opportunities, physicians provided consent for pharmacogenetic testing in 20,164 (50.1%) patients overall, and in 15,827 (49.2%) and 4,337 (53.8%) of warfarin and tamoxifen patients, respectively. Out of these cases, testing was also potentially appropriate from the patient perspective in 16,003 patients overall, and in 12,225 and 3,778 of warfarin and tamoxifen patients, respectively. Patients residing in all 50 states provided their consent in 6,850 (42.8%) instances —4,723 (38.6%) in the warfarin program and 2,127 (56.3%) in the tamoxifen program. Benchmark analysis of claims data in 2008 indicate that only 1.7% of warfarin patients and 2.3% of tamoxifen patients undergo any form of genetic testing in routine practice. These pharmacy-based pharmacogenetic testing programs provide convenient processes, access, and information to both physicians and patients and result in high rates of consent for testing. This demonstrates the capability of these programs to dramatically increase the adoption of pharmacogenetic testing by physicians and patients in the routine clinical practice setting on a nationwide scale.

397

Homologous recombination between nonhomologous chromosomes - Recurrent chromosomal translocations mediated by interchromosomal NAHR. *P. Stankiewicz[1], Z. Ou[2], Z. Xia[1], A.M. Breman[1], B. Hogan[1], J. Wiszniewska[1], P. Szafranski[1], M.L. Cooper[1], M. Rao[1], L. Shao[1], S.T. South[2], K. Coleman[3], P.M. Fernhoff[3], M.J. Deray[4], S. Rosengren[5], E. Roeder[6], V.B. Enciso[6], A.C. Chinault[1], A. Patel[1], S.-H.I. Kang[1], C. Shaw[1], J.R. Lupski[1], S.W. Cheung[1,7,8].* 1) Dept Molec & Human Gen, Baylor Col Med, Houston, TX; 2) Departments of Pediatrics and Pathology, University of Utah, Salt Lake City, UT; 3) Children's Healthcare of Atlanta, Atlanta, GA; 4) Department of Neurology, Miami Children's Hospital, Miami, FL; 5) Hartford Hospital, Hartford, CT; 6) Department of Pediatrics, UTHSCSA, San Antonio,; 7) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 8) Texas Children's Hospital, Houston, TX. *Authors 1 and 2 contributed equally.

Four unrelated families with the same unbalanced translocation der(4)t(4;11)(p16.2;p15.4) were identified. Both of the breakpoint regions in 4p16.2 and 11p15.4 were narrowed to large ~359-kb and ~215-kb low-copy repeat (LCR) clusters, respectively, by aCGH and SNP array analyses. DNA sequencing enabled mapping the breakpoints of one translocation to 24-bp within interchromosomal paralogous LCRs of ~130-kb in length and 94.7% DNA sequence identity located in olfactory receptor gene clusters, indicating nonallelic homologous recombination (NAHR) as the mechanism for translocation. To investigate the potential involvement of interchromosomal LCRs in recurrent chromosomal translocation formation, we performed computational genome-wide analyses and identified 5292 interchromosomal LCR substrate pairs, >5-kb in size and sharing >94% sequence identity that can potentially mediate chromosomal translocations. Additional proof for interchromosomal NAHR mediated translocation formation was provided by sequencing the breakpoints of another recurrent translocation, der(8)t(8;12)(p23.1;p13.31). The NAHR sites were mapped within 55-bp in ~7.8-kb paralogous subunits of 95.3% sequence identity located in the ~579-kb (chr8) and ~287-kb (chr12) LCR clusters. We demonstrate that NAHR mediates recurrent constitutional translocations throughout the human genome and provide a computationally determined genome-wide "recurrent translocation map".

398

GEN1 resolves cruciform-forming palindromic DNA leading to a recurrent translocation in humans. *H. Inagaki[1], T. Ohye[1], H. Kogo[1], M. Tsutsumi[1], T. Kato[1,2], M. Tong[1], B.S. Emanuel[2], H. Kurahashi[1].* 1) Div Molecular Genetics, ICMS, Fujita Health Univ, Toyoake, Aichi, Japan; 2) Division of Human Genetics, The Children's Hospital of Philadelphia, and Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA.

