American Society of Human Genetics 63rd Annual Meeting
October 22–26, 2013 Boston

PLATFORM ABSTRACTS

1

Whole exome sequencing of 94 matched brain metastases and paired primary tumors reveals patterns of clonal evolution and selection of driver mutations. *S.L. Carter[1], P.K. Brastianos[2,3], S. Santagata[4], A. Taylor-Weiner[1], P. Horowitz[4], K. Ligon[4], J. Seaone[5], E. Martinez-Saez[5], J. Tabernero[5], D. Cahill[3], S. Paek[6], I. Dunn[4], B. Johnson[2], M. Rabin[2], N.U. Lin[2], R. Jones[2], P. Hummelen[2], A. Stemmer-Rachamimov[3], D.L. Louis[3], T.T. Batchelor[3], J. Baselga[7], R. Beroukhim[2], G. Getz[1,3], W.C. Hahn[2].* 1) Cancer genome analysis, Broad institute, Cambridge, MA; 2) Dana-Farber Cancer Institute, Boston, MA; 3) Massachusetts General Hospital, Boston, MA; 4) Brigham and Women's Hospital; 5) Vall D'Hebron University Hospital; 6) Seoul National University College of Medicine; 7) Memorial Sloan-Kettering Cancer Center.

Cancer metastasis to the brain is the most common malignancy of the brain, affecting approximately 200,000 patients per year. Brain metastases are associated with high morbidity and mortality, with a median survival time of 3–4 months. Despite the high incidence, relatively little is known about the molecular mechanisms driving brain metastasis. We subjected 94 trios consisting of primary tumor, brain metastasis, and normal reference tissue to whole exome sequencing (WES). To analyze the data, we developed novel computational tools to derive high quality allelic copy-number profiles directly from the WES data. These were used to perform an integrative analysis of somatic copy-number alterations (SCNAs) and somatic single nucleotide variants (SSNVs). This analysis allowed us to estimate the clonal architecture of the primary and metastatic samples from each patient, and to reconstruct a single phylogenetic tree relating all of the subclones in both samples. Every metastasis developed from a single clone, consistent with a single cell of origin. In some cases, we determined that the primary sample represented an ancestral population with respect to the metastases, based on the presence of subclones present in the primary that became fully clonal in the metastasis. In other cases, we observed fully clonal mutations in the primary sample that were not present in the metastasis, indicating a sibling relationship between the two samples. Subclonal mutations in the metastasis by definition occurred within the brain; these mutations displayed different mutational signatures than those acquired in the primary tumor. These contrasts were most pronounced in cases of lung cancer or melanoma, with tobacco and UV signatures prominent in these primaries and nearly absent from the mutations acquired after metastasis. In order to understand the molecular drivers of clonal evolution and metastasis in our data, we annotated each subclone with driver mutations identified using large numbers of cancer samples analyzed by the cancer genome atlas (TCGA) consortium. This produced a detailed portrait of each patient's cancer, with nearly node in each phylogenetic tree associated with at least one driver mutation.

2

Pathogenic de novo SNVs, indels and CNVs in 1,000 children with undiagnosed developmental disorders. *M. Hurles[1], M. van Kogelenberg[1], T. Fitzgerald[1], W.D. Jones[1], D. King[1], P. Vijayarangakannan[1], S. Gerety[1], K. Morley[1], S. Gribble[1], D. Barrett[1], K. Ambridge[1], N. Krishnappa[1], E. Prigmore[1], D. Rajan[1], T. Bayzetinova[1], S. Al-Turki[1], A. Tivey[1], S. Clayton[1], R. Miller[1], P. Jones[1], N. Carter[1], C. Wright[1], J. Barrett[1], D. FitzPatrick[2], H. Firth[1,3], DDD Study.* 1) Wellcome Trust Sanger Inst, Cambridge, United Kingdom; 2) MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, Western General Hospital, Edinburgh, UK; 3) East Anglian Medical Genetics Service, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, UK.

To delineate the genetic architecture of severe undiagnosed developmental disorders in UK children we have deeply phenotyped over 6,000 affected children and their parents through a nationwide network of clinical geneticists, and recruited the families into a genetic research study entitled the Deciphering Developmental Disorders study. Seventy-five percent of the families are sporadic. We are interrogating the causal roles of coding and regulatory SNVs, indels and CNVs by applying exome-array comparative genomic hybridization (exome-aCGH) to detect deletions and duplications, and exome-sequencing to detect sequence variants, in all coding exons, known enhancers, and the most highly conserved non-coding elements. We have profiled over 5,000 probands using exome-aCGH, and over 1,000 parent-proband trios with exome sequencing. We are currently able to provide likely diagnoses for 15–20% of children. We have identified recurrent functional de novo mutations in 45 genes, of which only 19 are already known developmental disorder genes. The largest single contributing gene is ARID1B. We have identified four genes where exactly the same mutation occurs in 2 or more families, highly suggestive of gain-of-function mutations. In aggregation, these analyses have identified more than 10 likely novel developmental disorder genes. We have modeled some of these plausible candidate genes in zebrafish and identified concordant developmental phenotypes in morphant zebrafish for a subset of these. We will describe the breakdown of these pathogenic and putatively pathogenic variants by phenotype and family history.

3

Chromatin loops and CNVs: the complex spatial organization of the 16p11.2 locus. *M.N. Loviglio[1], M. Leleu[2], N. Ghedolf[1], E. Migliavacca[1], K. Männik[1], J.S. Beckmann[3,4], S. Jacquemont[3], J. Rougemont[2], A. Reymond[1].* 1) Center for Integrative Genomics (CIG), University of Lausanne, Lausanne, Switzerland; 2) EPFL, Lausanne, Switzerland; 3) Service de Génétique Médicale, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland; 4) Swiss Institute of Bioinformatics, Lausanne, Switzerland.

Hemizygosity of the 16p11.2 ~600kb BP4-BP5 region (29.5 to 30.1Mb) is one of the most frequent known genetic etiologies of autism spectrum disorder (ASD). It is also associated with a highly penetrant form of obesity and a significant increase in head circumference. Mirror phenotypes are observed in carriers of the reciprocal duplication, who present a high risk of being underweight, microcephalic and/or schizophrenic. The distal 16p11.2 220kb BP2-BP3 deletion is similarly associated with obesity and neuropsychiatric disorders. We assessed possible chromatin interplays between these regions via long-range acting regulatory elements using high-resolution Chromosome Conformation Capture Sequencing (4C-seq) technology. We compared the three dimensional organization at the 16p11.2 locus between normal copy number and 600kb deletion or duplication state using the promoters of the SH2B1, MVP, KCTD13, ALDOA, TBX6 and MAPK3 genes as "viewpoints". The analysis of normal copy number samples highlights complex chromatin looping between genes located in the 600kb and 220kb regions. In particular, blocks of regulators in chromosomal kontext of the 5 viewpoints from the 600kb interval encompass the genes CD19, LAT, RABEP2, TUFM and SH2B1, whose polymorphisms and mutations were previously associated with BMI, serum leptin, maladaptive behaviors and obesity. This interaction was reciprocally confirmed using SH2B1 as a viewpoint. To gauge whether the presence of a genomic rearrangement alters any of the identified chromatin interactions along chromosome 16, we compared interaction profile signals of deletions and duplications of the 600kb BP4-BP5 region and of controls. Considering all viewpoints we identified 342 and 378 regions whose looping intensities are significantly modified in deleted and duplicated samples, respectively. In parallel, we profiled the transcriptome of lymphoblastoid cell lines of 50 600kb BP4-BP5 deletion, 32 reciprocal duplication and 29 control individuals and identified 1188 differentially expressed (DE) genes using a numerical variable to reflect a dosage effect. 27 of the 74 DE genes (36.5%) mapping on chromosome 16 show concomitant significant changes in chromatin interaction. Our results show that relevant chromatin conformation changes may arise from copy number variants. They suggest a link between the observed chromatin perturbations and gene expression and a possible contribution of the chromosome conformation to the disease phenotype.

4

Fine-mapping GWAS followed by genome editing identifies an essential erythroid enhancer at the HbF-associated *BCL11A* locus. *D.E. Bauer[1], S. Lessard[3], S.C. Kamran[1,2], J. Xu[1], Y. Fujiwara[1], C. Lin[1], Z. Shao[1], M.C. Canver[1], E.C. Smith[1], L. Pinello[6], P.J. Sabo[4], J. Vierstra[4], R.A. Voit[5], G.C. Yuan[6], M.H. Porteus[5], J.A. Stamatoyannopoulos[4], G. Lettre[3], S.H. Orkin[1,2].* 1) Pediatric Hematology/Oncology, Boston Children's Hospital, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA., United States; 2) Howard Hughes Medical Institute, Boston, MA; 3) Montreal Heart Institute and Université Montréal, Montreal, Quebec, Canada; 4) Department of Genome Sciences, University of Washington, Seattle, WA; 5) Department of Pediatrics, Stanford University School of Medicine, Palo Alto, CA; 6) Harvard School of Public Health, Boston, MA.

Introduction: Genome-wide association studies (GWAS) have ascertained numerous trait-associated common genetic variants localized to regulatory DNA. The hypothesis that regulatory variation accounts for substantial heritability has undergone scarce experimental evaluation. Common variation at *BCL11A* is estimated to explain ~15% of the trait variance in fetal hemoglobin (HbF) level but the functional variants remain unknown. **Materials and Methods**: We use chromatin immunoprecipitation (ChIP), DNaseI sensitivity and chromosome conformation capture to evaluate the *BCL11A* locus in primary erythroblasts. We extensively genotype 1263 samples from the Collaborative Study of Sickle Cell Disease within three HbF-associated erythroid DNase I hypersensitive sites (DHSs) at *BCL11A*. We pyrosequence heterozygous erythroblasts to assess allele-specific transcription factor binding and gene expression. We conduct transgenic analysis by mouse zygotic microinjection and genome editing with transcription activator-like effector nucleases (TALENs). **Results**: Common genetic variation at *BCL11A* associated with HbF level lies in noncoding sequences decorated by an erythroid enhancer chromatin signature. Fine-mapping this putative regulatory DNA uncovers a motif-disrupting common variant associated with reduced transcription factor binding, modestly diminished BCL11A expression and elevated HbF. The surrounding sequences function *in vivo* as a developmental stage-specific lineage-restricted enhancer. Genome editing reveals that the enhancer is required in erythroid but dispensable in B-lymphoid cells for expression of BCL11A. **Conclusions**: We describe a comprehensive and widely applicable approach, including chromatin mapping followed by fine-mapping, allele-specific ChIP and gene expression studies, and functional analyses, to reveal causal variants and critical elements. We assert that functional validation of regulatory DNA ought to include perturbation of the endogenous genomic context by genome editing and not solely rely on *in vitro* or ectopic surrogate assays. These results validate the hypothesis that common variation modulates cell type-specific regulatory elements, and reveal that although functional variants themselves may be of modest impact, their harboring elements may be critical for appropriate gene expression. We speculate that the GWAS-marked *BCL11A* enhancer represents a highly attractive target for therapeutic genome editing for the major β-hemoglobin disorders.

5

Translating dosage compensation to Trisomy 21: a novel approach to Down syndrome. *J.B. Lawrence[1], J. Jiang[1], Y. Jing[1], C.J. Cost[2], J. Chiang[1], H.J. Kolpa[1], A.M. Cotton[3], D.M. Carone[1], B.R. Carone[1], D.A. Shivak[2], M. Byron[1], P.D. Gregory[2], C.J. Brown[3], F.D. Urnov[2], L.L. Hall[1].* 1) Department of Cell and Developmental Biology, University Massachusetts Medical School, Worcester, MA; 2) Sangamo BioSciences, Richmond, CA; 3) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada.

DS researchers have sought to define the more "DS critical" genes on Chr21, but this has proven difficult due to high genetic complexity and phenotypic variability of DS, confounded by normal variation between any individuals. There remains a critical need for better ways to understand the underlying cell and developmental pathology of human DS, key to design of therapeutics. Despite advances in strategies to correct single-gene defects of rare monogenic disorders *in vitro* and in some cases *in vivo*, genetic correction of chromosomal imbalance in living trisomic cells has been outside the realm of possibility, in any context. Several years ago we began work on a high-risk project to test the idea that functional correction of trisomy may be feasible by inserting a single gene that can epigenetically silence a whole chromosome. Nature has evolved a mechanism to dosage compensate X-chromosome dosage differences between mammalian males and females: the X-linked XIST silences one X-chromosome in female cells. We previously showed that XIST produces a large non-coding RNA that coats the whole chromosome in nuclei. The RNA induces a host of heterochromatin modifications that transcriptionally silence one X-chr *in cis*. While XIST has been intensely studied, its potential translational relevance for chromosome pathology has not been pursued.

Therefore, we first demonstrated that the targeted addition of a very large 21 kb XIST transgene could be efficiently achieved in human cells, using genome editing with zinc finger nucleases. With 99% accuracy, we inserted XIST into an intron of the DYRK1A locus on Chr21, thereby generating an inducible system to express XIST RNA on one Chr21 in DS patient-derived pluripotent stem cells. Remarkably, the RNA localizes across and comprehensively silences one of the three Chr 21s, as shown by eight different methods, including molecular, cytological, and genomic. Reversal of specific cellular phenotypes was demonstrated. Results show the clear promise of this new strategy as a novel approach to identify DS cellular pathologies and genome-wide pathways most directly perturbed by trisomy 21, distinct from pervasive genetic and epigenetic variation between cell isolates and subclones. Silencing of trisomy 21 by manipulation of a single gene in living cells *in vitro* surmounts the first major obstacle to development of potential "chromosome therapy".

6

Insights into population history from a high coverage Neandertal genome. *D. Reich[1], for.the. Neandertal Genome Consortium[2].* 1) Harvard Medical School, Department of Genetics, 77 Ave. Louis Pasteur, Boston MA 02115 USA; 2) Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig 04103 Germany.

We have sequenced to about 50-fold coverage a genome sequence from about 40 mg of a bone found in Denisova Cave in Southern Siberia. The genome of this female is much more closely related to the low-coverage Neandertal genomes from Croatia, Spain, Germany and the Caucasus than to the genome of archaic Denisovans, a sister group of Neandertals, and provides unambiguous evidence that both Neandertals and Denisovans inhabited the Altai Mountains in Siberia. The high-coverage Neandertal genome, combined with our earlier sequencing of a high quality Denisova genome, allows novel insights about the population history of archaic humans:

- We document recent inbreeding in this Altai Neandertal. The inbreeding coefficient of about 1/8 corresponds to about the homozygosity that would be expected from a mating of half siblings.
- The Altai Neandertal genome shares almost seven percent more derived alleles with present-day Africans than does the Denisova genome. This means that the Denisovans derived a proportion of their ancestry from a very archaic human lineage, and the amount of this ancestry they inherit is larger than in Neandertals.
- The Denisovan genome is affected by major recent gene flow from an Altai-related Neandertal.
- To further characterize the variation among Neandertals we sequenced the genome of a Neandertal from the Caucasus to about 0.5-fold coverage. Comparisons to present-day genomes show that the Neandertals who contributed genes to present-day non-Africans were more closely related to this Caucasian Neandertal than to the Neandertals we sequenced from the Altai.
- We built a map of Neandertal ancestry in modern humans, using data from all non-Africans in the 1000 Genomes Project. We show that the average Neandertal ancestry on chromosome X of present-day non-Africans is about a fifth of the genome average. It is known that hybrid incompatibility loci concentrate on chromosome X. Thus, this observation is consistent with a model of hybrid incompatibility in which Neandertal variants that introgressed into modern humans were rapidly selected away due to epistatic interactions with the modern human genetic background.

7

Annotation of pseudogenous gene segments by massively parallel sequencing of rearranged lymphocyte receptor loci. *R.O. Emerson[1], A.M. Sherwood[1], H.S. Robins[2], C.S. Carlson[2], M.J. Rieder[1].* 1) Adaptive Biotechnologies, Seattle, WA; 2) Public Health Sciences Division, FHCRC, Seattle, WA.

In order to generate a wide variety of functional T cell receptors and antibodies, lymphocytes undergo somatic rearrangement of the T cell receptor and immunoglobulin loci, each of which encodes dozens or hundreds of V, D and J gene segments in the germ line. As in many large gene families, many of these gene segments are classified as pseudogenes due to defects in primary sequence or motifs necessary for somatic rearrangement. Until now, a full annotation of pseudogene/functional status for each gene segment has proven elusive. Using next-gen sequencing of the T cell receptor beta (TCRB) and immunoglobulin heavy chain (IgH) loci in mature T and B cells from hundreds of healthy adults we have annotated the functional status of each V, D and J gene segment present in these loci and have identified functional genes that had been erroneously classified as pseudogenes. Briefly, random chance predicts that slightly less than one-third of somatic rearrangements at the TCRB and IgH loci will lead to in-frame transcripts with no premature stops; selection during lymphocyte maturation ensures that all mature T and B cells carry at least one rearrangement coding for a productive receptor, while the second allele also rearranges in some cells and can be out of frame, include a premature stop, or include a pseudogenous V, D or J gene segment. We have classified each gene segment as functional or pseudogene based on the proportion of in-frame rearrangements; in mature T and B cells the length of the CDR3 hypervariable region in rearrangements using functional gene segments has a pronounced periodicity at 3 nt, while no such feature exists in the case of pseudogenes. This discrepancy between functional and pseudogenous gene segments has allowed us to conclusively annotate the functional status of each gene segment in the complex TCRB and IgH immune receptor loci, and to examine the distribution of functional and pseudogenous alleles in hundreds of individuals.

8

High throughput sequence analysis of the TCR repertoire in glioma-associated immune dysregulation. *B. Grinshpun[1], J. Sims[2], Y. Feng[1,5], P. Canoll[3], P. Sims[1,4], J. Bruce[2], Y. Shen[1,5].* 1) Columbia Initiative in Systems Biology, Columbia University, New York, NY; 2) Department of Neurological Surgery, Columbia University, New York, NY; 3) Department of Pathology and Cell Biology, Columbia University, New York, NY; 4) Department of Biochemistry & Molecular Biophysics, Columbia University, New York, NY; 5) Department of Biomedical Informatics, Columbia University, New York, NY.

The adaptive immune system plays both causal and reactionary roles in the diseases of humans and other vertebrates. While overreaction leads to autoimmunity and allergy, and insufficiency leads to infection and occurs concomitantly with the development of many cancers, these dysfunctions stem from the ability to recognize and respond to specific self and non-self antigens. Diversity in the binding specificity of T cell receptors (TCRs) is generated by recombination of germline-encoded V, D, and J cassettes and the addition of random nucleotides between them. We have utilized high-throughput sequencing, combined with commercially-available systems for amplification of the complementarity-determining region 3 (CDR3) of the TCR-alpha and TCR-beta chains from the T cells, to examine the sequence repertoires of whole T cell populations. We have developed a computational pipeline for mapping TCR cassettes, in silico translation, and error analysis, and applied these tools to the development of glioblastoma (GBM), the most common and deadly brain tumor. We quantitatively compare the diversity of T cell populations between subjects and between the periphery and tumor tissue, assess clonal expansions, and investigate the relative contributions of recombination vs. junctional nucleotide diversity on the properties of the repertoires. We are interested in quantifying CDR3 diversity as a way to assess immune potential, and measuring clonal expansion to identify tumor specific TCRs. Recent advances in read length of the paired-end the Illumina MiSeq platform, allow us use sequence length-dependent alignment to reduce and in most cases eliminate uncertainty in V,J cassette mapping while leveraging read pairs to correct for errors in sequencing. Furthermore, we applied methods from ecology theory and information theory to describe CDR3 population size and select sequences prominent in GBM tissue. Combined with the preprocessing pipeline described above, this allows us to not only identify CDR3 sequences associated with or expanded in the tumor, but also to provide metrics for the contribution of cassette usage to the diversity of the CDR3 population. Quantitative, sequence-based characterization of the whole TCR repertoire, as well as the ability to tie specific CDR3s to the functional changes observed among T cells during disease states, promises new insights into the nature of T cell dysregulation and potential targets for vaccine and immunotherapy.

9

Extraction and analysis of clinical traits of multiple sclerosis using electronic medical records. *M.F. Davis[1], S. Sriram[2,3], W.S. Bush[1,4], J.C. Denny[4], J.L. Haines[1,2].* 1) Center for Human Genetics Research, Vanderbilt Univ, Nashville, TN; 2) Dept of Neurology, Vanderbilt Univ, Nashville, TN; 3) Vanderbilt Multiple Sclerosis Center, Vanderbilt Univ, Nashville, TN; 4) Dept of Biomedical Informatics, Vanderbilt Univ, Nashville, TN.

The clinical course of multiple sclerosis (MS) is highly variable, and research data collection is costly and time-consuming. We evaluated natural language processing techniques applied to electronic medical records (EMR) to identify MS patients and key clinical traits of disease course. We used four algorithms based on ICD-9 codes, text keywords, and medications to identify individuals with MS from a de-identified, research version of the EMR at Vanderbilt University. After identification of MS patients, we developed algorithms to extract detailed MS features capturing the clinical course of MS, including clinical subtype, presence of oligoclonal bands, year of diagnosis, year and origin of first symptom, Expanded Disability Status Scale (EDSS) scores, timed 25 foot walk scores, and MS medications. Algorithms were evaluated on a test set validated by two independent reviewers. We identified 5,789 individuals with MS. Positive predictive values for the clinical trait algorithms ranged from 87-99%. Recall values for clinical subtype, EDSS scores, and timed 25 foot walk scores were greater than 80%. DNA was available for 1,086 of the individuals through BioVU. These samples and 2,396 control samples were genotyped on the ImmunoChip. After extensive sample and SNP quality control, 1,031 cases, 2,226 controls, and 160,046 SNPs remained for analysis. At a nominal p-value of 0.05, 29 known MS loci were replicated in case-control analysis, further confirming the MS disease status of cases. Genome-wide analyses were conducted for each of the extracted MS features using linear regression for continuous measures, logistic regression for presence of oligoclonal bands, and Cox proportional-hazards regression for time to secondary progressive (SPMS). Analyses were adjusted for the first three principal components. No associations reached genome-wide significance, although multiple loci were associated in each analysis at a significance level $p < 1 \times 10\text{-}5$. The most significant result from time to SPMS analysis (127 individuals) was less than 100kb upstream from CADM3, which encodes a brain specific protein associated with inflammation, a hallmark feature of MS ($p=3 \times 10\text{-}7$). This work demonstrates that detailed clinical information is recorded in the EMR and can be extracted with high reliability, and that this data can be used to further understanding of the genetics of MS.

10

Haplotype of CpG related SNPs is associated with DNA methylation pattern. *Y. Ma[1], C.E. Smith[1], Y.C. Lee[1], L.D. Parnell[1], C.Q. Lai[1], J.M. Ordovas[1,2,3].* 1) Nutritional Genomics, Jean Mayer USDA-HNRCA at Tufts University, Boston, MA; 2) Department of Epidemiology, Centro Nacional Investigaciones Cardiovasculares (CNIC), Madrid, Spain; 3) Instituto Madrileño de Estudios Avanzados en Alimentación (IMDEA-FOOD), Madrid, Spain.

Background: DNA methylation occurs on CpG dinucleotides. Single nucleotide polymorphisms (SNPs) may affect DNA methylation by changing the formation of CpG dinucleotides. In this study, we defined those SNPs which could change the formation of CpG dinucleotides as “CpG related SNPs” (CGS). Each CGS has two types of alleles, which can either create or disrupt CpG dinucleotides. The condition when an individual carries the allele to create CpG dinucleotide is called CGS-C, while the condition when an individual carries the allele to disrupt CpG dinucleotide is called CGS-D. Methods: We applied a genome-wide scale and integrated bioinformatics analysis to publicly available datasets of both genotypes (HapMap project) and methylation patterns in B lymphocyte cell line (ENCODE project) to explore the relationship between these CGSs and DNA methylation from three perspectives: (1) whether the genotype of a single CGS affects the DNA methylation of that locus; (2) whether two CGSs in high linkage disequilibrium (LD) tend to be with the same type of allele (create or disrupt the CpG); (3) whether the haplotype consisting of multiple CGSs in high LD is associated with DNA methylation pattern of that region. Results: About 80% of CGS-Cs can be methylated. In addition, when two CGSs are in high LD, they tend to act as the same type of allele, which means that if the allele of one CGS is to create the CpG then the allele of another CGS in high LD tend to create the CpG also. This finding is highly consistent in both unrelated individuals and family members of HapMap CEU population. Finally, the haplotype of one LD block, consisting of multiple CGSs close to both CpG islands and promoter regions, is correlated with DNA methylation pattern according to both categorical and continuous analysis. Conclusion: CGSs which are close to CpG islands and promoters may affect DNA methylation pattern in the form of haplotype.

11

Comprehensive Blood Group Prediction Using Whole Genome Sequencing Data from The MedSeq Project. *W.J. Lane[1,2,3], I. Leshchiner[4], S. Boehler[1], J.M. Uy[1], M. Aguad[1], R. Smeland-Wagman[1], R.C. Green[3,6], H.L. Rehm[1,3,5], R.M. Kaufman[1], L.E. Silberstein[7] for The MedSeq Project.* 1) Department of Pathology, Brigham and Women's Hospital, Boston, MA; 2) Harvard Medical School Transfusion Medicine Fellow, Boston, MA; 3) Harvard Medical School, Boston, MA; 4) Genetics Division, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 5) Laboratory for Molecular Medicine, Partners Healthcare Center for Personalized Genetic Medicine, Cambridge, MA; 6) Department of Medicine, Brigham and Women's Hospital, Boston, MA; 7) Division of Transfusion Medicine, Department of Laboratory Medicine, Children's Hospital Boston, Boston, MA.