The t(11;22)(q23;q11) is the most frequent recurrent constitutional translocation in humans. The breakpoints on both chromosomes have been located within characteristic sequences, palindromic AT-rich repeats (PATRRs), which can form secondary structure by intrastrand annealing. Previously, we established a plasmid-based model system for this PATRR-mediated translocation using a human cell line as a host. In this system, we demonstrated that cruciform formation at the PATRR induced double-strand breaks and led to a translocation-like rearrangement. In the present work, we explored the enzyme responsible for cleavage of the cruciform structure of the PATRR using this system. We have found that GEN1, a recently identified Holliday junction (HJ) resolvase, participates in generation of PATRR-mediated translocations. When GEN1 was knocked down by siRNA, the amount of translocation-like rearrangement products was greatly decreased. Southern analyses also indicated that cleavage of the PATRR was suppressed by GEN1 knockdown. Indeed, recombinant GEN1 was found to cleave the cruciform of the PATRR in *in vivo* and *in vitro* experiments. On the other hand, these features were not observed for other HJ resolvases, such as MUS81 or SLX1. Combined with our previous finding that Artemis is also required for translocation-like rearrangement in this system, the PATRR-mediated translocation is induced by two structure-specific enzymes, GEN1 and Artemis. It is suggested that the PATRR is first processed by diagonal cleavage at the cruciform joint by GEN1, and subsequently by cleavage of the hairpin tips by Artemis at the ends of the broken DNA. Finally, the opened ends appear to be illegitimately fused by the non-homologous end joining pathway to form this unique chromosomal rearrangement in humans.

399

Telomere disruption leads to non-random formation of dicentric chromosomes involving acrocentric human chromosomes. *K.M. Stimpson[1,4], A. Jauch[2], J.M. Bridger[3], K.E. Hayden[1], B.A. Sullivan[1,4].* 1) Institute for Genome Sciences & Policy, Duke University, Durham, NC; 2) Institute of Human Genetics, University Hospital Heidelberg, D-69120 Heidelberg Germany; 3) Laboratory of Nuclear and Genomic Health, Centre for Cell and Chromosome Biology, Division of Biosciences, Brunel University, Uxbridge, UK; 4) Department of Molecular Genetics and Microbiology, Duke University, Durham, NC.

Dicentric chromosomes are abnormal chromosome rearrangements that physically link two centromeres on the same chromosome. Historically, dicentric chromosomes have been considered unstable and in many organisms, undergo breakage-fusion-bridge cycles. Dicentrics in humans can be remarkably stable due to functional silencing, or inactivation, of one centromere. However, the molecular basis of centromere inactivation is unclear. We developed a human cell culture system to create *de novo* dicentric chromosomes and monitor their mitotic behavior after formation. Using controllable, short-term expression of a mutant version of telomere protein TRF2 (dnTRF2), we induced genome-wide instability in human cells and enriched for chromosome fusions. We found that ~80% of the induced fusions occurred between the short arms of acrocentric chromosomes. In other words, our chromosome fusion assay produced the most common, naturally-occurring human chromosome rearrangement, the Robertsonian translocation (ROB). Long-term dicentric stability was then continuously monitored for nearly 150 cell divisions. After 20 divisions, >94% of induced dicentrics were still functionally dicentric. However, over a period of 18 additional weeks, the population of induced dicentrics that were functionally monocentric increased to 40%. This result suggested that centromere inactivation occurred in almost half the dicentric population between 40 and 150 divisions after formation. Small chromosome fragments, many containing both CENP-A and acrocentric α-satellite DNA, were observed during this time. The number of fragments peaked at 6-8 weeks after dicentric formation. We hypothesized that these fragments might have originated from centromeres that had undergone inactivation. To test this, we used quantitative FISH to measure and compare α-satellite FISH signal intensities on induced ROBs and free-lying chromosomes. Centromeric signal intensity was significantly reduced at the inactive 21 centromere in two independent induced rob(14;21), suggesting that the array had decreased in size after dicentric formation. Collectively, our results support a role for telomere function in acrocentric chromosome stability and show that dicentric human chromosomes are mitotically stable after formation for many cell divisions. Finally, we provide evidence for one model of centromere inactivation that involves partial deletion of α-satellite DNA that is assembled into unique centromeric nucleosomes.

400

Evaluation and Elucidation of Balanced Chromosome Rearrangements: An Understanding of Formation Utilizing a Whole Genome SNP Array. *S. Schwartz[1], R.D. Burnside[1], I. Gadi[1], V. Jaswaney[1], E. Keitges[2], R. Pasion[1], V.R. Potluri[3], H. Risheg[2], J. Smith[3], J. Tepperberg[1], B. Williford[1], P. Papenhausen[1].* 1) Lab Corp of America, RTP, NC; 2) Dynacare/LabCorp, Seattle, WA; 3) Dynagene/LabCorp, Houston, TX.

Approximately 7% of patients with de novo balanced rearrangements have an associated abnormal phenotype due to either a loss of genetic material at the rearrangement breakpoints or the disruption of gene(s) in the formation of the rearrangement. Our initial FISH studies on a limited population were able to confirm these hypotheses. To better understand underlying chromosome mechanisms, we have expanded our study to include high resolution whole genome SNP array analysis of 110 rearrangements (100 de novo and 10 familial) ascertained in individuals with phenotypic abnormalities. The vast majority of the rearrangements were translocations (~80%); however both inversions and insertions were also studied. These results show that 56% of the de novo rearrangements (56/100) also had a loss or gain of chromosomal material and a total number of 117 copy number changes were identified. The abnormalities demonstrated much greater complexity than expected with only 29% demonstrating a single deletion in addition to the rearrangement. These results provide important insights into the formation of these rearrangements and associated copy number changes including: (1) the frequency of a copy number change (56%) in the de novo rearrangements, detected by a whole genome SNP array is much greater than by FISH and initially expected; (2) the frequency of the complexity is greater than expected with 10% of the samples having deletions at two breakpoints and 61% having more than two chromosomes involved and one deletion; (3) in only 57% of the cases were the deletions or duplications adjacent to the breakpoints; the others involved changes on another part of the chromosome or a different chromosome; (4) 80% of the abnormalities were deletions; the other 20% duplications; (5) in 96% of the patients with gain or loss of material, the changes could be shown to be pathogenic; (6) FISH studies of the patients without copy number changes demonstrated gene breakpoint truncation in 94% of the patients; (7) comparison of the gene involvement in these cases with both our database of 3500 abnormalities detected by array analysis and the publically available CNV databases suggest that the genes involved are responsible for the phenotypic abnormalities. These studies clearly show the importance of array studies and demonstrate how a combination of array, FISH and standard cytogenetics has provided considerable knowledge of genetic imbalance and the complexity of rearrangement formation.

401

A two-hit model for phenotypic variability in multiple genomic disorders. *S. Girirajan[1], J.A. Rosenfeld[2], G.M. Cooper[1], B.P. Coe[1], L.G. Shaffer[2], E.E. Eichler[1,3].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Signature Genomics Laboratories, Spokane, WA; 3) Howard Hughes Medical Institute, University of Washington, Seattle, WA.

We recently proposed a two-hit model to explain the phenotypic variability associated with a 520-kbp microdeletion on chromosome 16p12.1, wherein, the microdeletion both predisposes to neuropsychiatric phenotypes as a single event and exacerbates neurodevelopmental phenotypes in association with other large (>500 kbp) copy number variants (CNVs). We now extend our model to include 34 genomic disorders and examine CNV data from 22,877 cases with intellectual disability and congenital malformation for the presence of two large CNVs. Of the cases with a known genomic disorder, 132/1386 (9.5%) cases carried a large CNV second hit and for 27/132 (20.4%) of these two-hit carriers, the second CNV was also associated with a genomic disorder. While the frequency of second hits was higher in CNVs associated with variable expressivity such as del15q13.3, del16p11.2, dup16p13.11, del16p12.1, and del and dup1q21.1, we found an inverse correlation (regression analysis, r=-0.51, p=0.0018) between the proportion of de novo cases and the prevalence of the second hit. We find that the two hits more frequently originated de novo or from different parents (52/61) compared to transmission of both events (9/61) from the same parent (p=1.8 × 10^{-8}). Pathway analysis of genes within the second hit CNVs shows disruption of genes involved in cellular signaling, neurological, and developmental functions. Our data provide strong support for the two-hit model to explain variable expressivity in genomic disorders and has potential implications for laboratory diagnosis and counseling of cases with neurodevelopmental phenotypes.