There are 339 phenotypically distinct red blood cell (RBC) blood group antigens. For 298 of these antigens, the molecular basis is known, comprising 48 genes and >1,100 alleles distributed across 34 blood group systems. Exposure to non-self RBC antigens during transfusion or pregnancy can lead to the development of alloantibodies, which on re-exposure can lead to clinically significant and even fatal complications. Therefore, it is vital to know which antigens are present on RBCs. However, traditional serologic phenotyping methods are labor intensive, costly, sometimes unreliable, and reagents are not always available. As such, routine antigen typing is only done for ABO and D antigens. A large percentage of blood is given for hematologic malignancies that will soon get routine whole genome sequencing (WGS). For a minor added cost this data could be used for RBC antigen prediction. However, there are no published reports of using WGS data to predict RBC antigens. This is likely for several reasons: (1) none of the existing WGS data sets have paired serologic RBC phenotypes, (2) there are no fully annotated and complete databases of genotypes to phenotypes, (3) all of the known alleles are defined using cDNAs numbered relative to the start codon without human genome coordinates, and (4) lack of software capable of RBC antigen prediction. We have created a fully interactive web site of all known blood group genotype to phenotype correlations, fully annotated with relevant information, and mapped to and visually overlaid to their corresponding human reference genome gene sequences, with algorithms to predict antigen phenotypes from inputted sequences. These predictions are part of the General Genome Reports for the 100 patients getting WGS as part of The MedSeq Project. We are also interpreting the antigen patterns to identify patients at risk of making difficult-to-match alloantibodies, potential rare donors, and those with blood group-associated resistance to malaria and norovirus. In addition, each patient is undergoing an extensive antigen phenotypic workup using traditional blood bank serology, which is being used to validate and improve our prediction strategies. As clinical WGS becomes pervasive we hope that comprehensive blood group prediction will be done on everyone, allowing for easy identification of rare donors and the prevention of alloantibody formation using extended upfront matching of antigens from sequenced recipients and donor.

12

pVAAST: A new method for family-based rare variant association testing. *C.D. Huff[1], H. Hu[1], H. Coon[2], S. Guthery[2], S. Tavtigian[2], J.C. Roach[3], Z. Kronenberg[2], J. Xing[4], G. Glusman[3], V. Garg[6], B. Moore[2], L.E. Hood[3], K.S. Pollard[5], D.J. Galas[7], D. Srivastava[5], M.G. Reese[8], L.B. Jorde[2], M. Yandell[2].* 1) Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, TX; 2) University of Utah School of Medicine, Salt Lake City, Utah, USA; 3) Institute for Systems Biology, Seattle, WA, USA; 4) Department of Genetics, Rutgers University, Piscataway, NJ, USA; 5) Gladstone Institute of Cardiovascular Disease and University of California, San Francisco, San Francisco, CA, USA; 6) Department of Pediatrics, The Ohio State University and Center for Cardiovascular and Pulmonary Research, Research Institute at Nationwide Children's Hospital, Columbus, OH, USA; 7) Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg; 8) Omicia, Inc., Emeryville, CA, USA.

Next-generation sequencing has become an important tool for identifying disease-causing variation in families. Although a variety of filtering methods have been successfully applied in family-based sequencing studies, few robust statistical methods are available to support such studies. The Variant Annotation, Analysis and Search Tool (VAAST) employs a variant association test that combines amino acid substitution and allele frequency information using a composite likelihood ratio test (CLRT). Here, we present a novel pedigree-based method, pedigree-VAAST (pVAAST), that expands VAAST to incorporate family data. The method evaluates the familial evidence using a model specifically designed for sequence data. This model is broadly similar to traditional linkage analysis but is more sensitive when the disease alleles are modestly rare (minor allele frequency 0.01-0.05), which is a critical parameter space for next-generation sequencing studies of common genetic diseases. The familial evidence at each locus from one or more families is incorporated directly into the CLRT to increase the accuracy and greatly decrease the bioinformatic complexity of disease-gene identification efforts. We calculate statistical significance using a combination of permutation and gene-drop simulation to account for both the family structure and the observed pattern of variation in cases and controls. pVAAST supports dominant, recessive, and de novo inheritance models, and maintains high power across a wide variety of study designs, from monogenic, Mendelian diseases in a single family to highly polygenic, common diseases involving hundreds of families. We also demonstrate pVAAST's utility on exome chip, exome sequence, and whole-genome sequence data for recessive, dominant (cardiac septal defects), and complex genetic diseases (breast cancer, autism, and familial suicide). Our results demonstrate that pVAAST is a powerful and highly flexible tool for identifying disease genes in family-based sequencing studies.

13

Functional genomic confounds of transgenic integration and methods for their delineation. *J.C. Jacobsen[1,2], C. Chiang[2], C. Ernst[2,3], A.J. Morton[4], C. Hanscom[2], S.J. Reid[1], R.G. Snell[1], M.E. MacDonald[2], J.F. Gusella[2], M.E. Talkowski[2].* 1) Centre for Brain Research, School of Biological Sciences, The University of Auckland, Auckland, New Zealand; 2) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 3) Department of Psychiatry, McGill University, Montreal, QC, Canada; 4) Department of Pharmacology, University of Cambridge, Cambridge, UK.

The integration of exogenous DNA into a host genome has been an important route to generation of animal and cellular models for mechanistic exploration into human disease and development of therapeutics. In most such models, little is known concerning the structural integrity of the transgene, the precise site of integration, or the impact of integration on the host genome. Here, we provide methodological optimizations for the delineation of transgenic sequences using next-generation sequencing and illustrate the potential impact of transgenic integration in biological systems. We sequenced the genome and delineated the transgenic structure of seven Huntington's disease (HD) transgenic animal models harboring the integration of two independent transgenes using targeted capture as well as whole-genome jumping libraries, followed by analyses of the integration site and assembly of the internal transgene structure. Our analyses revealed a transgenic architecture so complex as to be reminiscent of chromothripsis in some animals. A series of secondary experiments on the R6/2 mouse, the most widely used model system of HD, revealed significant structural rearrangement of the transgene, as well as insertion into a single location in chromosome 4, within intron 7 of the gene *Gm12695*, with coincident deletion of 5,444 bp of mouse DNA. Our analyses reveal that *Gm12695* is normally expressed at negligible levels in mouse brain, but is expressed at dramatically increased levels in the brains of R6/2 mice compared to non-transgenic (wild-type) animals. This effect was consistent across multiple R6/2 litters and animals with varying CAG repeat lengths and disease severity. These data suggest transgenic integration can represent a legitimate confound in biological models of disease, and vouches for sequence-level resolution of transgene insertions prior to extensive investment in phenotypic characterization. The molecular and bioinformatics approaches we have developed argue strongly that this base-pair level approach is now both feasible and advisable.

14

Improved exome prioritization of disease genes through cross species phenotype comparison. *D. Smedley[1], S. Köhler[2,3], A. Oellrich[1], K. Wang[4], C. Mungall[5], S.E. Lewis[5], S. Bauer[2,3], D. Seelow[6], P. Krawitz[2,3], C. Gilissen[7], M. Haendel[8], P. Robinson[2,3,9], Sanger Mouse Genetics Project.* 1) Wellcome Trust Sanger Institute, Cambridge, Cambridgeshire, United Kingdom; 2) Institute for Medical and Human Genetics, Charité-Universitätsmedizin Berlin, Berlin, Germany; 3) Max Planck Institute for Molecular Genetics, Berlin, Germany; 4) Zilkha Neurogenetic Institute, University of Southern California, Los Angeles, CA; 5) Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA; 6) Department of Neuropaediatrics, Charité-Universitätsmedizin Berlin, Berlin, Germany; 7) Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences, Nijmegen, The Netherlands; 8) University Library and Department of Medical Informatics and Epidemiology, Oregon Health & Sciences University, Portland, OR; 9) Berlin Brandenburg Center for Regenerative Therapies, Charité-Universitätsmedizin Berlin, Berlin, Germany.

Whole-exome sequencing has successfully identified over 100 new disease-gene associations in the last few years. However, many cases remain unsolved after exome sequencing. This is often due to the sheer number of candidate variants remaining after common filtering strategies such as removing low quality and common variants and those deemed non-pathogenic. The background level of ~100 genuine loss of function variants with ~20 genes completely inactivated in each of our genomes makes the identification of the causative mutation problematic when using these strategies alone. In some situations, further filtering may be possible by the use of multiple affected individuals, linkage data, identity-by-descent inference, identification of de novo heterozygous mutations from trio analysis, or prior knowledge of affected pathways. Where these strategies are not possible or have proved unsuccessful, we propose using an additional approach exploiting the wealth of genotype to phenotype data that already exists from model organism studies to assess the potential impact of these exome variants. We have developed an algorithm, PHenotypic Interpretation of Variants in Exomes (PHIVE), which integrates the calculation of phenotype similarity between human diseases and genetically modified mouse models with evaluation of the variants according to allele frequency, pathogenicity and mode of inheritance approaches. The approach can we used through our freely available web tool, Exomiser (http://www.sanger.ac.uk/resources/databases/exomiser). By large-scale validation using 100,000 exomes containing known disease associated mutations, we have demonstrated a substantial improvement (1.8-5.1 fold) over purely variant-based (frequency and pathogenicity) methods with the correct gene recalled as the top hit in up to 67% of samples, corresponding to an area under the ROC curve of over 95%. We conclude that incorporation of phenotype data can play a vital role in translational bioinformatics and propose that exome sequencing projects should systematically capture and utilize clinical phenotypes to take advantage of the strategy presented here.

15

Phased allele-specific expression analysis in integrated whole-exome and mRNA sequencing study in a family with non-random X chromosome inactivation. *S. Szelinger[1,2,3], V. Narayanan[2,4], J.J. Corneveaux[1,2], I. Schrauwen[1,2,5], A.L. Siniard[1,2], A.A. Kurdoglu[1,2], I. Malenica[1,2], K.M. Ramsey[1,2], M.J. Huentelman[1,2], D.W. Craig[1,2].* 1) Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ; 2) Center for Rare Childhood Disorders, Translational Genomics Research Institute, Phoenix, AZ; 3) Molecular and Cellular Biology Interdisciplinary Graduate Program, School of Life Sciences, Arizona State University, Tempe, AZ; 4) Pediatric Neurogenetics Center, Barrow Neurological Institute, Phoenix, AZ; 5) Department of Medical Genetics, University of Antwerp, Antwerp, Belgium.

We describe a diagnostic framework for high-throughput sequencing data to determine magnitude and mode of non-random X chromosome inactivation (XCI) in a clinical female patient diagnosed with mild cognitive impairment, using integrated, family trio-based whole-exome and mRNA sequencing. Female carriers of deleterious genetic variations causing X-linked disease can present with heterogeneous phenotypes due to XCI. Random XCI in females results in an approximate equal ratio of cells expressing maternal or paternal X chromosome genes. When XCI is non-random, this ratio can vary from mildly skewed XCI (e.g. 70:30) to completely skewed XCI (e.g. 0:100). Current screening methods in clinical practice to detect biased X chromosome inactivation rely on indirect observation of expression by DNA methylation status of a single genetic locus. Truseq exome sequencing of the trio on two Hiseq2000 lanes resulted in a mean base coverage of the captured exonic regions of 86.25X. TruSeq mRNA sequencing of the trio on a single lane of a HiSeq2000 run resulted in an average of 116 million paired, mapped reads. Heterozygous SNP genotypes on chromosome X were phased in the patient by following the transmission of alleles identified in the parental exome SNP set. Phasing uncovered the paternal X as the source of a de novo, interstitial, 1.7 Mb deletion on the Xp22.31 locus that is recurrently been identified as the source of mental retardation phenotype in X-linked diseases. Allele-specific mRNA expression analysis of phased SNP alleles identified 80:20 non-random XCI that favored the expression of the maternal, cytogenetically normal X and compared well to the 85:15 ratio determined by the standard DNA methylation assay. In conclusion, the integrated whole-exome and mRNA-seq data suggests that the deleterious effect of the deletion on the paternal copy may be offset by skewed XCI that favors expression of maternal X resulting in the mild phenotype. These results demonstrate the utility of the combined approach of high-throughput DNA and RNA sequencing, phase-by-transmission allelic expression analysis in understanding the molecular basis of X-linked disorders with moderate phenotypes.

16

The Utility of the Traditional Medical Genetics Diagnostic Evaluation in the Context of Next-Generation Sequencing for Undiagnosed Genetic Disorders. *V. Shashi[1], A. McConkie-Rosell[1], B. Rosell[1], K. Schoch[1], K. Vellore[1], M. McDonald[1], Y.-H. Jiang[1], P.-X. Xie[2], A. Need[2], D. Goldstein[2].* 1) Pediatrics, Division of Medical Genetics, Duke University Medical Center, Durham, NC; 2) Center for Human Genome Variation, Duke Universlty Medical Center.

The diagnostic process used by most medical geneticists is tiered, with clinical assessment followed by sequential laboratory testing, contingent upon the previous tests being negative. Although there have been advances in diagnostic tests there have been no studies of unselected patients to determine the diagnostic yield of this traditional approach. We lack the data required to permit a meaningful comparison of NGS to the traditional paradigm, to provide a clinically and economically grounded diagnostic algorithm for patients with suspected genetic disorders. The purpose of this study was 1) to determine the diagnosis rate, type of disorders diagnosed, number of clinic evaluations to diagnosis and an approximation of the number/cost of the diagnostic tests performed and 2) to determine the appropriate patients to study with NGS and at what stage in their clinical evaluation. Utilizing a retrospective analysis we studied a cohort of 500 unselected consecutive patients (Jan- Sept 2011) from a general genetics clinics at a tertiary medical center. We included all patients who did not have a diagnosis at the time of their initial evaluation and excluded patients referred with a diagnosis or those followed in one of the genetics specialty clinics. Results: Of the 500 patients, 39 did not have a genetic disorder, 212 of the remaining 461 (46%) were diagnosed with a genetic disorder; 72% were diagnosed on the first clinic visit. Diagnostic tests were: Karyotype 12%, FISH 33%, microarray 20%, targeted single gene 13%, gene panel 20%, and biochemical 1%. Conclusions: Based on our findings we have developed a diagnostic algorithm to help guide the clinician. The most critical component of this algorithm is the clinical evaluation that considers the evolving phenotype, emerging diagnostic studies, and the strengths and limitations of the tests. For the majority of diagnosed patients; this diagnosis was considered and diagnostic testing was performed within the first clinical evaluation. Based on the cost of subsequent tests performed on those not diagnosed at the first evaluation, we estimated that the cost per subsequent successful genetic diagnosis was $25,000. For these patients, NGS may be beneficial. Estimating a 50% success rate for NGS in undiagnosed genetic disorders, its application to cases that remain unidentified after the first clinical visit would make the use of NGS in a general genetics clinic clinically and economically beneficial.

17

The National Institutes of Health Undiagnosed Diseases Program (UDP): The First Four Years. *D.R. Adams[1,2], C.F. Boerkoel[1], R. Godfrey[1], G. Golas[1], C. Groden[1], A. Gropman[3], D. Landis[1], T.C. Markello[1], M. Nehrebecky[1], T. Pierson[4], M. Sincan[1], C.J. Tifft[1,5], C. Toro[1], C. Wahl[1], L. Wolfe[1], W.A. Gahl[1,2,5].* 1) Undiagnosed Diseases Program, NIH; 2) Medical Genetics Branch, NGHRI/NIH; 3) Children's National Medical Center, Washington, DC; 4) Departments of Pediatrics and Neurology, Cedars-Sinai, Los Angeles, CA; 5) Office of the Clinical Director, NHGRI, NIH, Bethesda, MD.

The NIH UDP was started in 2008 to diagnose and further investigate persons who remain undiagnosed despite an extensive medical workup. The UDP is currently developing a national network of similar sites. We reviewed progress toward our original goals, and examined the cohort of individuals accepted into the program to date. The UDP has received 7000 inquiries, reviewed 2800 medical records, and accepted 640 participants since 2008. Participants were selected based on objective findings, safety considerations, and available expertise. Selection criteria evolved over time, balancing the need to serve a diverse population with an understanding of the characteristics of potentially tractable cases. Evaluations included review of medical records, in-person medical testing, and specialty consultation. Cases remaining undiagnosed after the initial visit were evaluated for potential transition to research, and the medical workup was designed to clarify the quantifiable portions of the medical presentation. Applicants were classified into 21 broad categories; half of the cases were neurological; 56% were female; 33% were pediatric. Some diagnosis was offered in 24% of the evaluated cohort, ranging from descriptive to definitive. Exome sequencing was performed on 480 individuals (150 probands). In the pediatric cohort, 17 diagnoses were made by exome or SNP-array analysis, while 17 were made by conventional methods. Two new diseases were published and > 30 new-disease candidates are under ongoing study. At least seven cases resulted in treatments. Five new clinical protocols have arisen as a result of UDP evaluations. Evaluation of exome data from UDP patients has resulted in the development of new exome analysis approaches and technologies. There is a substantial population with undiagnosed, severe medical conditions. Most initial evaluations do not generate a definitive diagnosis. The majority of persons with undiagnosed disease, therefore, require a process for transition to medical research. To that end, future efforts will focus on refining such processes, including the creation of standardized phenotypic descriptions, the incorporation of novel hypothesis-generating screens, and the development of online collaboration tools to facilitate information sharing among clinicians, researchers and affected families.

18

First year experience of Clinical Exome Sequencing for rare disease diagnosis at UCLA. *H. Lee[1], J.L. Deignan[1], N. Dorrani[3], F. Quintero-Rivera[1], S. Kantarci[1], K. Das[1], T. Toy[2], S. Strom[1], R. Baxter[2], T. Hambuch[9], Y. Xue[8], L. Li[7], C. Louie[6], D. Cherukuri[1], E. Lin[1], B. Harry[2], M. Yourshaw[2], M. Fox[3], C. Palmer[2,5], D. Wong[3], B.L. Fogel[4], W.W. Grody[1,2,3], E. Vilain[2,3], S.F. Nelson[1,2].* 1) Dept Pathology and Laboratory Medicine, Univ California, Los Angeles, Los Angeles, CA; 2) Dept Human Genetics, Univ California, Los Angeles, Los Angeles, CA; 3) Dept Pediatrics, Univ California, Los Angeles, Los Angeles, CA; 4) Dept of Neurology, Univ California, Los Angeles, Los Angeles, CA; 5) Dept of Psychiatry and Biobehavioral Sciences, Univ California, Los Angeles, Los Angeles, CA; 6) Clinical Molecular Diagnostics Laboratory, City of Hope National Medical Center, Duarte, CA; 7) Department of Pathology & Laboratory Medicine, Children's Hospital Los Angeles, Los Angeles, CA; 8) Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, CA; 9) Illumina Clinical Services Laboratory, Illumina, San Diego, CA.

The CLIA/CAP-accredited UCLA Clinical Genomics Center launched Clinical Exome Sequencing (CES) in 2012 to improve genetic diagnosis of rare Mendelian disorders. Validated as a single test from DNA extraction to Genomic Data Board interpretation, we have found CES to offer high clinical utility with a higher diagnostic yield than most of the other single-gene or panel-based clinical molecular tests. Here, we report on the first sequential 252 cases within the first year of testing. CES is offered as a Trio CES, where both parents also have CES and as a Proband CES, where only the patient is sequenced. Overall, a conclusive molecular diagnosis (causative mutation(s) identified in a clinically relevant gene) was provided for 80 of the 252 total cases (32%). 105 of the 252 cases were Trio CES, and 42 (40%) of those 105 were provided with a conclusive molecular diagnosis. De novo and phased compound heterozygous (CH) variants are the two major categories not possible to deduce without sequencing the unaffected parents. Among these 42 trios with a conclusive result, 43% had a de novo variant, 17% had a homozygous variant, 29% had CH variants in trans, 5% had an inherited variant from a similarly affected parent, and 7% had an X-linked hemizygous variant. Of the 134 Proband CES, we established a conclusive molecular diagnosis in 34 cases (25%). Thus, Trio CES provided a significantly higher diagnostic yield than Proband CES (p=0.01), and it is therefore the recommended approach. For genetically heterogeneous disorders such as developmental delay, the improved diagnostic rate and more straightforward interpretation with Trio CES are critical. We have also observed diagnostic 'reclassification' in 15% of the cases post-CES, largely due to the phenotypic heterogeneity of many Mendelian disorders. In 12 of the cases (5%) novel variants were identified in genes that are not yet associated with any human disorder but have animal models available with significant phenotypic overlap with the patient's condition. Most of these variants (7/12) were highlighted because they were de novo, and these genes are now being investigated as novel disease genes. We note that many cases referred for CES have been performed as a 'last resort' after exhausting most other diagnostic options. However, more recently we have observed a shift, where physicians order CES as the first line genetic diagnostic tool, and as a result, we expect the diagnosis rate to increase further.

19

Clinical utility of the first one thousand clinical whole exome sequencing tests: molecular diagnostic rate, changes in medical management, and the impact of incidental findings. *C. Eng[1,3], D. Muzny[2,3], J. Reid[2], M. Bainbridge[2], A. Willis[1], M. Landsverk[1], J. Beuten[1], M. Leduc[1], P. Ward[1], A. Braxton[1], M. Hardison[1], Z. Niu[1], R. Person[1], F. Xia[1], M. Bekheirnia[1], J. Scull[1], S. Wen[1], J. Zhang[1], A. Hawes[2], C. Buhay[2], Y. Ding[3], M. Scheel[3], N. Saada[3], W. Liu[3], J. Ma[3], J. Chandarana[3], L. Dolores-Freiberg[3], W. Alcaraz[1], H. Cui[1], M. Walkiewicz[1], E. Boerwinkle[2], S. Plon[1], J. Lupski[1], A. Beaudet[1], R. Gibbs[2], Y. Yang[1].* 1) Dept Molec & Human Genetics, Baylor Col Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor Col Medicine, Houston, TX; 3) Whole Genome Laboratory, Baylor Col Medicine, Houston, TX.

We developed and optimized technical, bioinformatic and interpretive whole exome sequencing (WES) pipelines in a CAP and CLIA certified lab to identify causative mutations underlying disease phenotypes in undiagnosed patients being evaluated clinically for genetic disorders. Approximately 13 Gb of data were generated for each clinical sample and a mean coverage of 160X was achieved with >95% of the targeted bases covered at 20X or higher. Of the 18773 genes in our exome capture 90% have >90% of the coding regions covered by WES. Since October 2011, over 1700 WES tests have been submitted to our clinical lab on a fee-for-service basis. The majority of ordering physicians are medical geneticists and neurologists evaluating pediatric-aged patients with neurologic phenotypes who previously had a variety of genetic and other(e.g. imaging) tests without an etiologic diagnosis. To date, 1000 WES tests have been completed and results reported. We identified 367 causative mutant alleles in 265 patients, achieving an overall molecular diagnostic rate of 26%. The solved diagnoses include 149 patients with autosomal dominant, 81 with autosomal recessive and 27 with X-linked disorders. For AD conditions, 91 patients carry de novo mutations, and the remaining 58 patients have mutations either inherited from a parent or of unknown inheritance status. Eight patients were found to have two genetic diagnoses which may have confounded recognition of the underlying syndromes due to “blended” phenotypes. Recurrent diagnoses include six cases of Noonan spectrum disorder and four cases of Cornelia deLange, as well as 19 intellectual disability cases caused by mutations in SWI/SNF complex genes including 12 mutations in the ARID1B gene. Interestingly, 20% of the diagnoses have been made in genes identified and reported in the literature during the last 3 years. Examples of situations in which the WES led to changes in treatment include three patients diagnosed with congenital myasthenia. Medically actionable incidental results unrelated to the indication for testing were reported for 77 patients (7.7%);the majority of conditions reported in this category were mutations in cardiovascular or cancer predisposition genes. This study evaluates our first 1000 completed clinical WES cases and demonstrates the clinical utility of WES as an efficient diagnostic approach for patients with non-specific or unusual clinical presentations suggestive of underlying genetic disorders.

20

Discordant Karyotype Results Among Non-invasive Prenatal Screening Positive Cases. *K.W. CHOY[1,2], K.Y. KWOK[1], E.T. LAU[3], M.H. TANG[3], A. PURSLEY[4,5], J. SMITH[4,5], S.W. CHEUNG[4,5], T.Y. LEUNG[1], A. PETAL[4,5].* 1) Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong SAR, China; 2) CUHK-Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China; 3) Department of Obstetrics and Gynaecology, Tsan Yuk Hospital, The University of Hong Kong, Hong Kong SAR, China; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 5) Medical Genetics Laboratories, Baylor College of Medicine, Houston, TX, USA.

OBJECTIVE: To evaluate the performance of non-invasive prenatal screening (NIPS) for detection of fetal aneuploidies in a multi-center cohort of NIPS high-risk cases referred for prenatal diagnosis. METHODS: Cases included were those referred for an invasive prenatal diagnosis procedure to verify aneuploidies (trisomies 21, 13, 18 and sex chromosomes aneuploidy) among screen positive NIPS cases. The NIPS results were reported from Ariosa, BGI, Natera, Sequenom, and Quest. Cytogenetics studies were performed by two referral prenatal diagnostic laboratories in Hong Kong and the Medical Genetics Laboratories at Baylor College of Medicine.The karyotype results were compared against the NIPS results, and validated by birth follow-up when possible. RESULTS: A total of 80 prospective NIPS positive samples were referred, among which 55 cases were indicated as high risk for trisomy 21, 12 cases for trisomy 18, 7 cases for trisomy 13, 6 cases for sex chromosomes aneuploidy. The NIPS detection rates were as follows: T21: 52 of 55 (94.5%), T18: 6 of 12 (50.0%), T13: 4 of 7 (57.1%), and sex chromosomes aneuploidy: 4 of 6 (66.7%). Overall, NIPS correctly detected 66 of the 80 (82.5%) screen positive cases, including three mosaic karyotype results: two mosaic trisomy 21 and one mosaic 47,XXX[31]/45,X[19]. CONCLUSION: Consistent with previously published studies, the detection rates vary among the different chromosomes with chromosome 21 being the highest followed by sex chromosomes, chromosome 13 and 18. More specifically, close to half of the patients with a screen positive NIPS result for trisomy 13, 18 and sex chromosomes aneuploidy had NORMAL karyotype results. These results support the recommendation that non-invasive prenatal screening should be regarded as an advanced screening test and invasive prenatal diagnosis by CVS or amniocentesis should be performed to confirm the NIPS findings. This has important implications for counseling patients with positive NIPS results.