402

Acquired copy number analyses of matched tumor and germline DNA in multiple myeloma. *S. Kulkarni[1,3,4], V. Hucthagowder[1], M. Evenson[1], C. Mullins[2], M. Fiala[2], K. Stockerl-Goldstein[2,3], J. Dipersio[2,3], M. Tomasson[2,3], R. Vij[2,3].* 1) Cytogenomics and Molecular Pathology, Washington University School of Medicine, St Louis, MO; 2) Division of Oncology, Department of Internal Medicine, Washington University School of Medicine, St Louis, MO; 3) Genetics and Genomic Medicine, Department of Pediatrics, Washington University School of Medicine, St Louis, MO; 4) Siteman Cancer Center at Barnes-Jewish Hospital and Washington University School of Medicine.

Multiple Myeloma (MM) is the second most common hematologic malignancy in the United States accounting for approximately 10% of all hematologic neoplasms. Despite recent advances in our understanding of MM biology, virtually all patients eventually relapse and die from progressive disease. Chromosomal aberrations and genomic copy number alterations are detected in nearly all multiple myeloma (MM) patients. The molecular events involved in multiple myeloma development have been studied using classical cytogenetics, fluorescence in situ hybridization (FISH), comparative genomic hybridization, array comparative genomic hybridization (aCGH) and SNP arrays. Published approaches to identify global genomic changes in MM have relied solely on the high-throughput genomic scans on plasma cells selected using CD138 magnetic separations. However, this approach of studying tumor sample alone is inherently limited due to inability to carefully distinguish germ line non-tumor genomic alterations with tumor specific acquired gene changes. We performed genome-wide analyses using to detect copy number alterations (CNA) and loss of heterozygosity (LOH) in CD138+ malignant multiple myeloma cells and matched germline (skin) specimens from 23 newly diagnosed patients. The frequency of occurrence (expressed as a percentage) was determined for each of the aberrations. We identified 305 acquired changes (178 Gains; 121 Loss and 6 LOH regions). To identify the true “driver” pathogenic alterations in MM, that provides tumor with growth and survival advantage from the “passenger” CNAs, we performed integrated network analyses. All the MM patients analyzed revealed deletions and amplifications. To investigate if somatic alterations may play a role in the disparity in incidence and/or severity of MM in African Americans and males, we compared frequency of specific CNAs seen exclusively in these groups. Genomic analyses of large numbers of MM genomes will be required to develop a full understanding of the biological consequences of acquired somatic CNAs. These findings demonstrate the power, accessibility and adaptability of molecular karyotyping in cancer to identify novel acquired recurrent genetic alterations that contribute to MM pathogenesis. To our knowledge, this is the first ultra high-density chromosomal microarray study in matched tumor and germline samples to characterize global genomic alterations/aberrations in gene networks in multiple myeloma.

403

Expert Pathology and Cytogenetic Review Are Both Required for Accurate HER2 FISH Analysis and Interpretation. *A.E. Wiktor, W.R. Sukov, R.B. Jenkins, R.P. Ketterling.* Division of Laboratory Genetics, Mayo Clinic, Rochester, MN.

In 2007, a joint ASCO/CAP committee published guidelines for HER2 testing. The CAP Cytogenetics checklist requires both pathologist and cytogeneticist involvement when performing and interpreting HER2 FISH analysis. Pathology review of the H&E slide is necessary to identify areas of invasive tumor. Cytogeneticists interpret the underlying genetic mechanisms, more critical when equivocal results or complex signal patterns are observed. To better understand the complexity of HER2 testing, we instituted an IRB approved review of HER2 FISH analyses performed in the Mayo Clinic Cytogenetics Laboratory from July 2007 through April 2010. A total of 14,452 samples were submitted for HER2 FISH testing, including 13,188 (91%) breast tumors. While assessment of HER2 gene status is only FDA-approved for invasive breast cancer, physicians request testing for metastatic disease and other tumor types that may respond to Herceptin therapy. Metastatic tumors primarily to lymph node, liver and bone accounted for 986 (6.9%) samples, and 184 (1.3%) were other primary tumor types. Pathologist review identified 9443 (71.6%) breast samples with >90% invasive tumor present. More critical review was needed for 3163 (24%) cases with 50-90% invasive tumor present, 334 (2.5%) with 11-40%, and 248 (1.9%) with £10%. The latter cases require interaction between on-site pathologists and geneticists for accurate identification of the invasive tumor component for appropriate HER2 FISH scoring. Additionally, 40 samples had only intraductal carcinoma, 44 had insufficient invasive tumor and at least 10 were misdiagnosed as breast cancer and represented benign breast disease or other tumor subtypes. Of breast cancer samples, 358 (2.7%) had equivocal FISH results and 876 (6.6%) posed interpretive cytogenetic challenges. These 876 included 343 with two distinct tumor populations (heterogeneous), 176 with ratios >2.2 due to duplication of HER2 (not true amplification), and 357 with amplification of the CEP 17 control probe requiring reflex testing for accurate interpretation. Cases with challenging tumor architecture or complex FISH results comprise approximately 1/3 of our HER2 FISH test volume and underscore the complexity of the assay. Both a pathologist's detailed assessment of the H&E slide, and a cytogeneticist's interpretation of FISH signal patterns and underlying genetic mechanism, are critical in accurately assessing HER2 gene status as a predictive marker for Herceptin therapy.

404

Features of genomic architecture close to *PTEN* may facilitate acquisition of somatic genomic rearrangements and deletions in prostate cancer. *J. Squire[1], O. Ludkovski[2], A. Evans[2], K. Sircar[3], T. Bismar[4], P. Nuin[1,5], M. Yoshimoto[1].* 1) Dept Pathology & Molecular Med, Queen's University, Kingston, Canada; 2) University Health Network, Pathology, Toronto, Canada; 3) UT MD Anderson Cancer Center, Pathology, Houston, TX, USA; 4) University of Calgary, Pathology & Laboratory Medicine and Oncology, Calgary, Canada; 5) Ontario Cancer Biomarker Network, Toronto, Canada.

Genomic interstitial microdeletions of several hundred kb and focal rearrangements close to the *PTEN* locus at 10q23 are frequent and strongly associated with both initiation and progression in prostate cancer (CaP). However, the genomic mechanism leading to *PTEN* losses in this region of chromosome 10 are poorly understood. We propose that genomic architecture and clusters of sequence identity in this region of the chromosome may form recombinational hotspots that facilitate genomic rearrangements and loss of *PTEN* as part of CaP tumorigenesis. To investigate this hypothesis, the present study examines the role of segmental duplications (SD), copy number variations (CNV) and fragile sites in the vicinity of *PTEN* and flanking loci to determine whether significant regions of microhomologies could facilitate non-allelic homologous recombinational repair errors and initiate deletion events. We mapped the breakpoint regions associated with *PTEN* deletions in CaP by FISH (n=330) and available online SNP databases (n=117). FISH analyses showed that 41% of tumors had an interstitial genomic deletion of *PTEN* with a maximum size of 2Mb. Mapping the size of copy number transitions based on SNP array data indicated that interstitial deletions always centered on PTEN and typically encompassed ~700kb within this cytoband. Within this deleted region there were a rare fragile site (FRA10A), and at least three clusters of non-redundant regions of microhomology neighboring the deleted region by paired intrachromosomal SDs with sequence homology >95% and alignment >10kb in length. Thus, the most remarkable findings of this study were the location of high-density regions of SDs and CNVs flanking the genomic region frequently involved in CaP-associated breaks. Phylogenetic analysis of these SDs clusters when compared to other species indicates these genomic regions are evolutionary conserved. Thus the model being developed suggests that somatic defects of a homology-dependent repair pathways in undifferentiated prostate cancer progenitors utilizes tracts of microhomology to trigger the initial genomic rearrangements that lead to *PTEN* genomic losses in tumor initiating cells. Collectively these data highlight the role of sequence microhomology clusters surrounding *PTEN* locus on genomic stability in CaP in the predisposition to deletion events that lead to the decreased gene dosage of pten, essential for CaP initiation and clinical progression.

405

Interphase FISH and chromosome analyses on ETV6/RUNX1-positive childhood acute lymphoblastic leukemia reveal an increased risk in a patient with additional copies of both normal and derivative chromosomes 21. *Z. LOU, C. Tirado, W. Chen, R. Garcia, S. Iyer, J. Dollitle, P. Koduru.* Department of Pathology, The University of Texas Southwestern Medical Cente, Dallas, TX.