21

Genetic testing for mitochondrial disorders in the next-generation sequencing era: Targeted multiplex PCR gene panel or capture-based WES? *R. Bai[1], N. Smaoui[1], E. Haverfield[1], J. Higgs[1], S.F. Suchy[1], D. Arjona[1], K. Retterer[1], A. Shanmugham[1], F.D. Kendall[2], S. Parikh[3], A.L Gropman[4], R. Haas[5], A. Goldstein[6], J. Panzer[7], S. Yum[7], M.J. Falk[7], R.P. Saneto[8], G.M. Enns[9], W.K. Chung[10], S. Bale[1], G. Richard[1]*. 1) GeneDx Inc, Gaithersburg, MD; 2) Virtual Medical Practice, LLC, Atlanta, GA; 3) Neurogenetics/Neurometabolism, Neurosciences Institute, Cleveland Clinic, Cleveland, OH; 4) Department of Neurology, Children's National Medical Center, Washington, DC; 5) Department of Neurosciences, University of California San Diego, La Jolla, CA; 6) Division of Child Neurology, Children's Hospital of Pittsburgh, Pittsburgh, PA; 7) Department of Pediatrics, Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA; 8) Seattle Children's Hospital/University of Washington, Seattle, WA; 9) Pediatric Genetics, Stanford University School of Medicine, Palo Alto, CA; 10) Columbia University Medical Center, New York, NY.

Majority of primary mitochondrial disorders (MtD) are caused by mutations in nuclear genes. Current genetic testing for MtD includes different approaches utilizing next generation sequencing. Some diagnostic laboratories use PCR-based target enrichment (PCR-TE) panels for well-established MtD genes, others use hybridization-based target enrichment (capture) of some or all MitoCarta genes, or whole exome sequencing (WES). To determine the clinical utility, we compared results from 201 unrelated patients with a suspected MtD; 155 cases were evaluated using a 101-gene panel by PCR-TE (Mito101) and 50 cases by capture-based WES, including 20 patients previously tested (2 positive, 18 negative) by Mito101. The cases in each group were classified into definite, probable, possible MtD according to published diagnostic criteria for MtD. All identified sequence variants were evaluated and classified as mutation (Mut), likely mutation (VLM), VUS, likely benign (VLB) or polymorphism. Of the 155 cases tested by Mito101, pathogenic mutation(s) were identified in 26 (17%) and thus confirmed a MtD. The positive rate was 38% in patients with a definite/probable MtD and 7% in those with a possible MtD. Of the 50 patients tested by WES, pathogenic mutation(s) were identified in 18 (36%), with a positive rate of 25% for patients with a definitive/probable MtD and 37% for those with a possible MtD. 58% of WES cases had an undetermined result, and 6% were negative. In 61% of positive cases, including 4 that had a previous negative result by Mito101, patients were found to have mutations in non-MitoCarta genes associated with metabolic, movement or neurological disorders, emphasizing that many disorders can have clinical features overlapping those of a MtD, and will be missed if testing is limited to the latter. However, WES technology does not allow complete coverage of all genes due to low coverage and/or pseudogene issues, and about 7% of the reportable variants (Mut/VLM/VUS/VLB) by Mito101 were not detected by WES. When comparing the clinical sensitivity for patients with definite/probable MtD, Mito101 had a greater yield (38%) over WES (25%). However, in patients with a less specific phenotype (possible MtD), WES had a distinct advantage (37%) over Mito101 (7%). Based on our data, a targeted MtD gene panel by PCR-TE and reflex to WES is recommended for patients with definite/probable MtD, whereas WES is recommended as the first-line test for patients with possible MtD.

22

Comparison of results between karyotyping and microarray testing in postnatal and prenatal specimens: karyotyping is not dead yet. *D. Warburton[1], V. Jobanputra[2], V. Aggarwal[2], A. Sobrino[3], M. Macera[3], O. Nahum[2], B. Levy[2]*. 1) Gen & Development, and Pediatrics, Columbia University, New York, NY; 2) Pathology, Columbia University, New York, NY; 3) N.Y. Presbyterian Hospital, Columbia University Medical Center, New York, NY.

The advent of microarray analysis (CMA) to determine dosage changes in chromosomal complements has led to differences of opinion about whether standard cytogenetic analysis by G-banding should remain a routine part of clinical testing. We have compared results from approximately one year of laboratory testing when both procedures were performed on the same specimen. Aneuploidy is underrepresented among these cases, since no microarray was done if initial FISH detected aneuploidy or trisomy was indicated clinically. We divided results into those where (1) both tests gave normal results (2) an abnormality had the same diagnosis by either technique (3) a clinically significant abnormality was found only by CMA (4) both techniques were required to interpret the results and (6) an abnormality was found only by karyotype. CMA findings of uncertain significance have been scored as normal. These occurred in 9.8% of blood specimens and 4.6% of prenatal specimens. Of 480 blood specimens, studied because of clinical abnormalities in the patients,71 (14.8%)were abnormal. Of these 36 (7.5%) had dosage changes detectable only by CMA, demonstrating the power of this technique. In 20 specimens (4.2%) there was an abnormality diagnosable by both methods (mostly aneuploidy). In 9 cases both karyotype and CMA were necessary for a complete interpretation, and in 6 cases only the karyotype revealed the abnormality. Thus karyotyping was important for diagnosis in 15/71 (21.1%) of abnormal specimens. In prenatal specimens the abnormality rate, of course, was much lower. In 284 specimens there were 19 (6.7%) abnormalities, 4 diagnosable by either method, 3 diagnosable by CMA only, 8 where both methods were required and 4 where karyotype alone revealed the abnormality. Thus karyotyping was important for diagnosis in 12/19 = 63.2% of abnormal specimens. The abnormalities detected only by karyotyping consisted of a reciprocal translocation, a Robertsonian translocation, four complex rearrangements, and 4 mosaics, including an XY line in a girl with Turner syndrome. All are abnormalities expected to be missed by CMA but which had major consequences for the patient. Cases where both karyotype and CMA were necessary for interpretation were chiefly those where an abnormal karyotype was better delineated by CMA results (e.g. markers, visible deletions and derived chromosomes.) Our analysis clearly demonstrates that karyotyping is still a valuable part of cytogenetic analysis.

23

Parental Studies of 2,248 Chromosomal Microarray Analysis (CMA) Cases -Role of parental studies in facilitating the interpretation of copy number variants. *W. Bi, J. Pham, J. Denham, E. Roney, A.N. Pursley, P. Stankiewicz, A. Breman, S. Lalani, J. Smith, C. Bacino, A. Patel, S.W. Cheung.* Baylor College of Medicine, Houston, TX.

Background: As the resolution of clinical arrays has increased, the number of copy number variants (CNVs) of unclear clinical significance (UCS) detected is also increasing. Performing parental studies on every CNV of UCS is not cost-effective; therefore, a systematic study is needed to develop guidelines to determine the type of CNV for which parental studies are most useful for result interpretation. Methods: We retrospectively identified 1710 of the 2248 CMA cases with parental study results where a single CNV was detected from over 16,000 cases evaluated by a V8.1 array with exonic coverage of >1,700 genes and back bone coverage of 30kb/probe. We examined the correlation between the clinical significance, size, and gene content of these CNVs and whether they were inherited or de novo. Results: The CNVs were located on an autosomal chromosome in 1546 cases: 559 abnormal CNVs, many of which have reduced penetrance, 975 variants of UCS, and 12 heterozygous carrier changes. For abnormal CNVs, 56% (80/142) of the gains were inherited, whereas 39% (162/417) of losses were inherited. In contrast, for UCS variants, 96% (627/654) of the gains were familial while 92% (294/321) of the losses were inherited. The UCS CNVs were divided into 2 groups based on size: <0.5 Mb or >0.5 Mb. Of the 405 gains <0.5 Mb, 98% were familial events while 93% were familial for the gains >0.5 Mb. UCS deletions <0.5 Mb were seen in 274 cases, of which 93% were inherited, whereas only 87% of the deletions >0.5 Mb were inherited. Notably, one parent carried the same CNV for 4 of the 5 gains >2 Mb and all 4 of the losses >1 Mb in regions with a paucity of genes. For the 164 CNVs involving chromosome X, 47 abnormal changes, 116 UCS variants, and 1 heterozygous carrier change were detected. Similar to the findings for autosomal chromosomes, the vast majority of UCS variants were inherited (97/100=97% for gains and 16/16=100% for losses). Detailed analysis of the various categories of CNVs will be presented. Conclusions: Our data shows that the vast majority of single UCS gains (<1 Mb) and losses (<0.5 Mb) in both autosomal chromosomes and chromosome X are inherited. Most of these CNVs are within non-disease associated regions and likely represent benign CNVs. Therefore, for single UCS CNVs, one should first consider the size of the CNV and gene content before requesting parental studies as larger gene-rich gains and losses >0.5Mb are most likely to yield informative results.

24

Rapid and cost-effective whole exome sequencing for clinical diagnosis and personalized medicine. *D. Muzny[1], M. Wang[1], C. Buhay[1], Y. Han[1], H. Dinh[1], C. Kovar[1], H. Doddapanei[1], M. Bainbridge[1], J. Reid[1], E. Boerwinkle[1,2], R. Gibbs[1].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030; 2) University of Texas Health Science Center at Houston, School of Public Health, Houston, TX 77030.

For genome sequencing to migrate from the laboratory to clinical care, it should be capable of delivering high-quality sequencing results in a fast, consistent and effective way. A comprehensive approach towards this goal has incorporated protocol development, robotics applications and a robust LIMS so that goals for both high capacity and fast turn around time can be accomplished. To further reduce the turnaround time for whole exome sequencing (WES) in clinical environment, we have developed a fast workflow taking advantage of an experimentally optimized protocol for quick WES library preparation and capture methods as well as shortened sequencing time using the Illumina HiSeq2500 platform. The quick WES library construction protocol allowed us to complete the entire NimbleGen-based whole exome capture library construction within 2 work days (0.5 day for pre-capture library preparation and another 1.5 days for target enrichment), whereas sequencing on HiSeq2500 has helped reduce the data generation time to approximately 30 hours. Kapa HiFi enzyme has been evaluated and utilized in PCR amplification to achieve more uniform target coverage. Validation of the workflow involved multiple co-capture and single-capture tests using benchmark human DNA samples (i.e. HapMap NA12878 and HS1011) and the HGSC-designed VCRome rebalanced probe set (~44Mb capture size). In a recent 5-plex co-capture quick WES experiments, with ~6.0 Gbs of raw data generated for the individual sample, 92% of the total targeted bases were sequenced with a minimum depth of 20 fold. Nearly all (i.e. 99.3%) of the designed VCRome exome targets were present in sequence reads and 98.4% of the targeted bases were covered by at least one read. While the capture efficiency was slightly affected by the short 1-day hybridization scheme, we observed approximately 69% of the reads mapping to the target regions. These experiments indicate that without sacrificing sequencing performance, we are able to reduce the WES turnaround time from the previous three weeks to 3-4 days. Such rapid turnaround times are necessary for WES to be routinely used in prenatal and neonatal intensive care settings.

25

Prioritizing Sequence Variants Using Statistical Evidence: Not All Measures are Alike. *W. Li[1, 2], L. Strug[2, 1], D. Pal[3].* 1) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 2) Program in Child Health Evaluative Sciences, Research Institute, Hospital for Sick Children, Toronto, Ontario, Canada; 3) Department of Clinical Neurosciences, Institute of Psychiatry, King's College London, London, United Kingdom.

Genetic association studies of sequence variants require prioritization for follow-up. Statistical evidence weighs heavily in the prioritization and standard statistical methods rank individual rare variants based on ordering p-values from Fisher's exact tests (P_{exact}) or chi-square tests with continuity correction (P_{cwc}). But not all measures of statistical evidence result in the same rankings, and there has been debate over whether p-values adequately measure statistical evidence. For rare variants where the disease-variant distributions can be summarized into 2x2 tables, we propose to rank variants using the ratio of conditional likelihoods evaluated at the maximum conditional likelihood estimate (MCLE) of the odds ratio versus at an odds ratio of one (maxLRc). In the special but common case where a 2x2 table contains at least one empty cell, the MCLE does not always exist; however, we show analytically that the maxLRc is always well defined, and is equal to the inverse of the hypergeometric probability of the *observed* data. Using sequence data from a study of speech disorder in Rolandic Epilepsy with 10 cases and 17 controls, we show that the rankings by maxLRc, P_{exact} and P_{cwc} can be quite different for the same set of variants. Through simulation studies, we show that the maxLRc achieves better rankings than p-values based on several metrics: (1) the rankings assigned by the maxLRc correlate better with the "true" rankings, where the "true" rankings are defined by ordering the underlying effect sizes, (2) given K variants are to be selected for follow-up, the maxLRc results in a greater number of truly associated variants; and (3) the truly associated variants are, on average, ranked higher by maxLRc than by P_{exact} or P_{cwc}. The maxLRc uses only information in the observed data, while p-values further incorporate the probability of more extreme data that could have been observed. The maxLRc, a likelihood ratio, is a measure of statistical evidence as defined in the Evidential statistical paradigm as opposed to the Frequentist or Bayesian paradigms. Theoretical developments show it has good operational characteristics to measure evidence, even in small samples. Our findings suggest that the maxLRc outperforms p-value-based prioritizations for rare variants. It is straightforward to implement, and extends to the prioritization of common variants and data with different configurations.

26

Mixed model association methods: advantages and pitfalls. *A. Price[1], J. Yang[2], N.A. Zaitlen[3], M.E. Goddard[4], P.M. Visscher[2].* 1) Harvard Sch Pub Hlth, Boston, MA; 2) University of Queensland, Brisbane, Australia; 3) University of California, San Francisco, CA; 4) University of Melbourne, Melbourne, Australia.

It is widely known that mixed linear model association (MLMA) methods can prevent false-positive associations due to population or relatedness structure, and increase power by applying a correction that is specific to this structure. Here, we present new results including theoretical derivations, simulations and application to empirical data to highlight several advantages and pitfalls of MLMA. We provide an analytical derivation for the loss in power of MLMA with the candidate marker included in the genetic relationship matrix (MLMi) vs. linear regression. We also provide an analytical derivation for the increase in power of MLMA with the candidate marker excluded from the genetic relationship matrix (MLMe) vs. linear regression. In large data sets, MLMe will have average chi-square statistics > 1 (which is appropriate, due to polygenic effects) whereas MLMi will have average chi-square statistics = 1 (which is not appropriate and leads to a loss in power). Next, we investigate the previously proposed approach of including only a subset of top associated markers in the genetic relationship matrix. We show that this approach can increase power for some genetic architectures, but can suffer an insufficient correction for false-positives in the case of subtle population structure. Finally, we consider ascertained case-control traits, and show that MLMA methods (including MLMe) suffer a loss in power as a function of the sample size and level of case-control ascertainment (which depends on disease prevalence, in studies with an equal number of cases and controls). The above results were validated via extensive simulations (involving both simulated and real genotypes) and application to empirical WTCCC2 data sets with multiple sclerosis and ulcerative colitis phenotypes spanning >20,000 samples. We observed large difference in test statistics for different MLMA approaches, including >20% higher test statistics (P-value < 10^{-20}) for MLMe vs. MLMi at 99 published markers known to be associated to MS and UC, consistent with our analytical derivations and demonstrating the large impact of the choices we describe. Software implementing the MLMe approach is available at http://www.complextraitgenomics.com/software/gcta/mlmassoc.html.

27

Assessing multivariate gene-metabolome associations using Bayesian reduced rank regression. *P. Marttinen[1], M. Pirinen[2], A.-P. Sarin[2,3], J. Gillberg[1], J. Kettunen[2], I. Surakka[2], A.J. Kangas[4], P. Soininen[4,5], T. Lehtimäki[6], M. Ala-Korpela[4,5,14], O.T. Raitakari[7,8], M.-R. Järvelin[9,10,11,12,13], S. Ripatti[2,3,15,16], S. Kaski[1].* 1) Helsinki Institute for Information Technology (HIIT), Aalto University and University of Helsinki, Finland; 2) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland; 3) Unit of Public Health Genomics, National Institute for Health and Welfare, Helsinki, Finland; 4) Computational Medicine, Institute of Health Sciences, University of Oulu and Oulu University Hospital, Oulu, Finland; 5) NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio, Finland; 6) Department of Clinical Chemistry, Fimlab Laboratories, University of Tampere School of Medicine, Tampere, Finland; 7) Department of Clinical Physiology and Nuclear Medicine, University of Turku and Turku University Hospital, Turku, Finland; 8) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku and Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland; 9) Department of Epidemiology and Biostatistics, MRC Health Protection Agency (HPA) Centre for Environment and Health, School of Public Health, Imperial College London, United Kingdom; 10) Institute of Health Sciences, University of Oulu, Finland; 11) Biocenter Oulu, University of Oulu, Finland; 12) Unit of Primary Care, Oulu University Hospital, Finland; 13) Department of Children and Young People and Families, National Institute for Health and Welfare, Oulu, Finland; 14) Computational Medicine, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom; 15) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 16) Hjelt Institute, University of Helsinki, Helsinki, Finland.

A typical genome-wide association study searches for associations between SNPs and a univariate phenotype. However, there is a growing interest to investigate associations between genomics data and multivariate phenotypes, for example, gene expression or metabolomics data. The commonly used approach is to perform a univariate test between each genotype-phenotype pair, and then to apply a stringent significance cutoff to account for the large number of tests performed. However, this approach may have limited ability to uncover dependencies involving multiple variables. Another trend in the current genetics is the investigation of the impact of rare variants on the phenotype, where the standard methods often fail due to the lack of power when the risk allele is present in only a limited number of individuals. Here we propose a novel approach based on Bayesian reduced rank regression to assess the impact of multiple SNPs on a high-dimensional phenotype. Due to the method's ability to combine information over multiple SNPs and phenotypes, our method is particularly suitable for detecting associations involving rare variants. We demonstrate the potential of our method by analyzing every gene in a sample of 4,702 individuals from the Northern Finland Birth Cohort 1966, for whom whole-genome SNP data along with lipoprotein profiles comprising 74 traits are available. Using our new method, we discovered three putative loci without previously reported associations with the traits studied, which replicated in a sample of 2,390 individuals from the Cardiovascular Risk in Young Finns study.

28

Characterizing shared pathogenetics from genome-wide association studies via principal component analysis. *A. Keinan[1,2], D. Chang[1,2].* 1) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 2) Program in Computational Biology and Medicine, Cornell University, Ithaca, NY.

Shared pathogenesis between diseases has been extensively studied through comorbidity. In the era of genome-wide association studies (GWAS), many studies have also reported associations that are shared between diseases. Such findings are especially common across autoimmune diseases, and have led to meta-analyses that combine distinct diseases to improve power of detecting shared associations. A few recent studies have also investigated this at the level of summary statistics; such as testing the correlation of a vector consisting of the strength and direction of association of many SNPs across several autoimmune GWAS. Here, we present a novel method for studying the relationship between different diseases from GWAS datasets. It extends on the above correlation method and overcomes its main shortcomings, namely that all SNPs contribute equally and that signals may be missed due to heterogeneity in population and genotyping arrays across datasets. Our method is based on principal component analysis (PCA), which has been extensively employed for assessing and correcting for population structure. It accounts for heterogeneity and for associated markers not necessarily being causal by first considering gene-level tests of association. The method proceeds by applying PCA to a matrix of significance scores across genes and across GWAS datasets. It also controls for possible confounders that differ across GWAS such as sample size and genotyping array, thereby focusing on the main difference between datasets -the disease under study. We applied our method to 30 GWAS of a range of diseases, including autoimmune, neurological, psychiatric, and cancer. The first few resulting principal components led us to four main observations: (1) Different GWAS of the same diseases lie close together. (2) Inflammatory bowel diseases (IBD) form a separate cluster, distinct from other autoimmune diseases. (3) Cancer and autoimmune diseases form a distinct cluster each, but neurological diseases exhibit some overlap with both. (4) Genes that play a dominant role in defining the principal components (i.e. underlie the observed structure amongst diseases) are significantly enriched for genes previously associated with IBD ($P<10^{-9}$)—as well as with other autoimmune and psychiatric diseases—and for genes in pathways involved in immune, such as antigen processing and presentation. These results stress the utility of our method for characterizing and quantifying *"shared pathogenetics"*.

29

Power studies for the "extreme versus control" design for exome sequencing studies and identification of a variant of TMC6 as a deleterious modifier for age of onset of chronic *Pseudomonas* infection in children with cystic fibrosis. *M.J. Emond[1], T. Louie[1], J. Emerson[2,7], R.A. Mathias[3], M.R. Knowles[4], D.A. Nickerson[5], H.K. Tabor[2,6], K.C. Barnes[3], R.L. Gibson[2,7], M.J. Bamshad[2,5,8], NHLBI Exome Sequencing Project, Lung GO.* 1) Department of Biostatistics, University of Washington, Seattle, WA; 2) Department of Pediatrics, University of Washington, Seattle, WA; 3) Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD; 4) Cystic Fibrosis/Pulmonary Research and Treatment Center, University of North Carolina at Chapel Hill, Chapel Hill, NC; 5) Department of Genome Sciences, University of Washington, Seattle, WA; 6) Trueman-Katz Center for Pediatric Bioethics, Seattle Children's Research Institute, Seattle, WA; 7) Division of Pulmonary Medicine, Seattle Children's Hospital, Seattle, WA; 8) Division of Genetic Medicine, Seattle Children's Hospital, Seattle, WA.

As part of the NHLBI Exome Sequencing Project (ESP), we identified rare variants in DCTN4 that modify age-of-onset of chronic *Pseudomonas* (Pa) infection in cystic fibrosis (CF) using an extreme phenotypes design, by-gene tests, and a small number of individuals in each extreme (n< 50). However, this design and sample size will be under-powered for discovery of many variants underlying complex traits. We have devised a more powerful design for the same overall cost as an extreme phenotypes design by making use of a large set of extant exomes from individuals suitable as controls to which we compare exomes from one extreme of phenotype. The power of this extreme vs. population control design depends on the extent of enrichment of variants in the extreme. For example, consider a scenario in which extreme A is expected to harbor causal variants at 3 times the frequency of the overall population, the opposite extreme is expected to have zero variants (an ideal situation for an extreme phenotypes design) and the cumulative derived allele frequency (sum of causal allele frequencies) is 5% in the overall population. Using an extreme phenotypes design with 50 individuals per extreme (n=100 total) provides 26% power to detect the association, whereas use of 50 individuals in extreme A compared to 3000 controls provides 46% power, and 100 individuals in extreme A compared to 3000 controls provides 87% power. For a 4X enrichment in extreme A, we estimate the powers to be 40%, 91% and 99.9%, respectively. We show power gains of similar magnitude in other scenarios. Applying this design to exomes from extremes of individuals with early chronic Pa (n=86) in CF compared to exomes from 3316 ancestry-matched control individuals using by-gene tests (adj-SKAT-O), we identified TMC6 as having one or more variants significantly associated with age-of-onset of chronic Pa ($p= 9.6\times10^{-7}$) and validated an association between age-of-onset and a TMC6 variant among 556 individuals from the Early Pseudomonas Infection Control Observational Study (p=0.0005, HR=5.2, 95% CI [1.3-2.8]). This TMC6 variant was also associated with an 8.0 percentile decrease in FEV_1 (p= 0.01). The design also has high power for by-variant analyses, and preliminary by-variant results based on 156 CF exomes results in 7 significant variants in 7 genes. This extreme phenotype vs. controls study design can be a low-cost, powerful strategy for discovery of novel variants associated with risk for complex traits.

30

A Novel Statistical Framework for Using 'Out of Study' Control Groups in Association Studies with Next Generation Data. *A. Derkach[1], T. Chiang[2], L. Addis[3], S. Dobbins[4], I. Tomlinson[5], R. Houlston[4], D.K. Pal[3], J. Gong[2], L.J. Strug[2,6].* 1) Statistics, University of Toronto, Toronto, ON, Canada; 2) Program in Child Health Evaluative Sciences, the Hospital Sick Children, Toronto, ON, Canada; 3) Department of Clinical Neuroscience, Institute of Psychiatry, Kings College London, UK; 4) Institute of Cancer Research, London, UK; 5) Welcome Trust Centre for Human Genetics, Oxford, UK; 6) Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada.

Genome wide next generation sequence (NGS) data, such as that from the 1000 Genomes Project, is publicly available. However, these data have generally been used as a blunt comparative tool to identify novel or rare variants, and not properly exploited as controls for association studies. One explanation for the underutilization of these data for association may be the existence of several potential biases or confounding factors such as differences in sequencing platforms; alignment, SNP and variant calling algorithms; read depth, and selection thresholds. Here we focus on the effect of read depth and bioinformatic aspects of variant calling in comparing allele frequencies between cases and controls that were resequenced as part of different experiments. We assume that other potential confounding factors are reasonably well matched. We illustrate analytically, and by simulation, how differences in read depth and variant screening parameters affect Type 1 error. We propose a novel likelihood-based method that re-purposes and extends an approach by Skotte et al. (2012). We suggest substituting genotype calls by their expected values given the observed sequence data to eliminate read depth bias from estimation of minor allele frequency (MAF). We then incorporate read depth differences into the variance estimation to control between-study variation in read depth. We conduct a comprehensive simulation study to show that our method controls Type 1 error when cases and controls are resequenced at different read depth, and show this applies to association studies using single or multiple variants. We applied this method to NGS data from a 600kb linkage region for an epilepsy endophenotype present in ~2% of the population. We used long-range PCR and NGS on 27 epilepsy cases, with average read-depth of 197x. Using the BAM files from the 174 low read depth (LRD) 1000 genomes controls (release 21/05/11) and the 27 high read depth (HRD) epilepsy cases we calculated the expected genotypes given the observed sequence reads to compare the two groups. We compared our findings to an analysis using variant calls from the 27 HRD epilepsy cases and 200 HRD controls sequenced by Complete Genomics (~35x). We show that the proposed method removes bias and identifies the same associated variants as analysis with the HRD group. In conclusion, out-of-study control groups can be used in association studies as a way to prioritize variants for follow-up in more focussed studies.

31

Using Coalescent-Based Modeling for Large-Scale Fine Mapping of Complex Trait Loci using Sequencing Data in Large-Scale Case-Control Studies. *Z. Geng[1], P. Scheet[3], S. Zöllner[1,2].* 1) Dept Biostatistics, Univ Michigan, Ann Arbor, MI; 2) Dept Psychiatry, Univ Michigan, Ann Arbor, MI; 3) Dept Epidemiology, Univ Texas MD Anderson Cancer Center, Houston, TX.

Association mapping based on linkage disequilibrium (LD) is widely used to identify genomic regions containing disease variants. However, due to the complicated genetic dependence structure, identifying the underlying risk variants for complex diseases is challenging. By modeling the evolutionary process that produces our sequencing data, coalescent-based approaches may extract more information to improve such mapping. Such methods provide the genealogy at all sites in the region we have sequenced. Therefore, we can model the probability of carrying risk variants at all loci jointly, and obtain Bayesian confidence intervals (CIs) where true risk variants are most likely to occur. Additionally, the genealogy at each position provides more information about the shared ancestry of neighboring sites. Indeed, such careful modeling of the shared ancestry of sequences may also be beneficial in haplotyping and variant calling in regions of interests (ROI) where traditional hidden Markov approaches struggle. However, existing coalescent-based methods typically suffer from a major challenge: computational intensity. Here, we propose a novel approach to overcome such difficulty, so that it can be applied to large-scale studies. First, we infer a set of clusters from the sampled haplotypes so that haplotypes within each cluster are inherited from a common ancestor. Then, we apply coalescent-based approaches to approximate the genealogy of ancient haplotypes at different positions across the ROI. Doing so, the dimension of external nodes in coalescent models is reduced from the total sample size to the number of clusters. Finally, we evaluate the position-specific cluster genealogy and their descendants' phenotype distribution, to integrate over all positions and establish CIs where risk variants are most likely to occur. In simulation studies, our method correctly localizes short segments around true risk positions for both rare (1%) and common (5%) risk variants in datasets with thousands of individuals, as opposed to traditional coalescent-based approaches that typically restrict the sample size to a few hundreds. In summary, we have developed a novel approach to estimate the genealogy throughout sequenced regions. In fine mapping of complex trait loci, our method is applicable for large-scale case-control studies using sequencing data.

32

Quantifying and partitioning variation due to genetic effects and population stratification using within-family prediction analysis. *J. Yang[1,2].* 1) on behalf of the GIANT Consortium; 2) Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia.

Genome-wide association studies (GWAS) of human height have employed large sample sizes (~130,000) and identified a large number of associated variants (180) (Lango Allan et al. 2010), but have nonetheless only accounted for ~10% of phenotypic variation, in contrast to the predicted ~45% of variance explained by all common SNPs (Yang et al. 2010). As such, expansion of GWAS of height could provide continued insights into the genetic architecture of this model human complex trait, with potential implications for studies of other complex traits and diseases. It has been observed for GWAS on height and other complex traits that the genomic inflation factor (λ_{GC}) increases with increasing sample size, which is consistent with both population stratification and polygenic variation (Yang et al. 2011a). Here we report results from a meta-analysis of 79 GWASs, comprising ~250,000 individuals of European ancestry. We observe a large genome-wide 'inflation' factor of the test statistic for association (λ_{GC}=1.94) even after we corrected each study's test statistics by its individual inflation factor (Devlin and Roeder 1999). We developed a within-family prediction approach, which is able to quantify the variation due to real SNP effects, population stratification and errors in estimating SNP effects, by comparing the difference in SNP-based genetic predictor and phenotype between full sibs selected from independent families. The analyses with and without fitting principal components clearly show that our proposed approach is able to distinguish variance component due to true association signals from those due to stratification and estimation errors. We also show that variance attributable to population stratification is minor for SNPs that passed genome-wide significance. We confirmed the variance due to real SNPs effects, as inferred from the within-family prediction analysis, by the whole-genome estimation analyses as implemented in GCTA (Yang et al. 2011b). The results show that ~16% of phenotypic variance can be explained 697 genome-wide significant SNPs and that ~29% of variance is captured by the best ~9500 SNPs selected from a multiple SNPs association analysis (Yang et al. 2012). Together, these results suggest that the observed large genomic inflation is consistent with a genetic architecture for human height that is characterized by a very large but finite number of causal variants (thousands), spread out over the genome.

33

Meta-imputation: a simple and flexible method to combine multiple reference panels for imputing genetic variants. *P.K. Albers[1], G.R. Abecasis[4], M.I. McCarthy[1, 2, 3], K.J. Gaulton[1].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Oxford, UK; 3) Oxford National Institute for Health Research Biomedical Research Centre, Churchill Hospital, Oxford, UK; 4) Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, USA.

Methods for genotype imputation have become integral tools for genetic association studies, in which a reference panel is used to infer genetic variants that were not typed in a genotype sample. Although large panels are available in the public domain, e.g. HapMap or 1000 Genomes, recent large-scale sequencing studies have made it possible to use a wider variety of sequence data for imputation. To exploit this ever-increasing abundance of information, it is desirable to combine available reference sets, as this is likely to increase the chance of imputing lower frequency and rare variants more accurately. However, constructing a combined panel from data produced in independent sequencing studies is difficult, as data would ideally be called together to produce a single reference containing all haplotypes. Practically, either the subset of variants that are present in all sets can be considered, which reduces their total number, or missing variants have to be "cross-imputed" into each set, which decreases statistical certainty in downstream imputation. Also, by adding more haplotypes and variants to a reference set, a considerable higher computational burden is imposed on imputation. Here, an alternative solution is proposed; meta-imputation is a simple and flexible approach that integrates multiple reference panels without interfering in the imputation algorithm. After separately imputing different references into a genotype sample, inferred genotype likelihoods are combined at overlapping sites, using imputation-dependent certainty scores in a weighting function. For evaluation of our meta-imputation approach, we subdivided 1000 Genomes data into split reference panels, emulating the situation when several reference sets are to be combined, and compared meta-imputation of split references with imputation of the full reference. We assessed the results in a comprehensive cross-validation procedure, repeatedly taking out a random genotype for comparison with the corresponding imputed and meta-imputed variant. We show that meta-imputation compares well to imputation in terms of accuracy as measured by R^2, as well as allele and rare variant error rates. Notably, our approach is useful to reduce computation time, because it can impute references separately in parallel. The prime benefit, however, is that meta-imputation allows to make an informed choice as to which sequence-based data to include as reference sets in the overall imputation process.

34

Epistasis is widespread in the genetic control of transcription in humans. *J. Powell[1,2], G. Hemani[1,2], K. Shakhbazov[1,2], A. Henders[3], A. McRae[1,2], N. Martin[3], G. Montgomery[3], P. Visscher[1,2].* 1) Complex Trait Genomics, The University of Queensland Diamantina Institute, Brisbane, Qld, Australia; 2) Complex Trait Genomics, Queensland Brain Institute, University of Queensland, Brisbane, Qld, Australia; 3) The Queensland Institute of Medical Research, Brisbane, Qld, Australia.

A long standing question in evolution and genetics is the extent to which epistasis, the phenomenon whereby one polymorphism's effect on a trait depends on other polymorphisms present in the genome, contributes to complex trait variation. Though epistasis has been demonstrated in artificial gene manipulation studies in model organisms, and some examples have been shown in other species, few convincing examples exist for variation due to epistasis among natural polymorphisms in human traits. Its absence from empirical findings may simply be due to its unimportance in the genetic variation for complex traits, but we hypothesized that it has previously been too technically difficult to detect for pairs of loci due to statistical power and computational issues. Here we show that, using advanced computation techniques and a gene expression study design, evidence for abundant pairwise epistatic loci is found. In a cohort of 842 individuals with data on 7339 gene expression levels in whole blood, we found that after stringent correction for multiple testing the expression of 249 genes is influenced by 549 significant pairwise epistatic interactions. We attempted replication in two independent datasets and 421 show evidence of significance in at least one dataset. We provide evidence of functional enrichment for the interacting SNPs, for instance 49 of the genetic interactions are located within 500kb of known chromosome interactions ($p < 1.1*10-70$). 129 genes are controlled by multi-locus epistatic interactions whereby one cis-acting single nucleotide polymorphism (SNP) is modulated by several trans-acting SNPs. For example, a gene on chromosome 3 is controlled by a cis-additive effect which itself is controlled by trans-SNPs on 18 different chromosomes, with nearly identical genotype-phenotype maps for each cis-trans interaction. This study presents the first strong evidence for the widespread existence of epistasis effects on trait variance emerging from natural genetic variation in humans.

35

Low-pass whole-genome sequencing in Europeans identifies 1325 SNPs and indels associated with cis gene expression of which 4% are independent low frequency-large effect associations. *A.R. Wood[1], M.A. Tuke[1], H. Yaghootkar[1], D. Pasko[1], H. Lin[2], C.S. Xu[2], D.G. Hernandez[3,4], M.A. Nalls[3], J.R. Gibbs[3], L. Qibin[2], S. Juan[2], A. Murray[1], D. Melzer[5], M.N. Weedon[1], A.B. Singleton[3], L. Ferrucci[6], T.M. Frayling[1].* 1) Genetics of Complex Traits, University of Exeter Medical School, Exeter, United Kingdom; 2) Beijing Genomics Institute, Beishan Industrial Zone, Yantian District, Shenzhen, China; 3) Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, 35 Lincoln Drive, Bethesda, MD, USA; 4) Department of Molecular Neuroscience and Reta Lila Laboratories, Institute of Neurology, UCL, Queen Square House, Queen Square, London WC1N 3BG, United Kingdom; 5) Institute of Biomedical and Clinical Sciences, University of Exeter Medical School, Exeter, United Kingdom; 6) Clinical Research Branch, National Institute on Aging NIA-ASTRA Unit, Harbor Hospital, MD.

Initial results from whole genome sequencing, exome sequencing and exome microarray based experiments suggest that there are relatively few low frequency-large effect variants associated with common human phenotypes. However, these sequencing experiments have yet to reach sample sizes similar to those required to identify most common variant - phenotype associations. We aimed to test the role of low frequency variants in common human phenotypes using the same sample sizes where multiple common associations were detectable. As phenotypes, we used 11,132 cis gene expression profiles from the whole blood of 450 individuals from the population based InCHIANTI study. To identify low frequency variants we performed low-pass (mean 7X) whole-genome sequencing in 680 of the InCHIANTI individuals. We imputed missing genotypes using Beagle software, performed extensive QC of variants using the Genome Analysis Toolkit software and assessed the quality of our data by comparing it to 2Mb of deep sequence data (>100X) from 83 overlapping individuals. To identify variant - cis gene expression associations (cis-eQTLs) we inverse normalised all phenotypes and performed analysis using 9,720,795 SNPs and 2,018,182 indels observed at least 4 times. We used a *P*-value of $P<1\times10^{-06}$ that represented a false discovery rate of 5% based on the number of independent cis variants and phenotypes. Where we detected low frequency variant - phenotype associations we performed conditional analyses using the strongest common variant as a second variable. Using our deep sequence data, we estimated that we had detected 88% of SNPs and 79% of indels in our low pass sequencing data of which 0.6% and 16% respectively were false positives. We identified 1325 cis-eQTLs (1065 SNPs and 260 indels), of which 87 were low frequency (minor allele <5%) and had an average effect size of 1.36 standard deviations (SDs) (range: 0.80-2.32) compared to the 1238 common variants where the average effect size was 0.61 SDs (range: 0.32-1.81). Conditional analysis showed that common variants partially accounted for 37 low frequency signals but that 50 (4%) were independent of the strongest common variant signal at the cis locus. Our study shows that, using the same sample sizes, whole genome sequencing has the ability to identify low frequency variants with larger effect sizes than those observed for common variants, but that these low-frequency large effect signals may represent less than 5% of associations.

36

Deep whole-genome sequencing in pedigrees to quantify the contribution of private variants to type 2 diabetes and related metabolic traits. *G. Jun[1], M. Almeida[2], A. Manning[3], T. Teslovich[1], A. Wood[4], M. Zawistowski[1], S. Won[7], C. Fuchsberger[1], S. Feng[1], K. Gaulton[5], P. Cingolani[6], T. Frayling[4], G. Abecasis[1], J. Blangero[2], T2D-GENES Consortium.* 1) Dept Biostatistics, Univ Michigan, Ann Arbor, Ann Arbor, MI; 2) Texas Biomedical Research Institute, San Antonio, CA; 3) Broad Institute, Boston, MA; 4) University of Exeter, Exeter, UK; 5) University of Oxford, Oxford, UK; 6) McGill University, Montreal, Quebec, Canada; 7) Chungang University, Seoul, South Korea.

Sequencing in pedigrees can be a more effective approach to the analysis of rare and private variants than sequencing unrelated individuals. In T2D-GENES Project 2, we deeply sequenced whole genomes of 590 individuals from 20 large Mexican American pedigrees and imputed the identified sequence variants into additional 448 family members to better understand the role of rare and private variants in type 2 diabetes (T2D) risk and variability in T2D-related traits. Novel (not seen in 1000G or ESP), pedigree-specific singletons in founders are observed on average 2.9 (max. 39) times in the sequenced and imputed data; 16.9% are observed ≥5 times. To estimate our ability to detect rare variants affecting T2D-related quantitative traits in the given pedigree structure, we simulated singletons in founders that are transmitted into pedigrees, with various effect sizes. For a variant present in a single founder (0.125% frequency), power ranges from 5% when transmitted 5 times up to 43% and 93% when transmitted 10 and 20 times at a level of significance $p<5\times10^{-8}$. Assuming >100 variants with a frequency of 0.125% and an effect size of >2 standard deviations (SDs), accounting for >49% of variance, we have >96% power to detect at least one variant, with an expectation of 3.1 variants detected. Our genome-wide analyses identified no variant at $p<5\times10^{-8}$ with T2D-related traits, indicating that the number of variants with >2SD effect size would be smaller than 100. When focused on functional candidates and previously identified GWAS regions, we found enriched signals by rare private variants at lower significance levels ($10^{-3}<p<10^{-5}$). For example, the strongest associations for fasting glucose are driven by rare variants private to a family across a 13Mb region on chromosome 8, which overlap with a leading signal identified by GWAS (PPP1R3B). We also analyzed 30,567 gene expressions and found 21 transcripts associated with rare variants at $p<10^{-10}$. For 6 transcrips, the most significant association is to a pedigree-specific novel variant with mean effect size 2.6SDs, while the average for all expression transcripts is 0.78SDs. Using gene expression data as proof of principle, we show that deep whole genome sequencing in large pedigrees can identify private and rare associations that would not otherwise be feasible. Results are consistent with a genetic architecture where there are <100 variants with freq. ≥0.125% that influence diabetes related traits by >2SDs.

37

Inflammatory bowel disease associated genetic variants preferentially alter gene expression in neutrophil granulocytes. *H. Westra[1], D. Arends[2], T. Esko[3], M.J. Peters[4,5], R. Weersma[1], A. Metspalu[3], A.G. Uitterlinden[4,5,6], J. van Meurs[4,5], R. Jansen[2], L. Franke[1].* 1) Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, Netherlands; 2) Groningen Bioinformatics Centre, University of Groningen, Groningen, the Netherlands; 3) Estonian Genome Center, University of Tartu, Riia 23b, 51010, Tartu, Estonia; 4) Department of Internal Medicine, Erasmus Medical Centre Rotterdam, the Netherlands; 5) The Netherlands Genomics Initiative-sponsored Netherlands Consortium for Healthy Aging (NGI-NCHA), Leiden/Rotterdam, the Netherlands; 6) Department of Epidemiology, Erasmus Medical Center Rotterdam, the Netherlands.

Crohn's disease (CD) and ulcerative colitis (UC) are the two main forms of inflammatory bowel disease (IBD). IBD are immune-mediated chronic, relapsing intestinal inflammatory diseases. The most recent genome-wide association study in IBD has identified 163 genomic variants (SNPs) associated with the disease. We recently conducted *cis*-eQTL mapping in in whole blood of over 5,000 samples (Westra et al, Nature Genetics *in press*) and observed that 39% of these variants affect gene expression levels. However, it is still unclear what the primary cell type is in which these genes are expressed, because blood is a compound tissue that consists of many different cell-types. As eQTLs are often cell-type specific, identification of the cell-types where these eQTLs operate might help to identify the cell-type that is driving IBD. Although for some blood cell-types (e.g. monocytes, B-cells and CD4+ and CD8+ T-cells) eQTL studies in sample sizes of approximately 250 samples have now been presented, eQTL datasets for many other blood cell-types remain to be generated. However, generation of such data for e.g. neutrophil granulocytes is challenging: although they reflect a substantial proportion of circulating blood cells, they are difficult to purify and have a short lifespan after purification. In order to overcome these issues, we developed a new statistical method that uses whole peripheral blood eQTL data and individual blood cell count data to determine the cell-type specificity of *cis*-eQTLs. We applied this method to three independent peripheral blood datasets (n = 2,200) and observed that at least 5% of the cis-eQTLs that are detectable in whole peripheral blood are neutrophil specific. We then performed an enrichment analysis comparing unlinked ($r2 < 0.2$) IBD associated *cis*-eQTL SNPs to all other unlinked *cis*-eQTL SNPs that have been associated with complex traits and diseases, and observed that IBD *cis*-eQTL SNPs are strongly enriched for neutrophil specific cis-eQTLs (Fisher's Exact test p-value $< 1 \times 10^{-3}$). These findings reinforce the importance of neutrophils in the pathogenesis of IBD, which may provide clues for future treatment regiments. For the current study we have focused on neutrophils for this study, although our method is applicable to other cell types present in peripheral blood and other compound tissues as well. This enables cell-type specific eQTL mapping to prioritize cell types that are important in complex diseases or traits.

38

Multi-tissue eQTL and pathway analysis of genome-wide genetic association data helps uncover tissue-specific processes of complex disease. *A.V. Segrè[1], E.R. Gamazon[2], D.S. DeLuca[1], Y. Meng[1], L.D. Ward[3], T. Lappalainen[4,5], T. Flutre[6], X. Wen[7], E.T. Dermitzakis[4], M. Kellis[3], D.L. Nicolae[2], N. Cox[2], D.G. MacArthur[1], K. Ardlie[1], G. Getz[1], The GTEx Consortium.* 1) Broad Institute of Harvard and MIT, Cambridge, MA; 2) Department of Medicine, University of Chicago, Chicago, IL; 3) Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA; 4) University of Geneva, Department of Genetic Medicine and Development, Genève, Switzerland; 5) Stanford University School of Medicine, Stanford, CA; 6) Department of Human Genetics, University of Chicago, Chicago, IL; 7) Department of Biostatistics, University of Michigan, Ann Harbor, MI.

Genome-wide association studies (GWAS) have identified thousands of common variants (SNPs) associated with complex human diseases, and many more have yet to be found. GWAS interpretation faces multiple challenges, including identifying the causal genes and variants, and understanding their biological mechanisms and tissue-context. Given that pathogenic processes are tissue-specific and a lot of the association signals lie in noncoding regions, integrating genetic variation associated with gene expression changes (eQTLs) from a range of disease-relevant tissues with GWAS SNP data may be invaluable for overcoming these challenges. In support of this, NIH funded the Genotype-Tissue Expression (GTEx) project to generate a large, tissue-wide eQTL resource from ~30 human tissues per individual and hundreds of donors (final goal: 900 donors). To increase statistical and explanatory power, we developed a two-step eQTL-pathway analysis approach, and applied it to eQTLs from 9 tissues in the GTEx pilot project. The first step entails (i) testing whether a set of tissue-specific eQTLs are enriched for multiple modest to strong GWAS associations with a given complex disease or trait compared to a null distribution. If enrichment is found, (ii) genes affected by eQTLs that are top ranked based on their GWAS p-values are tested for enrichment of known biological processes, such as signaling pathways. The statistical framework includes pruning of GWAS SNPs to a set of independent SNPs, mapping eQTLs onto GWAS SNPs using linkage disequilibrium or other co-localization metrics, and evaluating enrichment with permutations. A comprehensive analysis of >10 GWAS and meta-analyses of metabolic, cardiovascular and autoimmune diseases will be presented, using cis-eQTLs from 9 tissues: whole blood, adipose, muscle, heart, artery, lung, skin, nerve and thyroid, taken from 80-185 individuals. Preliminary analysis shows enrichment of multiple modest associations amongst eQTLs in relevant tissues for some diseases, such as whole blood and autoimmune diseases, and less obvious tissue-disease connections for others. Whether eQTLs are the causal mechanisms or the altered function of the affected genes remains to be determined. In addition to proposing key pathogenic tissues for future functional studies, this eQTL-pathway method proposes potential causal genes and biological processes in specific tissues, and provides an initial step for fine-mapping of putative causal eQTL variants.

39

High-resolution Functional Analysis of eQTLs in Multiple Tissues. *X. Wen[1], R. Pique-Regi[2], T. Flutre[3], G. Moyerbrailean[2], F. Luca[2].* 1) Dept Biostatistics, Univ. of Michigan, Ann Arbor, MI; 2) Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI; 3) Dept of Human Genetics, University of Chicago, Chicago, IL.

Functional interpretation of eQTLs and their tissue specificity plays a critical role in understanding the transcriptional basis of complex traits. Most existing analyses of this kind are performed within a single tissue type and do not use fine-resolution maps of tissue-specific regulatory sequences. Recently, NIH GTEx and ENCODE projects provide valuable resources to enable the high-resolution functional study of eQTLs across multiple tissues. In this project, we develop novel statistical methods to annotate multiple-tissue eQTLs combining GTEx data from five tissues (blood, skeletal muscle, lung, skin and artery) with ENCODE data from 154 related cell-types.

Our methods hold two distinct advantages over the existing approaches: 1. We extend multiple-tissue eQTL mapping method by Flutre et al 2013 to fine-map potential multiple eQTL signals across multiple tissues simultaneously. This approach not only has high power and specificity in identifying eQTLs and their tissue specificity, but also naturally accounts for LD among candidate SNPs at any given locus. 2. We use tissue-specific single base-pair resolution annotations of genetic variants that are predicted to disrupt transcription factor (TF) binding events. These annotations are obtained by analyzing ENCODE DNase-seq footprints using the CENTIPEDE method (Pique-Regi et al 2011, Degner et al 2012) and a novel extension that learns a new sequence motif model and assesses the potential effect of a genetic variant.

Our preliminary results show that 1. Many genes have multiple cis-eQTLs. We identify examples of eQTLs, for which SNP-by-SNP analysis shows significant opposite effects in a pair of tissues. Through our multiple-SNP multiple-tissue analysis, we are able to convincingly demonstrate that this phenomenon is consistent with the scenario that two distinct tissue-specific eQTLs (active in different tissues) presented and in partial LD. 2. eQTLs are enriched with genetic variants affecting TF bindings. This pattern is consistently observed across examined tissues with high statistical significance (p-value 1.3×10^{-7}) 3. The tissue specificity of eQTLs are statistically significantly associated with the tissue-specific TF binding activity in each individual tissue.

40

Network QTLs: A new methodology for multi-tissue eQTL discovery. *B. Iriarte, M. Kellis, L. Ward, The GTEx Consortium.* MIT, Cambridge, MA.

Single nucleotide variants (SNVs) have been associated with gene expression changes in expression quantitative trait locus (eQTL) studies. However, a variant that disrupts or creates a regulatory element can have much more complex effects on gene expression programs than those profiled in single-tissue eQTL studies. To systematically discover such 'network-level' effects leading to more complex changes between different gene expression programs, we developed a new method we call 'network QTLs'. We applied this method to the unique resource of genome-wide gene expression levels across 45 post-mortem tissue samples in a cohort of 175 individuals generated by the Genotype-Tissue Expression project (GTEx). We first learn common expression patterns across the 45 tissues, by clustering the expression vectors of 19,604 protein-coding genes. We find 310 distinct gene expression patterns, which include tissue-specific, tissue-restricted, and ubiquitous expression patterns. We found remarkable enrichments for common gene functions among genes belonging in the same gene expression cluster, with the vast majority of clusters showing distinct GO enrichments, even when they were enriched in similar tissues, suggesting that the subtle differences discovered by the clustering approach are biologically meaningful. We then searched for instances where SNVs between individuals are associated with changes in multi-tissue gene expression patterns, by using module membership as a quantitative trait. After imputing missing data, imputing genotypes, controlling for sex-specific expression, and projecting module-membership coefficients onto the main PCA axes of variation across individuals, we discover 55k independent netQTLs at FDR<0.05 after LD pruning the results, and these are associated with 16,145 target protein-coding genes. The discovered netQTLs include eQTLs discovered independently in individual tissues using MatrixQTL, as expected since netQTLs are a generalization of the classical eQTL discovery approach. However, we also discover a large number of novel QTLs whose effects are too subtle in any individual tissue, but strongly detectable and statistically significant when multiple tissues are combined. These are strongly enriched for regulatory region and regulatory motif annotations by ENCODE and the Roadmap Epigenomics Project, confirming they are biologically meaningful, and suggesting potential regulatory mechanisms for their action.

41

Identification and characteristics of common genetic variants controlling transcript isoform variation in the Framingham Heart Study. *X. Zhang[1], R. Joehanes[1,2], T. Huan[1], P. Munson[2], A. Johnson[1], D. Levy[1], C. O'Donnell[1,3].* 1) NHLBI's Framingham Heart Study, Framingham, MA; 2) Center for Information Technology, National Institutes of Health, Bethesda, MD; 3) Division of Cardiology, Massachusetts General Hospital, Boston, MA.