The t(12;21)(p13;q22) [ETV6/RUNX1] in pediatric acute lymphoblastic leukemia is widely considered as a favorable prognostic factor. With the current treatment protocols these patients have superior overall survival (OS) and event free survival (EFS), but with frequent late relapse. We retrospectively collected clinical, cytogenetic and FISH data on 317 pediatric ALLs (0-18 years) ascertained during the period of January, 2007 to December, 2009 in this institution. In this cohort, 205 patients (64.7%) were examined for both chromosome abnormalities and for t(12;21) status with FISH; 53 (25.9%) of these were positive for (25.9%). We found that 13.6% (3/22) and 9.0% (8/89) of patients in groups of 0-1 years and 6-18 years are t(12;21) positive, respectively(P=0.453); this is in contrast to the 44.7% (42/94) of patients positive for t(12;21) in group of 2-5 years (P<0.001). The high incidence of t(12;21) especially in group of 2-5 years, suggests examining t(12;21) status by FISH should be routinely carried out at diagnosis. The frequency of abnormal karyotype increased with age of the patient; however, karyotypic complexity appeared to be not associated with age. FISH study revealed additional genetic events, such as del(12p), additional copy of 21, or additional copy of der(21)t(12;21) in majority (87.6%, 46/53) of patients that are positive for t(21;21). This suggests that these secondary additional events may play a role in the biological behavior of the disease. We also found that patients with either the additional copy of 21, or with the additional copy of der(21)t(12;21) responded well to the current treatment protocols. The deviation from this observation is one patient with both these abnormalities appearing in a secondary clone. This patient failed to achieve significant reduction in blasts at day 15 and expired. A previous study of patients with relapsed disease indicated poor outcome is associated with additional copy of chromosome 21, or der(21), or both. Data on additional patients with these changes at diagnosis is needed to confirm our inference.

406

Detection of clinically relevant genomic aberrations in hematological malignancies through routine application of multiplex ligation-dependent probe amplification (MLPA) based screening test coupled to conventional karyotype analysis. *C. Pangalos[1,2,3], S. Karapanou[2], M. Karambela[3], B. Hagnefelt[2], C. Konialis[2].* 1) Clinical Genetics, InterGenetics - Diagnostic Genetic Center, Athens, Greece; 2) Molecular Genetics, InterGenetics - Diagnostic Genetic Center, Athens, Greece; 3) Cytogenetics, InterGenetics - Diagnostic Genetic Center, Athens, Greece.

The presence of various large scale chromosomal abnormalities, both numerical and/or structural, is a frequent finding in clonal hematopoietic malignant disease, and these are traditionally revealed through routine karyotyping and/or FISH analysis of bone marrow or peripheral blood samples, thereby permitting disease classification and prognosis. More recently, the application of array CGH analysis has also uncovered many new cryptic genomic copy number imbalances, most of which are now recognized as clinically useful markers of hematological malignancies. In view of the severe limitations of both FISH and array CGH techniques, in terms of their routine application, we have recently included MLPA as routine testing in addition to classic karyotype analysis, for the detection of copy number changes involving chromosomal regions 2p24, 6q25-26, 8q24, 9p21, 10q23, 11q22.3, 13q14 and 17p13.1 (P037, P038, P040, MRC-Holland, The Netherlands). This test panel was applied to DNA extracted from 310 consecutive bone marrow samples, after 24hr short-term culture, referred to our center for routine karyotype analysis. Approximately half of the samples (158) originated from newly investigated patients with an undiagnosed hematological problem. MLPA results were obtained for 308 samples (99%) and interpretation of results was possible in 300 samples, as 8 samples exhibited multiple chromosomal abnormalities, interfering with MLPA normalization. We discovered genomic aberrations in a total of 14 patients (4.7%), with an otherwise normal karyotype or where the karyotype involved abnormalities of other (different) chromosomes. It is worth noting that 9 of these patients (~66%) belonged to the group of newly investigated patients. Copy number aberrations were also detected in 6 samples, where the MLPA results revealed genomic abnormalities which could be attributed to an abnormal karyotype (for example, 8q34 duplication and trisomy 8). The aberrations detected involved: 3 with loss of 9p21 (CDKN2A_B), 3 with loss of 13q14.2-3 (RB1, DLEU, MIRN15A), 3 with gain of 8q24 (myc), 2 with loss of 2p24 (MYCN), 2 with loss of 17p13.1 (TP53) and one with loss of 11q22.3 (ATM). Our data clearly indicate that application of this routine screening panel via MLPA, as an adjunct to routine karyotype analysis, provides a highly sensitive, rapid and relatively cheap approach for uncovering clinically important genomic abnormalities which would have remained otherwise undetected.