Introduction: Alternative splicing (AS) affects ~80% of human pre-mRNAs. Splicing quantitative trait loci (sQTLs) contribute to phenotypic differences among individuals and may have important roles in disease susceptibility. However, the available evidence from sQTL studies is derived from small sample sizes largely from Epstein-Barr virus transformed lymphoblastoid cell lines (LCLs) which may not represent *in vivo* environments. We performed genome-wide screening to identify SNPs controlling splicing in whole blood collected from the community-based Framingham Heart Study (FHS). **Methods:** 5,626 FHS participants were included. Total RNA was isolated from PAXgene whole blood samples. Expression levels of 17,873 genes and 283,805 exons were measured using Affymetrix Human Exon 1.0 ST arrays. Common SNPs (MAF ≥0.01) within a 50kb region flanking either side of the gene were selected from a 1000 genomes imputed SNP dataset. A *cis*-sQTL analysis for each of the genes and exons was conducted using an additive regression model adjusted for age, sex and family structure. Bonferroni-correction was used to account for multiple testing. **Results:** On average there were 553 SNPs located within 50kb of each gene, and 16 exons per gene. 6,137 genes were significantly associated with at least one *cis* SNP (P < 5.2e-09) at the whole gene level. For 2,865 genes without association with any SNP at the gene level, at least one of the exons harbor one *cis*-sQTL at P < 2.8e-10. There were a total of 672,846 *cis*-sQTLs, suggesting that a large proportion of genes are affected by splicing regulatory variants. ~70% of these sQTLs are in intergenic or intronic and 5% are near a 5' promoter region. We identified 32 sQTLs in a 5' splice donor or 3' splice acceptor site. When we examined the NHGRI GWAS catalog, 602 unique sQTLs are found to be associated with 263 disease traits, indicating that many disease variants might affect pre-mRNA splicing. Furthermore, 40% of these 2,865 genes with specific *cis*-sQTL associations overlap with known AS events. These 2,865 genes are enriched for RNA-binding and alternative splicing GO terms. **Conclusion:** Many strong common cis-acting regulatory variants affect the splicing patterns of genes in a large population. Many genes show significant genetically controlled differences in splice-site usage. Our study provides a splicing-QTL catalog for researchers to discover functional sQTLs that control transcript isoforms implicated in common diseases.

42

Identification of a Sjögren's syndrome-associated variant that influences *OAS1* isoform switching. *H. Li[1,2], J.A. Ice[1], J.A. Kelly[1], I. Adrianto[1], S.B. Glenn[1], K.S. Hefner[3], E. Vista[4], D.U. Stone[2], R. Gopalakrishnan[5], G.D. Houston[2], D.M. Lewis[2], M. Rohrer[5], P. Hughes[5], J.B. Harley[6], C.G. Montgomery[1], J. Chodosh[7], J.A. Lessard[8], J. Anaya[9], B.M. Segal[10], N.L. Rhodus[5], L. Radfar[2], R.H. Scofield[1], C.J. Lessard[1,2], K.L. Sivils[1].* 1) Oklahoma Medical Research Foundation, Oklahoma City, OK; 2) University of Oklahoma Health Sciences Center, Oklahoma City, OK; 3) Hefner Eye Care and Optical Center, Oklahoma City, OK; 4) University of Santo Tomas Hospital, Manila, Philippines; 5) University of Minnesota, Minneapolis, MN; 6) Cincinnati Children's Hospital Medical Center and the Department of Veterans Affairs Medical Center, Cincinnati, OH; 7) Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA; 8) Valley Bone and Joint Clinic, Grand Forks, ND; 9) Universidad del Rosario, Bogotá, Colombia; 10) Hennepin County Medical Center, Minneapolis, MN.

Sjögren's syndrome (SS) is a common, progressive autoimmune exocrinopathy characterized by symptoms of dry eyes and mouth. Our previous gene expression profiling (GEP) study using peripheral blood in 180 SS cases and 73 controls identified dysregulation of interferon (IFN) pathways in SS. Notably, *OAS1*, an IFN-inducible gene involved in the inhibition of virus replication, was overexpressed in SS patients. Through integration of GEP and genome-wide association study (GWAS) data, we identified multiple *cis*-expression quantitative trait loci (eQTL) in *OAS1*. The most significant eQTL in *OAS1* is a splice site variant, rs10774671, located at the intersection between intron 5 and exon 6. Support for association of rs10774671 with SS was observed in our GWAS dataset with $P=6\times10^{-3}$. Our objective was to further evaluate the association of this eQTL with SS and characterize its functional mechanism. We tested association of rs10774671 with SS in an independent set of samples consisting of 648 cases and 2927 controls of European ancestry. Meta-analysis was then performed using a weighted Z score that generated $P_{meta}=9\times10^{-6}$, with the A allele conferring risk. To determine the functional impact of rs10774671, we evaluated the alternative splicing events in *OAS1* using both GEP and RNA-sequencing (RNA-seq) data. In GEP, one probe specifically recognizes a truncated form of *OAS1* (p42). RNA-seq experiments were performed in 57 SS cases and 27 healthy controls on the Illumina platform. After quality control, *OAS1* transcripts were reconstructed using Cufflinks and the relative abundance of each isoform was compared across samples with different rs10774671 genotypes. Both GEP and RNA-seq showed that the risk allele A, which demolishes the splicing consensus sequence, was correlated with higher expression of p42 ($P_{micro}=2\times10^{-16}$ and $P_{seq}=1\times10^{-15}$). RNA-seq results also showed correlation of the A allele with higher proportions of p48 and p44 isoforms, but lower expression of the functionally normal isoform, p46. By using multiple genetic and genomic tools, we identified a SS-associated variant in *OAS1* that switches the primary isoform from p46 to p42. Interestingly, p42 has been reported to result in a decreased OAS1 enzyme activity. This IFN-inducible gene is involved in viral RNA degradation and the inhibition of virus replication. These results indicate the risk allele in rs10774671 may cause vulnerability to virus infection that contributes to SS susceptibility.

43

Heritability and familial risk of cancer: an update from the Nordic Twin Registry of Cancer (NorTwinCan). *L.A. Mucci[1], J. Kaprio[2], J. Harris[3], K. Czene[4], P. Kraft[1], T. Scheike[5], R. Graff[1], I. Brandt[3], N. Holmes[5], D. Havelick[1], M. Hartman[4], K. Penney[1], E. Pukkala[6], G. Parmigiani[7], A. Skytthe[5], H.O. Adami[1], J. Hjelmborg[5], Nordic Twin Study of Cancer (NorTwinCan).* 1) Epidemiology, Harvard School of Public Health, Boston, MA; 2) Finnish Twin Cohort Study, Department of Public Health, University of Helsinki, Helsinki, Finland; 3) Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway; 4) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 5) Department of Biostatistics, University of Southern Denmark, Odense, Denmark; 6) Finish Cancer Registry, Helsinki, Finland; 7) Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute, Boston MA.

Background. The landmark study by Lichtenstein *et al* in 2000 estimated cancer heritability using unique twin cohorts from the Nordic countries. We updated analyses within the Nordic Twin Registry of Cancer (NorTwinCan), expanding the original cohort, adding 10 years of follow-up and increasing number of cancer cases 3-fold to provide precise estimates of heritability and familial cancer risk. **Methods**. NorTwinCan includes 133,689 monozygotic (MZ) and dizygotic (DZ) twin pairs (N=267,378 total) from nationwide registries in Denmark, Finland, Norway and Sweden. We incorporate novel time-to-event analyses to estimate the concordance risk and heritability (95%; Confidence Intervals, CI) for 23 unique malignancies, with follow-up from cancer registration and accounting for censoring and competing risks of death through 2009. **Results**. During a median 40 year follow-up, 29,599 cancer cases were diagnosed. The heritability (95% CI) for prostate cancer was 58% (52-63%), the highest of any malignancy. The risk of prostate cancer in a twin given his cotwin also had prostate cancer (concordance risk) was 32%; in MZ and 16%; in DZ twins. For breast cancer among women, the difference in concordance between MZ (29%) and DZ (21%) twins was smaller, and heritability was 28%; (12-52%). There were notable differences in findings for colon and rectal cancer: the data supported a significant genetic component to colon cancer (16%, 2-63%) but not for rectal cancer. For testicular cancer, with a cumulative incidence of 0.4% in the population, the concordance risk was substantial amongst MZ (23%) and DZ (11%) twins, with heritability of 36% (2-95%). Estimates of heritability (95% CI) were also significant for kidney (23%, 11-42%), lung (25%, 12-44%), melanoma (39%, 8-81%), ovarian cancer (28%, 15-47%), stomach (24%, 5-65%), and uterine cancer (24%, 14-87%). **Discussion**. This expanded analysis in NorTwinCan provides more precise estimates of familial risk and heritability in cancer. Estimates of heritability for prostate cancer are even greater than previously estimated. For rare cancers such as testicular, the concordance risk was substantial and provides an accurate estimate for familial risk prediction. Twin studies can also provide context for genome wide association studies.

44

A comprehensive genetic analysis of common cancer risk through the development of the Oncochip. *C.I. Amos[1], A.C. Antoniou[2], A. Berchuck[3], G. Chenevix-Trench[4], F.J. Couch[5], R.A. Eeles[6], L.J. Esserman[7], S.A. Gayther[8], C.L. Goh[6], D.E. Goldgar[9], S.B. Gruber[8], C.A. Haiman[8], P. Hall[10], D.J. Hunter[11], Z. Kote-Jarai[6], P.K. Lepage[12], S. Lindstrom[11], J. McKay[13], R.L. Milne[14], U. Peters[15], P.D. Pharoah[2], C.M. Phelan[16], F.R. Schumacher[8], T.A. Sellers[16], J. Simard[17], Z. Wang[12], D. Seminara[12], S.J. Chanock[12], D.F. Easton[2], B.E. Henderson[8].* 1) Community and Family Medicine, Geisel School of Medicine at Dartmouth, Lebanon, NH; 2) University of Cambridge, Cambridge, UK; 3) Duke University, Durham, NC; 4) The Queensland Institute of Medical Research, Herston, Australia; 5) Mayo Clinic, Rochester, MN; 6) The Institute of Cancer Research, Sutton, UK; 7) University of California, San Francisco, CA; 8) University of Southern California, Los Angeles, CA; 9) University of Utah, Salt Lake City, UT; 10) Karolinska Institutet, Stockholm, Sweden; 11) Harvard School of Public Health, Boston, MA; 12) National Cancer Institute, Bethesda, MD; 13) International Agency for Research on Cancer, Lyon, France; 14) Spanish National Cancer Research Centre, Madrid, Spain; 15) Fred Hutchinson Cancer Research Insitute, Seattle, WA; 16) Moffitt Cancer Center, Tampa, FL; 17) CHU de Québec Research Center, Québec, Canada.

Identifying the genetic determinants of complex diseases through genome-wide association studies (GWAS) has been hampered by application of variety of different genotyping platforms, on often limited collections of cases and controls. In particular, evaluating the role that cancer subtype heterogeneity, gene-gene and gene-environment interactions have upon cancer susceptibility and subsequent risk profiling requires large data sets. To facilitate a more comprehensive evaluation of susceptibility to common cancers we formed a consortium comprising investigators from the Genetic Associations and Mechanisms of Oncology (GAME-ON), Personalised Risk Stratification for Prevention and Early Detection of Breast Cancer, The Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA), the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL), the Breast Cancer Case Control Consortium (BCAC), the Breast and Prostate Cancer Cohort Consortium (BPC3), Ovarian Cancer case control consortium (OCAC) and ATHENA Breast Health Network and the NIEHS Sisters studies. By bringing together multiple consortia, we have been able to design a custom Illumina array that incorporates approximately 600,000 markers and which will be genotyped on at least 400,000 samples, including patients affected with breast, prostate, ovarian, lung and colon cancers, respective controls, and individuals with BRCA1 or BRCA2 mutations. The array incorporates fine-scale mapping of more than 150 known cancer susceptibility regions, variants showing evidence for association from previous GWAS, putative risk variants identified through exome or whole genome sequencing, and other functional candidate variants potentially associated with disease risk. The array will include a backbone of 260K markers selected for high genome-wide coverage of common variants based on the 1000 genomes project, providing the basis for discovery of new associations for any cancer or associated trait. Additionally we selected markers associated with pharmacological outcomes, gene expression in multiple tissues, and traits (e.g. smoking behavior, menarche, BMI) that are associated with cancer risk. When completed, this analysis will be one of the largest conducted using a common custom genotyping array, and will allow comprehensive evaluation of the genetic architecture of multiple common cancer types.

45

Detection of Large Clonal Mosaic Events in Existing Genome-wide Association Study Data. *M. Machiela[1], W. Zhou[1,2], M. Yeager[1,2], K. Jacobs[1,2], S. Berndt[1], Q. Lan[1], N. Rothman[1,2], S. Savage[1], P. Taylor[1], N. Caporaso[1], I. DeVivo[3], R. Hoover[1], M. Tucker[1], S. Chanock[1].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 2) Cancer Genomics Research, SAIC-Frederick, Inc., NCI-Frederick, Frederick, MD; 3) Harvard School of Public Health, Boston, MA; Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115.

Mosaicism is defined as the presence of two or more genetically distinct cellular populations in an individual who developed from a single zygote. With the advent of genome-wide association studies, it is possible to evaluate the relative intensities of the genotype signals to detect large structural mosaicism in human populations. We performed an analysis of germline DNA derived from blood or buccal specimens from 24,849 individuals in 46 cancer-related studies using a modified mosaic alteration detection algorithm on renormalized B-allele frequencies and log2 relative probe intensity ratios from commercially available Illumina SNP arrays. Overall, we confirmed our previously reported findings (Jacobs et al Nat Gen 2012) including a similar distribution of types of events with respect to chromosomal position, an increase with age (overall and in cancer free-controls) and an increase in men versus women. Our segmentation algorithm detected 341 clonal mosaic events >2 Mb in size in 168 individuals, 91 in cases and 77 in controls. Among individuals with mosaicism, 27 had two or more events, including one individual with 34 events. The detected mosaic events include 163 (47.8%) copy-neutral losses of heterozygosity, 90 (26.4%) copy losses, 69 (20.2%) copy gains, and 19 (5.6%) complex events. The most common events detected were copy neutral losses of heterozygosity on chromosomes 4q and 9q and copy losses on 13q14 and 20q. The 13q14 deletion, commonly observed in chronic lymphocytic leukemia cases, was observed in our dataset among non-hematologic cancer cases suggesting a potential role of 13q14 deletions in these malignancies. A breakpoint analysis of the 13q14 deletion revealed that breakpoints clustered in genomic regions enriched with DNaseI peaks, transcription factors, and common repeat elements. An unadjusted analysis of age indicates the mosaicism increases with increasing age (p=5.3e-4), providing further support for age-related genomic senescence. Combined with independent data from our previous study of 57,835 individuals, analyses indicate mosaicism is associated with male sex with 353 of 37,287 males (0.95%) and 191 of 33,223 females (0.57%) having mosaic events (p=0.001). These results suggest that human mosaicism may be more common in the general human population than previously expected; underscore the importance of thorough characterization of germline DNA; and potentially represent the tip of the iceberg for smaller mosaic events.

46

Most common 'sporadic' cancers have a substantial germline genetic component. *S. Macgregor[1], W.E. Ek[1], Y. Lu[1], D. Whiteman[1], T.L. Vaughan[2], A.B. Spurdle[1], D.F. Easton[3], P.D. Pharoah[3], D.J. Thompson[4], A.M. Dunning[5], A.K. Hayward[1], G. Chenevix-Trench[1], Q-MEGA and AMFS Investigators, ANECS-SEARCH, UKOPS-SEARCH, BEACON consortium.* 1) Queensland Inst Med Res, Brisbane, Australia; 2) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; 3) Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK; Department of Oncology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK; 4) Department of Public Health and Primary Care, University of Cambridge, UK; 5) Department of Oncology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK.

Common cancers are frequently demarcated into 'hereditary' or 'sporadic' ('non-hereditary') types. Such distinctions initially arose from work identifying rare highly penetrant germline mutations causing 'hereditary' cancer. While rare mutations are important in particular families, most cases in the general population are 'sporadic'. Twin studies have suggested that many 'sporadic' cancers show little or no heritability. Our objective is to quantify the role of common germline genetic variants in cancer susceptibility. We apply a method for estimating the importance of common genetic variants (single nucleotide polymorphism, SNP, heritability, h2g) to several cancer genome-wide association studies (GWASs). Population samples of primarily sporadic cases (>1000 per cancer) and controls were obtained for the following cancers; bladder, breast, endometrial, esophageal, gastric, kidney, lung, melanoma, ovary, pancreatic and prostate.

The following cancers showed a significant (P<0.05) SNP heritability: melanoma USA set h2g =0.19 (standard error 0.09) and Australian set h2g = 0.30 (standard error 0.10); pancreatic h2g =0.18 (0.06); prostate h2g = 0.81 (0.25); kidney h2g =0.17 (0.07); ovarian h2g =0.30 (0.06); esophageal adenocarcinoma h2g =0.25 (0.05); esophageal squamous cell carcinoma h2g=0.19 (0.05); gastric h2g =0.11 (0.07); endometrial UK set h2g =0.23 (0.11) and Australian set h2g =0.39 (0.19). Two further cancers showed a positive but non-significant effect: breast h2g =0.13 (0.22); lung h2g =0.09 (0.06). One cancer showed a small effect: bladder h2g =0.01 (0.04). Amongst these cancers, previous twin studies were only able to clearly show heritability for prostate and breast cancer but we can now make much stronger statements for several common cancers which emphasize the important role of genetic variants in cancer susceptibility. As SNP heritability is only estimated from common SNPs, it provides a lower bound on the overall genetic component (heritability). In analyses with previously identified loci removed, in most instances a significant polygenic component remained. We have demonstrated that several 'sporadic' cancers have a substantial inherited component. Larger GWASs in these cancers will continue to find more loci which explain part of the remaining polygenic component. For most cancers examined here, the descriptor 'sporadic' or 'non-hereditary' should be replaced by 'polygenic'.

47

Meta-analysis of genome-wide association studies in 125,000 women identifies fourteen new breast cancer susceptibility loci. *K. Michailidou[1], P. Hall[2], P. Kraft[3], D.F Easton[1], Breast Cancer Association Consortium, DRIVE.* 1) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 2) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 3) Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA.

Genome Wide Association Studies (GWAS) and large scale replication studies have successfully identified 77 common variants associated with breast cancer. These variants explain ~15% of the familial risk of the disease. In an effort to identify new susceptibility loci, we performed a meta-analysis of 11 GWAS consisting of ~16K breast cancer cases and ~19K controls, and ~47K cases and ~43K controls from 41 studies genotyped on a 200K custom array (iCOGS). Analyses were restricted to women of European ancestry. Each study was imputed separately using the March 2012 release of the 1000 Genomes reference panel. Summary per-allele odds ratio estimates and standard errors were obtained by inverse variance fixed effect meta-analysis.

Approximately 11M SNPs were reliably imputed, with imputation r^2>0.3 and minor allele frequency (MAF)>0.005, in at least one of the studies. We identified more than 300 variants in 28 regions reaching $p<5\times10^{-8}$ not previously reported as being associated with breast cancer. In 13 of the regions, the most significant SNP has an imputation r^2>0.8 (in iCOGS), and supporting evidence was provided by genotyped SNPs in the same region. Three further regions lie within 1Mb of the top hit of regions previously found to be associated with the disease; these may reflect confounding with previous signals or secondary associations. In 3 of the regions the signal was provided by only one study and a further region from only three studies; in each case the MAF of the most significant SNP was <0.03. Finally, for 8 regions there was only one SNP in the region that reached the GWAS significance threshold, and/or the best SNP had imputation r^2<0.8. Further evaluation will be needed to determine whether these reflect true associations or imputation artefacts.

48

Identification of seven new ovarian cancer susceptibility loci; risk of ovarian cancer in the general population and in BRCA1 and BRCA2 carriers. *K. Kuchenbaecker[1], J. Tyrer[1], F.J. Couch[2], S.A. Gayther[3], P.P. Pharoah[1], S.J. Ramus[3], A.C. Antoniou[1], G. Chenevix-Trench[4], CIMBA and OCAC.* 1) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 2) Department of Laboratory Medicine and Pathology, and Health Sciences Research, Mayo Clinic, Rochester, MN, USA; 3) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, California, USA; 4) Department of Genetics and Computational Biology, Queensland Institute of Medical Research, Brisbane, Australia.

Twelve ovarian cancer susceptibility loci have been identified through genome wide association studies (GWAS), including a single nucleotide polymorphisms (SNP) at 4q32.3 which only appears to be associated with risk in *BRCA1* mutation carriers, but not in the general population or in *BRCA2* mutation carriers. Using data from two consortia participating in the Collaborative Oncological Gene-environment Study (COGS) we have carried out meta-analyses for ovarian cancer risk in the general population (n=15,437 invasive, including 9,627 serous, ovarian cancer cases and 30,845 controls) and in *BRCA1* (n=2,462 affected and 12,790 unaffected) and *BRCA2* (n=631 affected and 7,580 unaffected) mutation carriers using genotypes from the iCOGS chip and imputed data on 11.4M SNPs based on 1000 Genomes data as the reference panel. We assessed associations with ovarian cancer risk in mutation carriers with a retrospective likelihood approach, and in the unselected population by logistic regression. We identified (P< 5×10-8) three new loci for risk of invasive ovarian cancer in carriers and the general population (rs3820282 in WNT4 at 1p36.23, 9:136149709 in ABO at 9q34.2, rs8044477 in GOT2 at 16q21), three new loci for risk of ovarian cancer in carriers and serous invasive cancer in the general population (rs12039431 in RSPO1 at 1p34.3, rs17329882 in SYNPO2 at 4q26, rs115344852 in GPX5 at 6p22.1), and one new locus (17:29181220 in ATAD5 at 17q11.2) for risk of ovarian cancer in the general population and *BRCA2* mutation carriers (OR=1.11, 95%CI: 1.08-1.14) for which there was no evidence of association in *BRCA1* mutation carriers (OR=0.98, 95%CI: 0.91-1.05). Although the target of these associations will not be known until fine mapping and functional analyses have been done, it is interesting that WNT4 is a GWAS hit for endometriosis, and RSPO1 regulates Wnt4 expression in the ovary. The ABO blood group has previously been reported to be associated with risk of ovarian cancer, so is likely to be the target of the association with 9:136149709. These results demonstrate the, mostly, shared genetic basis of ovarian cancer susceptibility in the general population and in *BRCA1/2* carriers and illustrate the benefits of combining data from different cancer consortia. The identification of multiple loci modifying ovarian cancer risk may be potentially useful for counselling women with *BRCA1* and *BRCA2* mutations about their risk of ovarian cancer.

49

Breast cancer risk and survival at the 2q35 locus are mediated through chromatin looping and regulation of IGFBP5 expression. *M. Ghoussaini[1], S.L. Edwards[2], J.D. French[2], K. Michailidou[3], S. Nord[4], J. Dennis[3], Q. Guo[3], M.K. Schmidt[5], K. Hillman[2], S. Kaufmann[2], E. Dicks[3], S. Ahmed[1], M. Maranian[1], C.S. Healey[1], C. Baynes[1], C. Luccarini[1], M. Bolla[1], J. Wang[1], V.N. Kristensen[4], P.D.P. Pharoah[1,3], G. Chenevix-Trench[2], D.F. Easton[1,3], A.M. Dunning[1], The Breast Cancer Association Consortium.* 1) Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK; 2) Department of Genetics, Queensland Institute of Medical Research, Brisbane, Australia; 3) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 4) Department of Genetics, University of Oslo, Oslo, Norway; 5) The Netherlands Cancer Institute, Division of Molecular Pathology, Amsterdam, the Netherlands.

Genome wide association studies (GWAS) have previously identified a breast cancer susceptibility locus on 2q35 (rs13387042). In an effort to identify the likely causal variant(s), we performed fine-scale mapping of this region by genotyping 276 SNPs across a 210 Kb region in 89,050 subjects of European ancestry and 12,893 subjects of Asian ancestry from 50 case-control studies in the Breast Cancer Association Consortium (BCAC). Genotypes for 1,284 SNPs were imputed using the 1000 genomes project as a reference. The most strongly associated variant in Europeans was rs4442975 (OR =0.86; 95% CI [0.84-0.88]; p=2.3×10-40); the association was confined to estrogen receptor positive disease. All but two perfectly correlated SNPs (rs4442975 and rs6721996) could be excluded as the likely causal variant (likelihood ratio more than 100:1). Both variants were much rarer in Asians than in Europeans (13% versus 49%); the association in Asians was in the same direction as in Europeans but not statistically significant. Follow-up experiments found no evidence for functional activity for rs6721996. SNP rs4442975 lies just outside an enhancer mark in breast cancer cell lines; using chromatin conformation studies, we identified a long-range physical interaction between the enhancer/flanking sequence encompassing rs4442975 and the IGFBP5 promoter located more than 350 Kb away. Expression analysis show that the rare allele of rs4442975 is associated with increased IGFBP5 levels in normal breast tissue (n=123, p=0.04). Consistent with this, we found that the rare allele of rs4442975 is associated with increased chromatin looping and interaction with the IGFBP5 promoter. The rare allele was also associated with poorer breast cancer survival in a meta-analysis including the BCAC dataset and nine GWAS (p=7×10-3). IGFBP5 plays an important role in breast cancer cell growth both in vivo and in vitro. These results suggest strongly that the 2q35 susceptibility locus is mediated through regulation of IGFBP5.

50

Expression quantitative trait loci analysis in breast cancer tumor and normal adjacent FFPE specimens from the Nurses' Health Study. *A. Quiroz-Zarate[1], B.J. Harshfield[2], N. Knoblauch[3], S. Christe[3], R. Hu[2], D.J. Hunter[2,4], A.H. Beck[3], R.M. Tamimi[2,4], J. Quackenbush[1,4], A. Hazra[2,4], U19/GAME-ON DRIVE Consortium.* 1) Center for Computational Biology, Dana Farber Cancer Institute, Boston, MA; 2) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 3) Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA; 4) Harvard School of Public Health, Boston, MA.

Genome-wide association studies (GWAS) of breast cancer have identified 70 risk alleles. To identify regulatory variants we conducted expression quantitative trait loci (eQTL) analysis in clinically applicable formalin-fixed paraffin embedded (FFPE) breast tissue to identify potential regulatory effects of those risk alleles. We identified 450 invasive breast cancer cases in the Nurses' Health Study (NHS) diagnosed from 1990-2004 with GWAS data and sufficient RNA for expression profiling in paired breast tumor and normal adjacent breast tissue. RNA was extracted using the Qiagen AllPrep kit, amplified using the NuGen Ovation FFPE WTA kit, and profiled using the Affymetrix Human Transcriptome Array (HTA 3.0v1). The HTA includes 6,892,960 features for measuring gene expression, alternative splicing, coding SNPs, and noncoding transcripts. After filtering, we had data from 45,560 probes sets representing over 17,000 genes. 70 risk alleles were included in the linear regression model as dosages and analyzed to identify associations with gene expression. Analyses were conducted separately for tumor and normal adjacent samples and performed for breast cancer as a single disease as well as for five intrinsic molecular subtypes. We further evaluated the stepwise effect of batch-to-batch variation, patient's age at diagnosis, and year of diagnosis and did not observe an effect of age or year of diagnosis at a genome-wide level. We identified SNPs associated with expression levels using a genome-wide model adjusted for batch-to-batch variation in tumor and normal tissue (stratified analysis). Finally, we performed an innovative functional QTL (fQTL) analysis to gain additional functional insight into genetic variants important in breast cancer. In the fQTL analysis we tested for the association between SNPs and the expression of gene functional classes and pathways, evaluating the hypothesis that SNPs may also be associated with regulation of processes in addition to individual genes. Here we will present significant association with the goal of understanding functional changes in genetically-regulated cellular processes that occur during the development and progression of breast cancer. Our data show that susceptibility eQTLs can be identified in archival samples. Identification of gene transcripts in FFPE tissue that are affected by breast cancer loci is critical in understanding the mechanism by which these variants affect risk and mediate disease processes.

51

Finding Genes in Animal Models of Histiocytic Sarcoma. *E.A. Ostrander[1], J.L. Rowell[1], G.R. Rutteman[2], H.G. Parker[1].* 1) Cancer Genetics Branch, Bldg 50, NHGRI/NIH, Bethesda, MD; 2) Department of Clinical Sciences of Companion Animals, Faculty of Veterinary Medicine, Utrecht University, The Netherlands.

Histiocytic sarcoma (HS) is a rare but extremely aggressive cancer arising from antigen-processing phagocytes (histiocytes). Though the disease can appear at any age, ~15% of the diagnoses published in the last ten years were in children. HS does not respond well to treatment and nearly 50% of those diagnosed die of the disease. This number approaches 100% for those presenting with disseminated disease. Because the human disease is poorly defined, and may furthermore represent multiple related disorders, a reliable model system would inform the human condition and aid in the development of treatment and diagnostics. Dogs offer such an opportunity. Much like the human disease, canine HS is a highly aggressive and lethal dendritic cell neoplasm with survival times of 2-15 months. Commonly, HS is divided into two subtypes: localized and disseminated. Recent work has questioned whether the two subtypes of HS represent the same disease, are two different diseases, or operate on a continuum (disseminated being a metastatic version of localized). While HS is relatively uncommon among dog breeds, it is present at high frequency within Flat-Coated Retrievers (FCR). Here, we performed a genomewide association analysis on 204 FCR samples. Strikingly, we have discovered one locus associated with disseminated HS and a second novel locus for localized HS in FCR. For disseminated HS, the most prominent signal was located on canine chromosome 7 within a 200kb region highly conserved between human and dog, but with low conservation with mouse and chicken. We also identified a signal for localized HS on canine chromosome 5, interestingly demonstrating two peeks; one that overlaps TP53 and the second that represents a gene regulator of TP53. This work presents the first identification of two novel loci for localized and disseminated HS in FCR and will hopefully lead to clear therapeutic targets in humans and companion animals.

52

Using brain molecular QTLs to identify novel risk genes shared by multiple psychiatric diseases. *C. Liu[1, 2, 4], C. Zhang[1, 4], C. Chen[1, 2, 4], J.A. Badner[3], N. Alliey-Rodriguez[3], E.S. Gershon[3], E.R. Gamazon[5], . IOCDF-GC., . TSAICG., N.J. Cox[5].* 1) Dept Psychiatry, Univ of Illinois, Chicago, Chicago, IL; 2) State Key Laboratory of Medical Genetics, Central South University, Changsha, PR China; 3) Dept of Psychiatry, Univ of Chicago, Chicago, IL; 4) Inst. of Human Genetics, Univ of Illinois at Chicago, Chicago, IL; 5) Dept of Medicine, Univ of Chicago, Chicago, IL.

Genome-wide association studies (GWASs) have detected some common variants associated with psychiatric diseases. Many common SNPs have been found to be associated with gene expression or DNA methylation levels in human brain by quantitative trait loci (QTL) mapping. We and others have shown that SNPs associated with expression or DNA methylation were enriched in GWAS signals. Using expression QTL (eQTL) and DNA methylation QTL (mQTL) mapping results from multiple postmortem brain collections, we studied QTL SNPs ($p < 0.001$) in GWAS signals ($p<0.01$) of bipolar disorder (BD), schizophrenia (SCZ), major depression (MDD), autism (ASD), attention deficit and hyperactivity disorder (ADHD), obsessive compulsive disorder (OCD) and Tourette syndrome (TS) for their abilities explaining disease heritability. We further compared lists of eSNPs in disease GWAS signals across two or three diseases to identify putatively functional SNPs that are shared across multiple diseases. After we showed that GWAS signals of psychiatric diseases were significantly enriched with brain eQTL and mQTL SNPs, we further found that 19-50% of psychiatric disease heritability captured by GWAS could be explained by brain eQTLs. Different diseases shared different amount of expression QTL SNPs (eSNPs) in their GWAS signals. SCZ and BD shared the most comparing to other disease combinations. Cerebellum, parietal, and temporal cortex data, gene level and exon level analyses showed consistent pattern of sharing. The pattern clearly indicated a spectrum of genetic relatedness among the seven psychiatric diseases. By examining eSNPs shared by three diseases, we found one eSNP shared by SCZ, BD and MDD; and eight eSNPs shared by SZ, BD and OCD. Fifteen genes were regulation targets of these SNPs. Most of these genes were associated with mouse behavioral phenotypes that may link to psychiatric diseases. Interestingly, type II diabetes, as a control disease, has the fewest brain eSNPs in its GWAS signals than psychiatric diseases. But it still shares one eSNP with SCZ. The shared SNP is associated with expression of CACNA1C, which is a known risk genes of both SCZ and BD. Brain eQTL and mQTL data helped to identify novel functional SNPs and their target genes for multiple diseases. Different diseases share different amount of eSNPs that can be captured by GWAS. Multiple novel risk genes were identified as shared risk genes of psychiatric diseases, even non-psychiatric disease.

53

Functional enrichment analysis on genome-wide epistasis patterns reveals pathway interactions in Bipolar Disorder. *S. Prabhu, N. Clinger, B. Burnett, I. Pe'er.* Computer Science, Columbia University, New York, NY.

In the context of GWAS, pathway based functional enrichment (FE) methods aggregate marker associations (like SNPs) at functionally annotated loci (like genes), and have been extremely useful at finding biological "modules" driving disease. So far, mainstream FE approaches have focused on mining single-locus GWAS hits (i.e. marginal associations), and developing statistics that implicate a single ontology, a single pathway, or a single set of genes with phenotype. In this work, we present the first systematic and unbiased framework for genome-wide FE of interactions (i.e. epistatic associations) in case-control GWAS datasets. First, we develop a statistical framework in which SNP-SNP epistasis p-values are percolated upwards into gene-gene scores. In the subsequent step, we combine gene-gene epistasis scores across gene sets, in a search for ontology-ontology epistasis. At each stage, we either use powerful statistics that have been thoroughly characterized for type I and II errors, or rely on an extensive (and so far impractical) permutation procedure to establish empirical significance. We also describe how to account for any confounding that might arise from variable gene size, ontology size, population structure and LD between loci. Finally, we apply our method to the WTCCC Bipolar Disorder (BD) dataset (2K cases, 3K controls, 113K genic SNPs). Considering the computational burden of even a single genome-wide interaction scan, our robust permutation analysis (5000 genome-wide scans for interaction in 1 week on a small compute cluster) was only made feasible by recent computational advances by our group (SIXPAC ultrafast interaction scan, Genome Research 2012). For BD, we report an ontology interaction graph containing 17 connected components (FDR<0.01). The largest component highlights the epistatic links across, but not within, two important gene sets. The first set contains genes involved in serine/threonine kinase activity and SMAD signaling proteins: targets of Lithium, the most widely prescribed compound for BD Disorder. The second set (LDL and HDL remodeling, triglyceride lipase activities and cholesterol transport) highlights the role of lipid levels, whose role in the pathophysiology of this disorder is less well understood. Other connected components highlight epistasis involving cerebellar purkinje cell development (implicated in schizophrenia and depression) and histamine and glutamate neurotransmitter catabolism, among others.

54

Role of the Wnt signaling pathway in bipolar disorder susceptibility: Gene-set analysis of SNP-BMI interaction effects. *M.A. Simonson[1], S.J. Winham[1], G.D. Jenkins[1], A. Cuellar Barboza[2], M.H. Jang[3], S. McElroy[4], M.A.F. Frye[1, 5], J.M. Biernacka[1, 5].* 1) Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Universidad Autonoma de Nuevo Leon, Monterrey, Mexico; 3) Neurologic Surgery, and Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN; 4) Lindner Center of Hope, Mason, OH; 5) Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN.

The Wnt signaling pathway is essential for the maturation and functioning of the central nervous system and may play a role in the etiology of bipolar disorder (BD), although the precise genetic mechanisms of this relationship remain unknown. BD is associated with higher body mass index (BMI) and increased metabolic comorbidity. In a previous GWAS that examined interaction effects between BMI and potential BD risk alleles, we identified a significant SNPxBMI interaction in the gene TCF7L2, a critical member of the Wnt signaling pathway. Here we built upon those results by performing a gene-set analysis of this pathway to elucidate relevant genetic factors that influence BD susceptibility, potentially via an interaction with BMI. Using data from the Genetic Association Information Network (GAIN) genome-wide association study of BD, we performed gene-set analyses to evaluate both genetic main effects and SNP-BMI interaction effects. The analysis included 388 cases and 1,020 controls of European American ancestry with available BMI data. The gene set consisted of 3,761 SNPs in 148 genes in the pathway. Gene-set analysis was performed using Interval Based Enrichment Analysis (INRICH), and a modification of the PC-gamma method previously developed for testing SNP main effects at the pathway level. In addition to pathway-level results, the PC-gamma approach provides gene-level tests of association, and INRICH identifies specific genes that contribute to pathway enrichment for associations, which were used to evaluate both gene main effects and BMI-gene interactions. Both methods provided evidence of association for both the SNP main effects and BMI-SNP interactions at the pathway level. Additionally, both methods identified several of the same genes or gene-BMI interactions as top associations with BD. In addition to the gene TCF7L2, FZD7 was identified as having for both main and interaction effects by INRICH and the PC-gamma approach. These genes also contain SNPs previously associated with psychiatric or neurological disorders. Further functional work is needed to determine the neurobiological mechanisms underlying the identified associations with BD, particularly the identified interaction with BMI.

55

A novel variant in the HERG3 voltage-gated potassium ion channel gene (KCHN7) is associated with bipolar spectrum disorder among the Old Order Amish. *A.R. Benkert[1,2], E.G. Puffenberger[1,2], S. Markx[3], S.M. Paul[4], R.N. Jinks[2], B. Georgi[5], T. Hoshi[6], A. McDonald[3], M.B. First[3], W. Liu[4], A. Heaps[1], Y. Tian[6], A. Chakravarti[7], D.H. Morton[1,2,8], M. Bucan[5], K.A. Strauss[1,2,8].* 1) Clinic for Special Children, Strasburg, PA; 2) Franklin & Marshall College, Lancaster, PA; 3) Dept. of Psychiatry, Columbia University, New York, NY; 4) Depts. of Neuroscience, Psychiatry, and Pharmacology, Weil Cornell Medical College of Cornell University, New York, NY; 5) Dept. of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 6) Dept. of Physiology, University of Pennsylvania, Philadelphia, PA; 7) Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 8) Lancaster General Hospital, Lancaster, PA.

To identify genetic variants associated with bipolar spectrum disorder among the Pennsylvania Amish, we conducted blinded psychiatric assessments of 26 Amish subjects from four families. Potentially pathogenic variants were identified by whole exome sequencing and tested for association with psychiatric phenotypes within the larger Amish Study of Major Affective Disorder (ASMAD, N=394) (Am J Psychiatry 140:56). Fourteen of 26 Amish subjects had bipolar spectrum disorder. The only rare allele shared among all of them was a missense variant in KCNH7 c.1181G→A (p.Arg394His), which we predicted to be pathogenic based on conservation, control allele frequencies, and in silico modeling. Ten rare variants with potentially high pathogenicity were subsequently tested within the ASMAD cohort (N=394); KCNH7 c.1181G→A had the highest enrichment among individuals with bipolar spectrum disorder and the strongest family-based association with bipolar 1 (P=0.021), bipolar spectrum (P=0.031), and any major affective disorder (P=0.016). KCNH7 (HERG3/Kv11.3) wild-type and Arg394His voltage-gated potassium ion channels were studied using immunofluorescence, western blotting, and patch-clamp electrophysiology. Heterologously expressed KCNH7 Arg394His channel subunits traffic normally to the plasma membrane in neuroblastoma cells in vitro without loss of abundance relative to wild-type HERG3 subunits. However, the histidine substitution at the highly conserved N-terminal cytoplasmic arginine394 shifts voltage dependent channel activation in the positive direction, slows activation kinetics, and is predicted to increase neuronal excitability in vivo. Penetrance and severity of mental illness were similar among KCNH7 c.1181G→A heterozygotes and homozygotes. This may reflect the heterotetrameric nature of ERG channels and/or a high degree of potassium channel redundancy in the nervous system that attenuates the biological impact of modest functional abnormalities of any one channel subunit. The novel KCNH7 c.1181G→A missense variant alters the electrophysiological properties of voltage-gated HERG3/Kv11.3 potassium ion channels and may be a specific risk allele for bipolar spectrum disorder among the Pennsylvania Amish.

56

Exome sequencing in sporadic cases of schizophrenia and functional studies identify 21 putative candidate genes for schizophrenia and establish the RGS12 gene as a strong candidate. *M. Guipponi[1], F. Santoni[1], V. Setola[2], E. Oestreich[3], C. Gehrig[1], M. Rotharmel[4], M. Cuenca[4], O. Guilin[4], D. Dikeos[5], G. Papadimitriou[5], A. Méary[6,7,8,9], F. Schürhoff[6,7,8,9], S. Jamain[6,7,8,9], M. Leboyer[6,7,8,9], D. Rujescu[10], A. Pulver[11], S. Moy[12], D. Campion[4], A. Malafosse[1], D. Siderovski[2], S.E. Antonarakis[1].* 1) Dept of Genetic Medecine and Development, University of Geneva Medical School and University Hospitals of Geneva, Switzerland; 2) Dept of Physiology and Pharmacology, West Virginia University School of Medecine, USA; 3) Pacific Northwest University of Health Sciences, USA; 4) Centre Hospitalier du Rouvray, Sotteville les Rouen et INSERM U 1079, France; 5) Athens University Medical School, 1st department of Psychiatry, Eginition Hospital, Greece; 6) Inserm U955, Psychiatrie Génétique, Créteil, France; 7) Université Paris Est, Faculté de Médecine, Créteil, France; 8) Chenevier - H. Mondor, Pôle de Psychiatrie, Créteil, France; 9) Fondation Fondamental, Créteil, France; 10) Division of Molecular and Clinical Neurobiology at the University of Munich, Germany; 11) Epidemiology and Genetics Program in Psychiatry, John Hopkins School of Medicine, USA; 12) Carolina Institute for Developmental Disabilities, University of North Carolina, USA.

Schizophrenia is a severe, debilitating mental illness with a strong genetic component. Identification of the genetic factors related to schizophrenia has been challenging and its genetic architecture is still largely unknown. To evaluate the contribution of damaging de novo mutations to schizophrenia, we sequenced the exomes of 58 sporadic cases of schizophrenia and their healthy parents. Exomes were captured and sequenced using the Agilent and Illumina technologies, respectively and genetic variants were called using our analytical pipeline that integrates BWA, Samtools, Pindel2 and Annovar. We identified 55 validated de novo variants, 21 of which were predicted as deleterious including 16 missense, 2 conserved splice site, 2 nonsense mutations and 1 frameshift deletion. The observed exonic point mutation rate of 0.91 events per trio or 1.7 × 10-8 per base per generation and the non-synonymous to synonymous ratio of 2.25 did not differ from neutral expectations. We did not observe multiple independent de novo variants in the same gene in our cohort. However, when combined to the other studies (Girard et al. 2011; Xu et al. 2012), our data allowed the identification of a second proband carrying a missense de novo SNV in the RGS12 gene, a finding unlikely to have occurred by chance (Monte Carlo simulation; P<0.0001). To further explore a potential role for RGS12 in schizophrenia, homozygous Rgs12 knock-out mice were tested for sensorimotor gating capabilities as measured by prepulse inhibition of acoustic startle responses, one of the well-documented biological markers for schizophrenia in humans. Remarkably, these mice showed a significantly impaired sensorimotor gating when compared to wild-type and heterozygous mice indicating that Rgs12 may underlie certain aspects of schizophrenia pathogenesis. Overall, this study provides a list of 21 putative candidate genes for schizophrenia and establishes the RGS12 gene as a strong candidate.

57

Rare duplications in RB1CC1 are associated with schizophrenia in large datasets from Europe. *F. Degenhardt[1,2], M.A. Pfohl[1,2], L. Lennertz[4], L. Priebe[1,2], J. Strohmaier[3], S.H. Witt[3], A. Hofmann[1,2], T. Becker[5,6], R. Mössner[4], W. Maier[4,5], I. Nenadic[7], S. Meier[3], J. Buizer-Voskamp[8,9], R.A. Ophoff[10,11], D. Rujescu[12,13], I. Giegling[12,13], A. Ingason[13], M. Wagner[4], A. Meyer-Lindenberg[14], H. Walter[15], S. Moebus[16], A. Corvin[17], H. Stefánsson[18], T.G. Schulze[3,19], M. Rietschel[3], S. Cichon[1,2,20,21], M.M. Nöthen[1,2,5], GROUP Consortium, Wellcome Trust Case Control Consortium 2; International Schizophrenia Consortium.* 1) Institute of Human Genetics, University of Bonn, Bonn, Germany; 2) Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany; 3) Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim / Heidelberg University, Mannheim, Germany; 4) Department of Psychiatry, University of Bonn, Bonn, Germany; 5) German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany; 6) Institute for Medical Biometry, Informatics, and Epidemiology, University of Bonn, Bonn, Germany; 7) Department of Psychiatry and Psychotherapy, Jena University Hospital, Jena, Germany; 8) Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands; 9) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 10) Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA; 11) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience & Human Behavior, University of California Los Angeles, Los Angeles, California, USA; 12) Molecular and Clinical Neurobiology, Department of Psychiatry, Ludwig-Maximilians-University, Munich, Germany; 13) Department of Psychiatry, University of Halle-Wittenberg, Halle, Germany; 14) Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany; 15) Department of Psychiatry and Psychotherapy, Charité Campus Mitte, Berlin, Germany; 16) Institute of Medical Informatics, Biometry, and Epidemiology, University Duisburg-Essen, Essen, Germany; 17) Department of Psychiatry, Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland; 18) deCODE Genetics, IS-101 Reykjavik, Iceland; 19) Section of Psychiatric Genetics, Department of Psychiatry and Psychotherapy, University Medical Center, Georg-August-University, Göttingen, Germany; 20) Institute of Neuroscience and Medicine (INM-1), Structural and Functional Organisation of the Brain, Genomic Imaging, Research Centre Juelich, Juelich, Germany; 21) Division of Medical Genetics, University Hospital Basel and Department of Biomedicine, University of Basel, Basel, Switzerland.

Schizophrenia is a severe neuropsychiatric disorder with heritability estimates of ~80%. Recently, the first exome-sequencing studies (NGS) in patients with schizophrenia were published. They provided evidence that de novo mutations were more frequent in patients compared to controls (Girard et al., 2011, Xu et al., 2011). These results suggested that de novo mutations contribute to the risk of developing schizophrenia. However, a large number of genes were hit by de novo mutations and therefore it is not easy to pinpoint specific genes as actually being relevant for the disease process. Support for a specific gene can be provided by the identification of additional alterations - including copy number variants (CNVs) - in independent study cohorts. We screened the genome-wide SNP array data from 1,637 patients with schizophrenia and 1,627 controls for the presence of CNVs in 55 candidate genes, suggested from NGS studies. Duplications in RB1CC1 on chromosome 8 were overrepresented in patients. The CNVs were technically verified using quantitative PCR. The duplications were followed-up in independent European samples. In the combined analysis, comprising of 8,461 patients and 112,871 controls, duplications in RB1CC1 were found to be associated with schizophrenia ($P = 1.29 \times 10\text{-}5$; odds ratio = 8.58). To obtain additional evidence for RB1CC1 being a strong candidate gene for schizophrenia we are currently sequencing selected regions of RB1CC1 for the presence of rare mutations in 2,200 patients. RB1CC1 is a brain expressed gene and has been implicated in cell cycle progression and neurodegeneration. Our study is the first to provide evidence that rare duplications in RB1CC1 are a risk factor for schizophrenia. Results from our targeted resequencing of RB1CC1 wil be presented.

58

Variants in NRG3 Associated with Delusion Have Regulatory Potential and Differentially Bind to Nuclear Proteins. *M. Zeledon[1,2,3], M. Taub[5], N. Eckart[1,2], M. Beer[1,6], R. Wang[3], M. Szymanski[1,2], P. Chen[7], A.E. Pulver[3,4], J.A. McGrath[3,4], P. Wolyniec[3,4], D. Avramopoulos[1,3], A. Sawa[3], D. Valle[1].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Predoctoral Training Program in Human Genetics, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD; 4) Epidemiology-Genetics Program, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 5) Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 6) Whitaker Biomedical Engineering Institute, Johns Hopkins University, Baltimore, MD; 7) Department of Medical Genetics, National Taiwan University Hospital, Taipei City, Taiwan.

Schizophrenia (SZ) is a severely disabling psychiatric disease that affects 1% of the world's population. Symptoms include dysfunctions in cognition, thought/belief (delusions), perception/sensation (hallucinations) and affect. Previously, our group reported a linkage peak for SZ (NPL of 4.7) at 10q22 in the Ashkenazi Jewish (AJ) population and, in a follow-up fine mapping association study in the AJ, found strong evidence of association between a quantitative phenotypic trait "delusion" and three intronic SNPs in the 5' end of NRG3. Two other independent groups have replicated our findings, making NRG3 a strong candidate gene for a subtype of schizophrenia with delusions. To identify causative variants, we sequenced the 162 kb LD block covering the 5' end of NRG3 and containing the three associated SNPs in 47 AJ SZ patients at either extreme of the delusion quantitative trait. We identified 5 SNPs with minor alleles on the implicated haplotype and significantly overrepresented in high delusion patients. We tested these for regulatory potential and found that the delusion-associated alleles of rs10883866 and rs60827755 significantly and reproducibly decreased and increased (3-4 fold) expression of a reporter gene respectively, compared to the common allele in a variety of cell types including primary rat cortical neurons. To test whether these changes are due to differential binding of DNA regions surrounding the variants to nuclear protein(s), we performed electrophoretic mobility shift assays. We determined that the 21 bp surrounding the variants consistently bind to nuclear proteins and by competing with cold probe, found that binding of rs60827755 is influenced by the genotype of the SNP. In silico predictions of disrupted binding sites pointed to CNOT4 (CCR4-NOT transcription complex, subunit 4) as a candidate binding protein, and supershift data suggest that the shift is due in part to CNOT4 binding. In summary, we have identified cis-acting regulatory motifs that modify NRG3 expression and that account for the association with delusions in SZ have regulatory potential and bind to nuclear proteins, one of which may be CNOT4. Several proteins binding CNOT4 have been implicated in SZ by GWAS. We anticipate that these experiments will allow us to move from genetic observations to deciphering the mechanistic pathways that may lead to risk for SZ or modifications of the SZ phenotype.

59

Testing genetic associations with addiction phenotypes using moderate-depth whole genome sequencing. *S.I. Vrieze[1], S. Feng[1], X. Zhan[1], M.B. Miller[2], G. Jun[1], M.K. Trost[1], A. Tan[1], J. Bragg-Gresham[1], M. Flickinger[1], L. Scott[1], A. Locke[1], H.M. Kang[1], S. Levy[3], R.M. Myers[3], M. Boehnke[1], W.G. Iacono[2], M. McGue[2], G.R. Abecasis[1].* 1) Biostatistics, University of Michigan, Ann Arbor, MI; 2) Psychology, University of Minnesota, Minneapolis, MN; 3) HudsonAlpha Institute for Biotechnology, Huntsville, AL.

We report results from genetic association analyses of smoking, alcohol, and drug use and dependence behavior in 1800 individuals of European ancestry with moderate-depth whole genome sequencing from the Minnesota Center for Twin and Family Research. We developed and validated a novel selection method to prioritize pedigrees likely to carry rare disease-associated alleles. The method was applied to the full sample of 7188 individuals from 2400 nuclear family pedigrees, delivering a selected sample of 1480 individuals from 596 pedigrees. To date, DNA samples from 1343 individuals have been sequenced to 10x mean depth, which is estimated to provide >80% power to detect singletons. SNP calling and LD-based genotype refinement is complete on a preliminary data freeze of 686 samples. After quality control filtering, we have identified 21,190,919 autosomal SNPs, including 9,439,824 singletons and 4,117,829 doubletons. Fully 408,879 SNPs were predicted to be nonsynonymous and 18,158 putative loss of function. We evaluated more closely 172 candidate addiction regions tagged by previously-identified loci, which contain 8,090 nonsynonymous and 238 putative loss of function variants. These genes include 11 genes with known nicotine- and alcohol-relevant loci, as well as 161 other candidate addiction genes related to nicotinic receptors, nicotine metabolism, alcohol metabolism, and neurotransmitter systems known to be involved in addictive behaviors. Single nucleotide, CNV, and indel calls on the full sequenced sample will be presented, and will provide detail about the extent of deleterious variation across the genome and within key addiction genes. To test for association we conduct single variant and burden tests across the genome, and present in finer detail results for promising candidate addiction genes. Burden tests include a variable threshold burden test and the sequence kernel association test, separately for nonsynonymous variants and loss of function variants. These experiments will provide a comprehensive evaluation of genetic association in substance use, abuse, and dependence phenotypes, and the family-based design will allow additional traction in the study of very rare variation.

60

A Genomewide Association Study of Alcohol Dependence in the Irish Affected Sib Pair Study of Alcohol Dependence. *A.E. Adkins[1], L.M. Hack[2], T.B. Bigdeli[1], B.T. Webb[2], J.C. Bettinger[3], A.G. Davies[3], M.S. Grotewiel[4], C.A. Prescott[5], D.M. Dick[6], K.S. Kendler[6], B.P. Riley[6].* 1) Dept. of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA; 2) Dept. of Human and Molecular Genetics, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA; 3) Dept. of Pharmacology and Toxicology, Dept. of Psychiatry, Virginia Commonwealth University, Richmond, VA; 4) Dept. of Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA; 5) Dept. of Psychology, University of Southern California, Los Angeles, CA; 6) Dept. of Human and Molecular Genetics, Dept. of Psychiatry, Virginia Commonwealth University, Richmond, VA.

Background: We report results from a genomewide association study (GWAS) in an ethnically homogeneous Irish sample (N=710 related cases,1755 population controls) with strong supporting evidence from VCU Alcohol Research Center (VCU ARC) model organism (MO) studies. **Methods:** GWAS cases from the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD) were diagnosed using DSM-IV criteria. Affymetrix V6.0 arrays were genotyped at 3 separate core facilities and BeagleCall was used to call genotypes. IMPUTE2 and the 1000 Genomes reference haplotype panel (March 2012 freeze) were used to impute unmeasured genotypes. After QC filtering, imputation, and post-imputation filtering, 710 AD cases, 1755 controls and 8.2 million SNPs remained. Probabilities were converted to dosages with MACH2. Case/control association analysis was run using MQLS to correct for the non-independence of siblings. A sex weighted prevalence estimate of 8.9% was used for controls. We used a significance threshold of p<3.06E-8. FDR q-values were calculated with QVALUE in R. **Results:** Site effects were well corrected by BeagleCall and p-values showed little inflation (λ=1.05). SNPs in the collagen 6A3 (*COL6A3*, N=12) gene on chromosome 2 and an intergenic region of chromosome 3 (N=1) were genomewide significant. 725 SNPs in 103 independent loci had q-values ≤0.5. Preliminary experimental data using multiple MOs support 3 of the top 5 genes: *COL6A3* (top SNP p=6.18E-9,q=0.07), the Krueppel-like factor 12 (*KFL12*) gene on chromosome 13 (top SNP p=1.16E-7,q=0.08), and the Ryanodine receptor 3 (*RYR3*) gene on chromosome 15 (top SNP p=1.69E-7,q=0.08). Inactivation of one of three genes in *C. elegans* showing homology to human COL6A3 results in an ethanol resistance phenotype. The *C. elegans klf-3* mutant (orthologous to human *KLF12*) does not develop acute functional tolerance to ethanol and RNAi knockdown of the *D. melanogaster* homolog of *KLF12*, *luna*, results in enhanced sensitivity to ethanol. Finally, a loss of function allele of *unc-68*, the *C. elegans* homolog of *RYR3*, confers resistance to ethanol. **Discussion:** Our case-control GWAS of AD detected significant association signal in *COL6A3* based on 12 non-independent SNPs. Emerging evidence from the VCU ARC MO investigations provides strong additional support for 3 of the top 5 genes in our p-value ranked SNP list. Replication is underway for our top 725 SNPs in 4 independent samples of European descent (N>11,000).

61

NIH Study "Clinical and Molecular Investigations into Ciliopathies": Findings on Alström Syndrome. *J.D. Marshall[1], J. Han[2, 3], J.K. Naggert[1], D. Yildirimli[2, 4], J. Bryant[2, 4], R. Fischer[2, 4], W.M. Zein[2, 5], K. Daryanani[2, 6], B. Turkbey[2, 7], E. Turkbey[2, 6], C.-Y Liu[2, 6], P. Choyke[2, 7], T. Heller[2, 8], D. Rosing[2, 9], A. Brofferio[2, 9], V. Sachdev[2, 9], L. Olivieri[2, 9], N. Bridges[2, 10], J. Graf-Myles[2, 6], S. Bernstein[2, 6], K. Olivier[2, 10], B. Shamburek[2, 9], M. Huizing[2, 4], W.A. Gahl[2, 4], M. Gunay-Aygun[2, 4].* 1) Jackson Laboratory, Bar Harbor, ME; 2) National Institutes of Health; 3) National Institute of Child Health and Human Development; 4) National Human Genome Research Institute; 5) National Eye Institute; 6) NIH Clinical Center; 7) National Cancer Institute; 8) National Institute of Diabetes and Digestive and Kidney Diseases; 9) National Heart, Lung, and Blood Institute; 10) National Institute of Allergy and Infectious Diseases.

Alström syndrome (AS) is a ciliopathy characterized by childhood retinal dystrophy, obesity, diabetes, hyperlipidemia, cardiomyopathy, hearing loss, and renal and hepatic involvement. The causative gene ALMS1 encodes a large protein that localizes to the centrosome and basal body. The exact function of the ALMS1 protein is not known; it may play a role in centrosome assembly and/or function and endosome recycling. The frequency and nature of kidney, liver, cardiac, hormonal/metabolic, and other organ system involvement in AS is not well defined. Under our ciliopathy study (www.clinicaltrials.gov, trial NCT00068224), ongoing since 2003, we have evaluated a total of 275 patients including AS patients fulfilling the clinical diagnostic criteria. Evaluations included abdominal ultrasonography and magnetic resonance imaging (MRI) to quantitate visceral adiposity, echocardiogram, biochemical and hormone testing, ALMS1 sequencing, dual-energy X-ray absorptiometry to measure total body bone mineral content, lean mass, fat mass, MR-spectroscopy for liver and thigh muscle fat content, breakfast mixed-meal test to assess hormones related to energy homeostasis and resting metabolic rate measurement. Our AS cohort included 25 patients from 20 families; 3 families had 2 affected children and 1 contributed 3 affected siblings. Ages ranged from 1.7 to 38.9 years (15.7 + 10.9); there were 8 adults and 13 females. Early onset retinal degeneration and progressive sensorineural hearing loss were constant features. Even among siblings, there was considerable variability in the frequency and severity of individual components of the syndrome including cardiomyopathy and obesity, suggesting the presence of modifier genes. For example, in the family with 3 affected sons, the oldest had infantile-onset cardiomyopathy while the other 2, monitored from birth, had normal echocardiograms at ages 5 and 1.7 years. Three patients from 2 families including two 12-years-old non-identical twin brothers (with 2 truncating ALMS1 mutations) had no obesity (weight at 25th % since birth); potentially explainable by ALMS1 isoform-specific effects. We continue to enroll new AS patients to define the full phenotypic spectrum of this disorder, and to provide the groundwork for more focused studies and future therapeutic interventions. These data might potentially contribute to development of novel treatments for non-syndromic forms of obesity, diabetes and hyperlipidemia.

62

Genotype/epigenotype/phenotype correlations define Beckwith-Wiedemann syndromes. *A. Mussa[1], A. De Crescenzo[2], S. Russo[3], L. Calzari[3], N. Chiesa[1], C. Molinatto[1], G. Baldassarre[1], D. Melis[4], D. Milani[5], M.F. Bedeschi[5], L. Tarani[6], A. Selicorni[7], M. Cirillo Silengo[1], L. Larizza[3,8], A. Riccio[2], G.B. Ferrero[1].* 1) Department of Pediatrics, University of Torino, Italy, Torino, Torino, Italy; 2) Genetics and Biophysics Institute A. Buzzati-Traverso, CNR, Naples, Italy; 3) Laboratory of Cytogenetics and Molecular Genetics, Istituto Auxologico Italiano, Milan, Italy; 4) Department of Pediatrics, University Federico II, Naples, Italy; 5) Medical Genetics Unit, IRCCS Ca' Granda Foundation, Ospedale Maggiore Policlinico, Milan, Italy; 6) Genetica Clinica, Dipartimento di Pediatria, Università "La Sapienza" di Roma, Policlinico Umberto I, Roma; 7) Pediatric Clinical Genetics, Department of Pediatrics, University of Milano Bicocca, S. Gerardo Hospital MBBM Foundation, Monza, Italy; 8) Medical Genetics, Department of Medicine, Surgery and Odontoiatry, University of Milan, Italy.

Background — Beckwith-Wiedemann syndrome (BWS) is characterized by cancer predisposition and a variable association of overgrowth, macroglossia, abdominal wall defects, renal anomalies, nevus flammeus, ear malformations, hypoglycemia, hemihyperplasia, and organomegaly. BWS molecular bases are heterogeneous as several mechanisms lead to the disruption of the transcription of genes regulated by the two Imprinting Center (IC1 and IC2) in the 11p15 chromosomal region responsible for the syndrome. Objective — To search for genotype-phenotype correlations in a large cohort of BWS patients with positive molecular tests. Methods — 281 patients with BWS and proven molecular defect were included and characterized clinically. Patients' characteristics were compared among BWS molecular subclasses: KvDMR1 hypomethylation (IC2, n=172), H19/IGF2 hypermethylation (IC1, n=30), chromosome 11p15 paternal uniparental disomy (UPD, n=68), and CDKN1c mutation (n=11). Results — Each group showed different growth patterns: neonatal macrosomia was typical and almost constant in IC1 patients (90%), postnatal overgrowth in IC2 patients (50%), and hemihyperplasia more common in UPD patients (88%, $p<0.001$). Exomphalos was more common in IC2/CDKN1c patients (63% and 29%), whereas minor abdominal wall defects as diastasis recti and umbilical hernia were associated with IC1 defects (40% and 50%, $p<0.001$), consistent with organomegaly (20-25%) and polyhydramnios (11%). Renal defects were typical of UPD/IC1 patients (33%), and uretheral malformations of IC1 cases (23%, $p<0.001$). Ear anomalies and nevus flammeus were associated with IC2/CDKN1C genotype (52-63%, $p<0.001$). Macroglossia, almost always present (90% of cases), was less common among UPD patients (72%, $p<0.001$). Wilms' tumor was associated with IC1 defects (20%) or UPD (4%) and never observed in IC2 patients ($p<0.001$). Hepatoblastoma occurred only in UPD cases (6%) and other tumors were randomly scattered among molecular subclasses that displayed a different cancer risk, lower in IC2/CDKN1c defects, intermediate in UPD, and very high in IC2 cases ($p=0.050$). Conclusions — In BWS is definable a clear phenotype-(epi)genotype correlation is definitely present allowing to define three different syndromes with large clinical overlap. These clinical molecular correlations will likely allow a tailored follow-up and cancer screening procedures.

63

Myhre and LAPS syndromes : clinical and molecular review of 32 patients. *C. Michot[1], C. Le Goff[1], A. Afenjar[2], A.S. Brooks[3], P.M. Campeau[4], A. Destree[5], M. Di Rocco[6], D. Donnai[7], R. Hennekam[8], D. Heron[9], S. Jacquemont[10], P. Kannu[11], A.E. Lin[12], S. Manouvrier-Hanu[13], S. Mansour[14], S. Marlin[15], R. McGowan[16], H. Murphy[7], A. Raas-Rothschild[17], M. Rio[1], M. Simon[3], I. Stolte-Dijkstra[18], J.R. Stone[19], Y. Sznajer[20], J. Tolmie[21], J. van den Ende[22], N. Van der Aa[22], T. van Essen[18], A. Verloes[23], V. Cormier-Daire[1]*. 1) INSERM U781, Department of Genetics, Paris Descartes University - Sorbonne Paris Cité, Institut IMAGINE, Necker Enfants Malades Hospital, Paris, France; 2) Department of Neuropediatry, Centre de Référence Maladies Rares "anomalies du développement et syndromes malformatifs- Ile de France", Armand-Trousseau CHU, Paris, France; 3) Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, the Netherlands; 4) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, USA; 5) Department of Human Genetics, Institute of Pathology and Genetics, Gosselies, Belgium; 6) Unit of Rare Diseases, Department of Pediatrics, Gaslini Institute, Genoa, Italy; 7) Genetic Medicine University of Manchester, Manchester Academic Health Science Centre, St Mary's Hospital, Manchester, M13 9WL, UK; 8) Department of Pediatrics H7-236, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands; 9) Genetics and cytogenetics Departement, GRC-upmc, Pitié-Salpétrière CHU, Paris, France; 10) Department of Genetics, CHUV, CH-1011 Lausanne, Vaud, Switzerland; 11) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Canada; 12) Medical Genetics, Massachussets General Hospital for Children, Boston, Massachusetts; 13) Department of Clinical Genetics, University Hospital, Lille, France; 14) Clinical Genetics, St George's Healthcare NHS Trust, Tooting, London SW17 0QT, UK; 15) Genetic and Medical Embryology unit, Centre de Référence des Surdités Génétiques, Armand-Trousseau CHU, Paris, France; 16) North Scotland Regional Genetics Service, Clinical Genetics Centre, Ashgrove House, Foresterhill, Aberdeen, UK; 17) Institute of Human Genetics, Meir Medical Center, Kfar Saba, Israel; 18) Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands; 19) Department of Pathology, Massachussets General Hospital for Children, Boston, Massachusetts; 20) Center for Human Genetics, Cliniques Universitaires St-Luc - UCL B 1200 Brussels, Belgium; 21) Ferguson-Smith Department of Clinical Genetics, Yorkhill Hospital, Glasgow, UK; 22) Department of Medical Genetics, University and University Hospital Antwerp, Edegem, Belgium; 23) Clinical Genetics Functionnal Unit, Department of Genetics, Robert Debré CHU,Paris, France.

Myhre syndrome is characterized by short stature, brachydactyly, dysmorphic facial features, pseudomuscular hypertrophy, joint limitation and hearing loss. We previously identified *SMAD4* mutations as the cause of Myhre syndrome. More recently, *SMAD4* mutations have also been identified in Laryngotracheal stenosis, Arthropathy, Prognathism and Short stature syndrome (LAPS). This study aimed to review the features of Myhre and LAPS patients to better define the clinical spectrum of *SMAD4* mutations. 17 females and 15 males, ranging in age from 8 to 48 years, were included in the study. Among them, 30 cases were diagnosed with Myhre syndrome and 2 with LAPS. The entire *SMAD4* coding sequence was analyzed by Sanger sequencing. Details of clinical and radiological features were collected from a questionnaire completed by each referring physician. All patients displayed a typical facial gestalt, thickened skin, joint limitation and pseudomuscular hypertrophy. Growth retardation was common (22/32 - 68.7%) and was variable in severity (-5.5 to -2 SD). Intellectual deficiency of mild to moderate degree was observed in 28/32 (87.5%) and associated with behavioral troubles in 18 cases (56.2%). Significant health concerns included obesity (n=11), arterial hypertension (n=9), bronchopulmonary insufficiency (n=7), laryngotracheal stenosis (n=5) and pericarditis (n=3). Early death occurred in 4 patients (at 9, 19, 22 and 28 years). 29 patients (90%) had a de novo heterozygous *SMAD4* mutation, including both patients with LAPS. In 27 cases, mutation affected Ile 500 and in 2 cases Arg 496. The three patients without *SMAD4* mutations had typical findings of Myhre syndrome. We conclude that Myhre and LAPS syndromes are a unique clinically homogenous condition with life threatening complications in the course of the disease, supporting specific management recommendations. Our identification of *SMAD4* mutations in 29/32 cases confirms that *SMAD4* is the major gene responsible for Myhre syndrome. The finding of the Ile 500 change in *SMAD4* in the majority of cases and of a significant increase in paternal age may suggest a mechanism of protein-driven selfish selection in sperm.

64

The phenotype combining Treacher Collins Syndrome features with Diamond-Blackfan Anemia is a heterogeneous ribosomopathy. *K. Sol-Church[1], D. Stabley[1], A. Haskins Olney[2], C. Curry[3], J. Fisher[3], L. Pilchman[1], C. Schanen[1], J.X. Chong[4], D.E. Ward[1], K.W. Gripp[5], UW Center for Mendelian Genomics, Seattle WA*. 1) Biomedical Research, Alfred I duPont Hospital for Children, Wilmington, DE; 2) Munroe-Meyer Institute for Genetics and Rehabilitation, University of Nebraska Medical Center, Omaha, NE; 3) Genetic Medicine Central California, Fresno/UCSF, CA; 4) University of Washington Center for Mendelian Genomics, Seattle, WA; 5) Medical Genetics, A. I. duPont Hospital for Children, Wilmington, DE.

Diamond-Blackfan anemia (DBA) and Treacher Collins syndrome (TCS) are two distinct congenital disorders caused by defective ribosome biogenesis. DBA is a bone marrow failure syndrome, which typically present with macrocytic anemia in early infancy, short stature with craniofacial anomalies and a risk for cancer. DBA is caused by heterozygous mutations in ribosomal proteins, most commonly *RPS19*. TCS is characterized by severe craniofacial dystosis resulting in conductive hearing loss and typically results from heterozygous *TCOF1* mutations affecting RNA polymerase. A combination of facial features resembling DBA and TCS was reported as a distinctive phenotype (Gripp, AJMG 2001; Handler, J Craniofac Surg 2009). Patients with the combined DBA/TCS phenotype were negative for mutations in these candidate genes. Material and Methods: Whole exome sequencing was performed on DNA isolated from the members of five unrelated families including seven individuals with a DBA/TCS phenotype. Data analysis generated candidate variants that were validated using Sanger sequencing and segregation analysis. Results: A heterozygous germline *RPS26* mutation c.259C>T, predicting p.Arg87Ter, is present in the affected family members reported in Handler 2009, but not in the unaffected mother. This variant identified by rs148942765 is not seen in the 1,000 genome or UW-CMG databases. In two families, a novel c.1A>G transition in exon 1 of *RPS28* (p.Met1Val), is heterozygous in one proband; and apparently mosaic in another unrelated proband. This mutation is not reported in the publicly available databases. In a fourth family, two cousins inherited an X-linked c.191A>G mutation in the *TSR2* gene from their mothers. The resulting p.Glu64Gly change affects a conserved residue of a pre-rRNA-processing protein. Data analysis is ongoing for the last family. Discussion: We identified novel pathogenic mutations in genes involved in RNA processing in 4 families with a complex DBA/TCS phenotype. Though the *RPS26* mutation is novel, this gene is known to be involved in DBA. In contrast, this is the first evidence that a mutation in *RPS28* causes this phenotype. It is noteworthy that a similar p.Met1Val variant in RPS26 causes DBA10. Most strikingly, we identified an X-linked form of this syndrome in cousins harboring a mutation in yet another ribosomal gene *TSR2*. These data suggest that the DBA/TCS phenotype is a rare expression of the DBA spectrum, rather than an independent phenotype.

65

Weaver syndrome is caused by loss-of-function mutations in *EZH2*. *A.S.A. Cohen[1,2], D.B. Yap[3,4], X. Han[5], S.M.E. Lewis[1,2,6], C. Chijiwa[1,6], M.A. Ramos-Arroyo[7], D.D. Weaver[8], C.J.D. Ross[1,2,5], S. Aparício[3,4], W.T. Gibson[1,2].* 1) Department of Medical Genetics, University of British Columbia, Vancouver, BC V6T 1Z3, Canada; 2) Child and Family Research Institute, Vancouver, BC V5Z 4H4, Canada; 3) Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, V6T 2B5, Canada; 4) Department of Molecular Oncology, British Columbia Cancer Research Centre, Vancouver, BC V5Z 1L3, Canada; 5) Centre for Molecular Medicine and Therapeutics, Vancouver, BC V5Z 4H4, Canada; 6) British Columbia Children's and Women's Health Center, Vancouver, V6H 3N1, Canada; 7) Department of Medical Genetics, Complejo Hospitalario de Navarra, Pamplona 31008, Spain; 8) Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN 46202, USA.

In late 2011, we and others found that constitutional mutations in the epigenetic regulator *EZH2* (enhancer of zeste homolog 2) cause Weaver Syndrome (WS). WS is characterized by overgrowth, increased height, large head, accelerated bone growth and maturation, intellectual disability, neuronal migration defects and susceptibility to various cancers. Having established a diagnostic test, we are now investigating the link between *EZH2* mutations and the clinical features of WS, including cancer development.

In addition to the mutations we previously reported, we have identified *EZH2* mutations in three other patients from our cohort of 29 individuals with Weaver-like phenotypes: two novel mutations and one reoccurring mutation. Our goal is to look at phenotype/genotype correlations, with particular emphasis on determining characteristics that would help us predict the likelihood of these patients developing cancer. This would allow for early screening to detect both solid tumours and haematological malignancies, which should increase the patients' chances of survival.

Furthermore, by collaborating with the Aparício lab at the BC Cancer Agency, we have investigated the effects of WS-associated mutations on protein function using *in vitro* assays. Normally, EZH2 acts as a histone methyltransferase in the polycomb-repressive complex 2 (PRC2), and silences transcription through methylation of histone H3 lysine 27 (H3K27). In contrast to the enhanced trimethylation activity seen in some somatic mutations that cause leukemia, the WS-associated EZH2 mutants show reduced methylation activity. These results support the hypothesis that WS is caused by loss-of-function mutations in *EZH2*. Although strategies to reduce EZH2 activity are currently being explored in common cancers such as leukemias, prostate cancer and breast cancer, our data suggest that these strategies may not be effective in rare cancers associated with WS.

66

Novel ELN Mutations and Vascular Phenotype in Autosomal Dominant Cutis Laxa. *E. Lawrence[1], M. McGowan[1], K. Levine[1], C. Lorenchick[1], S. Alkan[1], H. Salvaggio[2], A. Zaenglein[2], M. Bodzioch[3], A. Kiss[4], M. Siefring[5], Z. Urban[1].* 1) Department of Human Genetics University of Pittsburgh; 2) Department of Dermatology Penn State Milton S. Hersey Medical Center; 3) Jagiellonian University Medical College; 4) Dept. of Clinical Genetics Universidade Federale de Ciencias de Porto Allegre; 5) Stamford Skin and Medical Centre.

Cutis laxa (CL) is an inherited skin disease with remarkable locus heterogeneity. Mutations in the elastin gene (ELN) cause autosomal dominant cutis laxa (ADCL). We have sequenced ELN for exons 30-34 in 89 consecutive probands with CL and a subset of 20 of these probands for all 34 exons in the gene. Mutations were identified in 11 probands. Three families had a c.2296delA in exon 32, one presented with a c.2177delC in exon 30 and one with a c.2137delG in exon 30, all of which were previously published for other families. Several families had previously unreported mutations, a c.2184delT in exon 30 and a c.2351delG in exon 34. The more interesting of the unreported mutations were splice site changes that were found in the remaining four families. Three unrelated probands had the same de novo change a c.2132-7C>A located between exons 33 and 34 which activated a cryptic splice site at position -5 and caused an addition of five nucleotides between exons 33 and 34. This insertion led to a frame-shift which extended the open reading frame. The remaining proband not only had a splice change but it was located outside the canonical region of exon 30-34 where most of the previously published mutations had been located. This family has a c.133+1delG in intron 2 which also activated a cryptic splice site and led to a 28-amino acid in frame insertion between exons 2 and 3. An examination of the origin of the mutations was possible in 10 families, 2 showing inherited, and the remaining 8 de novo mutations. In addition to CL, 4/11 affected individuals had mild-moderate obstructive pulmonary disease and one had a mild aortic root dilatation. There were also 5 cases of arterial tortuosity and two of venous tortuosity both of whom had the mutation c.2132-7C>A. Immunoblotting showed elevated canonical, but unaffected non-canonical transforming growth factor-beta signaling despite unaltered extracellular TGFβ activity in fibroblasts from patients with ELN mutations. We conclude that ELN-related CL explains approximately 12% of cutis laxa cases and elevated TGFβ signaling is a disease mechanism in ADCL shared with other connective tissue diseases.

67

Mutations in *Lrp5* improve bone properties in a mouse model of Osteogenesis Imperfecta. *C.M. Jacobsen[1,2,3,4], L.A. Barber[3], U.M. Ayturk[3,5], H.J. Roberts[3], M.A. Schwartz[3], M. Weis[6], D. Eyre[6], D. Zurakowski[7], A.G. Robling[8], M.L. Warman[3,5,9].* 1) Division of Endocrinology, Boston Children's Hospital, Boston, MA; 2) Division of Genetics, Boston Children's Hospital, Boston, MA; 3) Orthopaedic Research Laboratories, Department of Orthopaedic Surgery, Boston Children's Hospital, Boston, MA; 4) Department of Pediatrics, Harvard Medical School, Boston, MA; 5) Department of Genetics, Harvard Medical School, Boston, MA; 6) Department of Orthopedics and Sports Medicine, University of Washington, Seattle, WA; 7) Department of Anesthesia, Children's Hospital Boston, Boston, MA; 8) Department of Anatomy and Cell Biology, Indiana University, Indianapolis, IN; 9) Howard Hughes Medical Institute, Boston, MA.

The cell surface receptor, low-density lipoprotein receptor-related protein 5 (LRP5) has emerged as a key regulator of bone mass and strength. Heterozygous missense mutations in LRP5 can increase bone mass and strength in humans. These mutations reduce binding to LRP5 by its endogenous inhibitors. Mice with a *Lrp5* missense mutation that is orthologous to a human high bone mass (HBM) causing mutation have increased bone formation and as a consequence increased bone mass and bone strength. Osteogenesis Imperfecta (OI), a disorder most frequently caused by mutations in Type I collagen, is characterized by increased skeletal fragility which can lead to deformity, pain, and disability. Therapies are needed that can reduce fractures and improve bone health in patients with OI.

We mated mice with a dominant *Lrp5* HBM-causing knockin allele to mice with a dominant OI-causing *Col1a2* allele. We evaluated the offspring at 12-weeks-old. We analyzed skeletal properties by DEXA, μCT, whole-bone 3-point bending and quantitative histomorphometry. As expected, offspring with OI alone have lower bone density and weaker bones than wild-type siblings, and mice with a HBM allele alone have higher bone density and stronger bones than wild-type siblings. Compared to siblings with OI alone, siblings with OI and a HBM allele have significantly increased bone density and bone strength ($p < 0.05$). RNAseq and collagen analysis from the bones of mice with OI and a HBM allele demonstrated that the improved bone properties were not due to altered mRNA expression of type I collagen or its chaperones, nor were they due to changes in mutant type I collagen secretion.

Increased LRP5 signaling improves bone density and strength in a mouse model of moderate OI. This proof of principle experiment indicates that promoting bone anabolism may benefit patients with OI, even though the underlying collagen defect responsible for their skeletal fragility is unchanged. Therefore, therapies that target the LRP5 signaling pathway could be effective in improving outcomes for patients with OI.

68

Teriparatide, the first anabolic agent for treatment of osteogenesis imperfecta improves bone mineral density at the hip and spine: a randomized, blinded, placebo-controlled trial. *SC. Sreenath Nagamani[1], J. Shapiro[2], S. Veith[3], Y. Wang[3], J. Lapidus[3], JL. Reeder[3], TM. Keaveny[4], D. Lee[4], MA. Mullins[1], B. Lee[1,5], ES. Orwoll[3].* 1) Baylor College of Medicine, Houston, TX, USA; 2) Kennedy Krieger Institute, Baltimore, MD, USA; 3) Oregon Health & Science University, Portland, OR, USA; 4) ON Diagnostics, Berkeley, CA, USA; 5) Howard Hughes Medical Institute, Houston, TX, USA.

Osteogenesis Imperfecta (OI) refers to a group of disorders that are characterized by bone fragility and recurrent fractures. Therapy for OI is presently limited to bisphosphonates and there have been few controlled studies for treatment of OI. As mutations in OI alter the bone matrix, affect osteoblast function and bone remodeling, treatment with anabolic bone agents may be useful to increase bone mass and strength. In the largest placebo-controlled trial in adults with OI, and the first trial of anabolic therapy in this disease, we evaluated the effects of teriparatide treatment. We performed a randomized, double-blind, placebo-controlled trial of teriparatide in 77 adults with OI (33 men, 44 women). The mean age of subjects was 41 years (range 18-75). Fifty-one had OI type I, 14 had OI type III, and 12 had OI type IV. Subjects were treated for 18 months with either teriparatide (20µg/day) or placebo. Baseline and 6-monthly measures of bone mineral density (BMD), i.e. DXA scans and vertebral QCT scans were obtained. Self-reported fractures at any site were recorded. Serum P1NP and urine NTX levels increased (135$\pm$14% and 64$\pm$10%, respectively) with teriparatide therapy but remained stable in the placebo group suggesting increased bone remodeling with therapy. Total hip areal BMD (2.5$\pm$1.3 % vs. -3.4$\pm$1.1; $p< 0.001$) and spine BMD (6.0$\pm$1.2 % vs. 0.9$\pm$1.2 %; $p< 0.05$) increased significantly more in the teriparatide treated than in the placebo group. Volumetric spine BMD increased in those receiving teriparatide therapy (18$\pm$6%) but declined in those receiving placebo (-5.0$\pm$6%; $p< 0.05$). The estimated vertebral strength as assessed by finite element analysis increased 14% in the teriparatide group and decreased 2.5% in the placebo group ($p= 0.001$). The BMD increase was robust in patients with OI type 1 whereas there was no significant increase in those with OI types III and IV. The study could not definitively assess the effects on fracture outcomes as it was not adequately powered to detect a change in fracture rates. There was no statistically significant difference in fractures (RR 0.8; CI 0.42-1.55). There were no drug-related serious adverse events. Our study proves that anabolic therapy with teriparatide in adults with OI is well-tolerated and increases BMD at the hip and the spine, particularly in those with OI type I. Teriparatide could be a potential therapy for OI though further studies would be needed to assess its effects on fracture rates.

69

Integrin Modulating Therapies Prevent Fibrosis and Autoimmunity in Genetic Mouse Models of Scleroderma. *E.E. Gerber[1], E.M. Gallo[1], S.C. Fontana[1], E.C. Davis[3], X. Zhong[1], F.M. Wigley[4], D.L. Huso[5], H.C. Dietz[1,2].* 1) Institute for Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA; 3) McGill University, Montreal, Quebec H3A 2K6, Canada; 4) Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 5) Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.

Systemic sclerosis (SSc), the most common form of scleroderma, affects up to 1 in 5,000 individuals and is characterized by the acquired predisposition for autoantibodies and progressive dermal and visceral fibrosis in young adults, culminating in death. The study of SSc is hindered by the absence of a defined genetic contribution and a lack of animal models that mechanistically recapitulate human disease. In an effort to overcome these barriers, we focused on the study of a rare but tractable childhood presentation of scleroderma called stiff skin syndrome (SSS) that is caused by heterozygous missense mutations in the gene (FBN1) encoding the extracellular matrix protein fibrillin-1. All SSS mutations localize to a single domain in fibrillin-1 that mediates cellular attachment via the binding of integrins to a specific Arg-Gly-Asp (RGD) sequence. We generated two mouse lines harboring either a heterozygous SSS-associated amino acid substitutions in fibrillin-1 (W1570C) or one that results in obligate loss of integrin-binding (RGD to RGE). Both mouse lines recapitulate fully penetrant dense dermal fibrosis by three months of age with the added predisposition for aggressive visceral fibrosis upon minimal environmental provocation. Mutant mice show tissue infiltration and activation of pro-inflammatory immune cells including plasmacytoid dendritic cells (pDCs), Th2 and Th17 T helper cell subsets, and plasma cells in association with high circulating levels of anti-nuclear antibodies and anti-topoisomerase I antibodies that are more characteristic of and specific for SSc. Activation of pDCs is associated with high expression of many pro-fibrotic cytokines including transforming growth factor β (TGFβ), interleukin-6 (IL-6) and interferon α (IFNα). The combination of IL-6 and TGFβ is sufficient for Th17 skewing (with IL-17 production), while the combination of IL-6 and IFNα is sufficient for plasma cell infiltration and activation with autoantibody production. Furthermore, naïve pDCs show an enhanced propensity to attach and activate in culture when plated on the matrix elaborated by SSS cells, when compared to controls. All phenotypic abnormalities including fibrosis and autoimmunity are fully prevented by integrin-modulating therapies and reversed by TGFβ antagonism in both SSS mouse models. These data show that alterations in cell-matrix interactions are sufficient to fully phenocopy autoimmune scleroderma and highlight novel therapeutic strategies.

70

Large-scale parent-child trio sequencing highlights factors influencing spontaneous human mutation. *S. Sunyaev[1], P. Polak[1], L. Francioli[2], W. Kloosterman[2], P.I.W. de Bakker[2], Genome of Netherlands (GoNL) Consortium.* 1) Brigham & Women's Hospital, Harvard Medical School, Boston, MA; 2) University of Utrecht Medical Center.

Characterization of context-dependency and regional variation of human mutation rate is important for studies focusing on the role of de novo mutations in Mendelian and complex phenotypes. Study design based on identifying genes with recurrent de novo mutations associated with a phenotype requires an accurate model of local mutation rate. Further, quantitative description of context-dependency and regional variation of human mutation rate is a key step towards understanding biology of spontaneous mutagenesis. It is also an important ingredient in models of genome evolution and evolution of genetic disease. Current knowledge of properties of human mutation rate primarily comes from indirect inference from comparative genomics and population genetic variation data. Whole genome sequencing of multiple parent-child trios enables direct characterization of properties of spontaneous mutagensis in humans. The Genome of the Netherlands (GoNL) is an effort to characterize genomic variation in the Dutch population through whole-genome sequencing of 250 families (231 trios, 19 twin quartets) at 12x using Illumina HiSeq. Sequencing was performed by BGI (China). We developed a Bayesian algorithm to detect de novo mutations from pedigree data and implemented it as the PhaseByTransmission module in the Genome Analysis Toolkit (GATK). In total, we called more than 19,600 de novo mutations, of which 44% were called with high confidence. 92% of the high-confidence calls were independently validated. We used the resulting set of high confidence de novo mutations to characterize context dependency and regional variation of human mutation rate. Our analysis was assisted by newly developed statistical methods to correct for false-positive and false-negative mutation calls. We observed excellent agreement of context-dependent point mutation rates with earlier predictions from comparative genomics. We also detected a striking strand asymmetry in transcribed regions with the rate of A>G transitions elevated by 40% in the non-transcribed strand compared to the transcribed strand, suggesting the impact of transcription-coupled repair on human germ-line mutagenesis. Finally, we quantified regional variation in mutation rate and specifically addressed influence of epigenetic variables such as replication timing and chromatin architecture. This analysis generated hypotheses on specific factors shaping the landscape of human mutation.

71

Palindromic *GOLGA* core duplicon promotes 15q13.3 microdeletion, inversion polymorphisms, and large-scale primate structural variation. *M.Y. Dennis[1,7], F. Antonacci[1,7], J. Huddleston[1,2], P.H. Sudmant[1], K. Meltz Steinberg[3], T.A. Graves[3], M. Malig[1], L. Vives[1], L. Denman[1], C. Baker[1], C.T. Amemiya[4], A. Stuart[4], W.J. Tang[4], B. Munson[4], J.A. Rosenfeld[5], L.G. Shaffer[5,6], R.K. Wilson[3], E.E. Eichler[1,2].* 1) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA; 2) Howard Hughes Medical Institute, Seattle, WA, USA; 3) The Genome Institute at Washington University, Washington University School of Medicine, St. Louis, MO, USA; 4) Benaroya Research Institute at Virginia Mason, Seattle, WA, USA; 5) Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA, USA; 6) Paw Print Genetics, Genetic Veterinary Sciences, Inc., Spokane, WA, USA; 7) These authors contributed equally to this work.

Microdeletions of chromosome 15q13.3 are one of the most common recurrent structural variants associated with intellectual disability, schizophrenia, autism, and epilepsy. Non-allelic homologous recombination of flanking segmental duplications (SDs) at this locus mediates the formation of structural variants; however, the complex architecture of these SDs has yet to be fully described. Utilizing Illumina short-read, capillary, and PacBio sequencing, we comprehensively characterized these SDs in humans and nonhuman primates. We discovered and characterized five alternate structural configurations of the 15q13.3 region in humans ranging in size from 2 to 3 Mbp; these configurations arose as a result of human-specific expansions of SDs in conjunction with two independent evolutionary inversion events—a 2 Mbp inversion (γ) and a smaller 203 kbp inversion (β), never before reported. Interestingly, the β inversion is population stratified with frequencies ranging from less than 5% in Asian to as high as 40% in European populations. We show that both inversion polymorphism breakpoints map to a *GOLGA* core duplicon—a primate-specific chromosome 15 repeat of 15 kbp—embedded within a larger 58 kbp inverted repeat (or palindrome). These changes led to the formation of two different structural configurations predisposing to 15q13.3 and *CHRNA7* microdeletions, respectively. Remarkably, the same *GOLGA*-flanked palindrome that promoted two independent inversion events in humans also demarcated the breakpoints of the recurrent 15q13.3 microdeletions as well as an expansion of SDs in the human lineage and an independent inversion in chimpanzee. Overall, these results provide direct evidence for the role of this core duplicon and its palindromic architecture in evolutionary and disease-related instability of chromosome 15.

72

Large-scale genotyping of polymorphic inversions in human populations. *S. Villatoro[1], C. Aguado[1], D. Vicente[1], D. Izquierdo[1], M. Puig[1], M. Cáceres[1,2].* 1) Institut de Biotecnologia i de Biomedicina, Universitat Autonòma de Barcelona, Barcelona, Spain; 2) Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

In the last years, different types of structural variants (SVs) have been discovered in the human genome and their importance in human diseases, complex traits and evolution has become increasingly clear. Among the different types of SVs, inversions are the less known due to the difficulty to analyze them and with few exceptions, their frequency and distribution in human populations has not been widely studied. As part of a larger project towards the complete characterization of inversions in humans (INVFEST), we have set up a custom assay based on hybridization of probes that allows us to interrogate inversions with simple breakpoints, as well as inversions flanked by long repetitive sequences. Using this new technique, we have developed a high-throughput protocol to genotype 41 inversions in a large number of samples. In this study, first, inversions were genotyped in 90 individuals of European origin (CEU) and compared with previous results obtained by PCR in order to calculate the accuracy of our method, which was established in a 99%. Then, we analyzed additional samples up to a total of 550 individuals from different populations included within the 1000 Genomes Project: CEU, Toscani (TSI), Yoruba (YRI), Luhya (LWK), Chinese (CHB), Japanese (JPT) ,and Gujarati Indians (GIH), with an European, African and Asian origin, respectively. Using the data obtained, the genotyping-success rate of this method was estimated in a 98%. Besides, in all cases the genetic transmission in 60 mother-father-child European and Yoruba trios was correct. Next, we estimated the global inverted allele frequencies on the unrelated samples, which ranged between 0.5-98.8% and all inversions were in Hardy-Weinberg equilibrium in each population. Finally, significant differences among populations were found for some of the inversions, with Fst statistic values between 0.01-0.49. In conclusion, this assay provides a powerful, reliable and fast high-throughput method to genotype a wide range of inversions, which has allowed us to create the largest map of the frequency and distribution of these SVs in human populations. In addition, this information has opened the possibility to explore the association of inversions with nucleotide variation, gene expression, and phenotypic features. Support: European Research Council (ERC) Starting Grant (INVFEST) under the European Union Seventh Research Framework Programme (FP7).

73

***SCRIB* and *PUF60* are primary drivers of the multisystemic phenotypes of the 8q23.4 CNV.** *C. Golzio[1], A. Dauber[2], F.M. Jodelka[3], C. Guenot[4], J.S. Beckmann[4], J.N. Hirschhorn[5,6,7], M.L. Hastings[3], S. Jacquemont[4], N. Katsanis[1,8].* 1) Center for Human Disease Modeling and Department of Cell Biology, Duke University, Durham, NC; 2) Boston Children's Hospital, Boston, MA; 3) Chicago Medical School, Department of Cell Biology and Anatomy, IL; 4) Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland; 5) Program in Medical and Population Genetics, Broad Institute, Boston, MA; 6) Division of Endocrinology, Boston Children's Hospital, Boston, MA; 7) Departments of Genetics and Pediatrics, Harvard Medical School, Boston, MA; 8) Department of Pediatrics, Duke University, Durham, NC.

Copy number variants (CNVs) are frequent contributors to human genetic disorders. These lesions also represent a significant interpretative challenge, given that each CNV typically affects the dosage of multiple genes. Here we report five patients with coloboma, microcephaly, short stature, craniofacial, cardiac, and renal defects who harbor overlapping microdeletions on 8q23.4. Fine mapping localized a commonly-deleted 78kb region that contains three genes: *SCRIB*, *NRBP2,* and *PUF60*. *In vivo* dissection of the CNV showed discrete contributions of *SCRIB* and *PUF60* to the observed syndromic phenotype, while the combinatorial suppression of both genes exacerbated some, but not all, phenotypic components. Consistent with these findings, we identified a patient with microcephaly, short stature, intellectual disability and heart defects with a de novo p.H169Y change in the splicing factor *PUF60*. Functional testing of this allele *in vivo* and *in vitro* showed that the mutation perturbs the relative dosage of two *PUF60* isoforms and, subsequently, the splicing efficiency of downstream *PUF60* targets. These data inform the functions of two genes not associated previously with human genetic disease and demonstrate how CNVs can exhibit complex genetic architecture, with the phenotype being the amalgam of both discrete dosage dysfunction of single transcripts and of binary genetic interactions.

74

Functional dissection of the recurrent reciprocal 1q21.1 autism-associated CNV. *N. Katsanis[1,2], I. Blumenthal[3,4], A. Ragavendran[3,4], M.E. Talkowski[3,4], C. Golzio[1].* 1) Center for Human Disease Modeling and Department of Cell biology, Duke University, Durham, NC; 2) Department of Pediatrics, Duke University, Durham, NC; 3) Molecular Neurogenetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 4) Departments of Neurology and Genetics, Harvard Medical School, Boston, MA.

Copy number variants (CNVs) are frequent lesions involved in both rare and complex human traits. This has raised the challenge of identifying which genes within the CNV drive observed clinical traits. We have previously shown how the combinatorial use of neuroanatomical surrogate phenotypes in zebrafish embryos and genomic studies can dissect a single contributory locus to these phenotypes within the 600kb CNV on 16p11.2 associated with multiple developmental and psychiatric disorders. These findings prompted us to ask whether this approach could be useful for the systematic dissection of other CNVs that manifest similar defects. The 1q21.1 CNV is associated with congenital heart defects, head size defects, and represents the second most common lesion in autism. We noted that the minimal nine-gene deletion of this CNV is associated with microcephaly, while the reciprocal duplication is associated with macrocephaly, rendering it tractable by our methods. Systematic overexpression and suppression of all genes in the CNV provided a striking result that only one gene, the chromodomain-helicase-DNA-binding protein CHD1L, gave significant head size changes; overexpression of human CHD1L mRNA led to macrocephaly, while suppression of chd1l lead to a significant decrease in head size. These phenotypes were likely driven by defects in neurogenesis; chd1l morphants exhibited reduced rates of neuronal proliferation while, conversely, overexpressants showed increased rates of neuronal proliferation in vivo. Given the role of chromodomain genes in transcriptional regulation, and their emerging roles in neuropsychiatric disorders, we next asked which genes are regulated by CHD1L during brain development and whether there are any common transcriptional networks relevant to autism. We thus performed a transcriptomic evaluation of RNA-seq data from heads of control, CHD1L RNA-, and chd1l MO-injected embryos at 5 days post-fertilization. Using a linear regression model, we identified 257 loci with significant reciprocal differential expression correlated with gene dosage. An intriguing subset these loci have been associated with autism or connected to known loci by a first-order interaction. We are evaluating this dysregulated network in vivo with regard to neurogenesis and head size regulation and we will present data on the ability of this gene set to contribute to neurodevelopment and neurodevelopmental disorders.

75

Discovery of Genes Responsible for Neurocognitive Disease by Large Scale Integration of Sequence and Copy Number Data. *B.P. Coe[1], K.T. Witherspoon[1], C. Baker[1], B. O'Roak[1], J. Schuurs-Hoeijmakers[2], J. Shendure[1], B. deVries[2], J. Gecz[3], M. Fichera[4], C. Romano[5], L.G. Shaffer[6], J.A. Rosenfeld[7], E.E. Eichler[1].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 3) School of Paediatrics and Reproductive Health, The University of Adelaide, Adelaide, Australia; 4) Laboratory of Genetic Diagnosis, I.R.C.C.S. Associazione Oasi Maria Santissima, Troina, Italy; 5) Unit of Pediatrics and Medical Genetics, I.R.C.C.S. Associazione Oasi Maria Santissima, Troina, Italy; 6) Paw Print Genetics, Spokane, WA; 7) Signature Genomics, Spokane, WA.

Copy number variants (CNVs) and sequence variants, including indels and single nucleotide variants (SNVs), have been associated with a variety of neurocognitive disorders; however these events are typically individually rare and thus require very large populations to identify with case-control significance. Large scale CNV screens of patients and controls allow sensitivity in identifying pathogenic events, but the large size of typical pathogenic CNVs results in the identification of multiple candidate genes per locus. In contrast, sequence variation is gene specific, and typically restricted to smaller study populations. Here we combined large scale CNV analysis with targeted sequencing of high priority candidates to enhance sensitivity and specificity of gene discovery. We compared the CNV landscape of 29,206 children referred to diagnostic labs with developmental delay and intellectual disability (DD/ID) to 19,584 healthy controls. This identified 66 regions including 31 novel regions that show an excess of large deletions or duplications in cases when compared to controls. This large scale case-control approach has yielded precise estimates of clinical significance for pathogenic copy number variants, as well as the identification of new loci and case-control significance for rare previously described loci such as 15q24, 3q29 and 2q11.2. We next targeted 36 genes which had been highlighted by large CNVs and de novo variants in studies of DD/ID and autism cohorts with molecular inversion probes. These genes were then sequenced in 3,249 additional cases of ID/DD and autism, and 2,600 controls, with follow up in parents when gene disruptive events were identified. The integration of CNV and sequence data has allowed us to specifically identify several genes including statistical enrichment of both CNVs and loss of function mutations in ZMYND11 in the 10p15.3 deletion syndrome, as well as statistical enrichment of loss of function mutations in additional genes including SETBP1 and ARID1B. In conclusion, this combined approach has allowed for the rapid discovery of potentially new syndromes and genetic causes of neurocognitive disease.

76

Mind the gap - Exomes and CNVs testing in primary immunodeficiencies. *A. Stray-Pedersen[1,2], H.S. Sorte[2], P.S. Samarakoon[2], L. Forbes[3,4], T. Gambin[1], O.K. Rødningen[2], I.C. Hanson[3], L.M. Noroski[3], C. Davis[3], F. Seeborg[3], S.K. Nicholas[3], J.W. Caldwell[5], C.R. Beck[1], T.J. Vece[6], W. Wiszniewski, SJ Penney[1], S.N. Jhangiani[1], L. Mæhle[1], A. Patel[7], H.C. Erichsen, TE Abrahamsen[8,9], G.E. Tjonnfjord[9,10], B.E. Kristiansen, M Kulset[2], L.T. Osnes[11], W.T. Shearer[3], B. Fevang[12], K.R. Heimdal[2], D.E. Undlien[2,9], R.A. Gibbs[1,13], R. Lyle[2], J.S. Orange[3,4], J.R. Lupski[1,7,13,14], The Centers for Mendelian Genomics.* 1) Baylor-Hopkins Center for Mendelian Genomics at Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Medical Genetics, Oslo University Hospital, Oslo, Norway; 3) Immunology, Allergy and Rheumatology, Texas Children's Hospital, Houston, TX; 4) Center for Human Immunobiology, Texas Children's Hospital and Department of Pediatrics, Baylor College of Medicine, Houston, TX; 5) Section Pulmonary, Critical Care, Allergy and Immunologic Diseases, Wake Forest Baptist Medical Center, Winston-Salem, NC; 6) Department of Pulmonology, Texas Children's Hospital, Houston, TX; 7) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 8) Department of Pediatrics, Oslo University Hospital, Oslo, Norway; 9) University of Oslo, Norway; 10) Department of Hematology, Oslo University Hospital, Oslo, Norway; 11) Department of Immunology, Oslo University Hospital, Oslo, Norway; 12) Section of Clinical Immunology and Infectious Diseases, Oslo University Hospital, Oslo, Norway; 13) Human Genome Sequencing Center HGSC, Houston, TX; 14) Pediatrics, Baylor College of Medicine and Texas Childrens Hospital, Houston, TX.

Primary immunodeficiencies (PIDDs) constitute a heterogenous group of genetic diseases affecting the immune system. Dependent on the genetic etiology, symptoms range from mild to severe and life threatening. Knowledge of the molecular genetic cause and disease mechanism is important and can direct targeted and curative therapy. However, categorization of the subtypes is challenging as patients with different immunodeficiencies may have overlapping immunological and clinical phenotypes. In addition, more than 200 causal genes have been reported, and few are offered for clinical diagnostic genetic testing. We examined the utility of combining exome sequencing (exomeSeq) and array comparative genomic hybridizaton (aCGH) in the diagnostic workup and research of PIDDs. As of June 2013, 120 patients with extensive immunological and genetic testing from 90 families have been recruited from Oslo University Hospital (Norway) and Texas Childrens Hospital (Houston, USA). Initial analyses were individually tailored based on clinical data and immunophenotyping and family history. ExomeSeq data was systematically screened for variants in all reported PIDD genes. Generally, exomeSeq data does not directly detect copy number variants (CNVs), but a computational CNV prediction pipeline was applied to predict potential PIDD-causing CNVs from the exomeSeq data. Predicted PIDD-causing variants were validated using custom high-resolution aCGH containing exon-wise probes. After analysing the exomeSeq data alone in the first 60 families (78 patients), five novel genes were identified. PIDD relevant variants were detected in 60 percent (41 families), 21 of these attaining a definitive molecular PIDD diagnosis. The remaining (20) patients had either a heterozygous potential disease-causing variant in a putative recessive condition or a previously reported PIDD-causing variant, but an unexpected or extended phenotype, or potential causative variants were detected in OMIM genes not previously reported in PIDD. No potential causative variants identified in 19 families. In 6 out of 60 families more than one PIDD-causing gene variant was involved, i.e. one family with 3 affected males with X-linked *FANCB* mutation combined with X-linked *SH2D1A* mutation. This study shows that exomeSeq is an efficient method to detect disease-causing variants in a large set of candidate genes. Combining data from exomeSeq and aCGH has proven useful and important to identify PIDD causing variants.

77

Absence of heterozygosity accompanying complex human genomic rearrangements: further evidence for replicative mechanisms. *C.M.B.C Fonseca[1,2], R. Pfundt[3], L.W. Zuccherato[1], P. Liu[1], P. Stankiewicz[1], C.W. Brown[1], C.A. Shaw[1], G. Ira[1], P.J. Hastings[1], H.G. Brunner[3], J.R. Lupski[1,4,5].* 1) Dept Molecular Human Genetics, Baylor Col Medicine, Houston, TX; 2) Centro de Pesquisas René Rachou - FIOCRUZ, Belo Horizonte, MG, Brazil;; 3) 3Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 4) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 5) Texas Children's Hospital, Houston, TX.

Complex genomic rearrangements (CGRs) consist of more than one simple rearrangement with at least two breakpoint junctions formed during the same mutational event. They can consist of deletions, duplications and triplications as well as inversions and complex combinations thereof. Replicative models, such as break-induced replication (BIR), microhomology-mediated break induced replication (MMBIR) and fork stalling and template switching (FoSTeS), feature prominently, but the extent to which they contribute to disease etiology via gene dosage effects, gene interruption at breakpoint junctions, or other mechanisms remains to be unraveled. These replicative mechanisms can more parsimoniously explain the experimental data we observe associated with such complexities. We recently studied four cases of complex rearrangements constituted by duplications interspersed with triplications and associated with absence of heterozygosity (AOH) in the genomic interval distal to the CGR and continuing to the telomere on the same chromosome. These regions of AOH were from 6 Mb to 50.6 Mb in length. Extensive AOH was shown to occur in BIR in yeast if the broken end invades and copies a homologue instead of a sister molecule; it was also predicted to occur in the MMBIR model. We now provide experimental evidence that in humans, complex rearrangements generated postzygotically can lead to regional uniparental disomy (UPD) and that replication-based mechanisms may underlie formation of diverse types of genomic alterations implicated in both constitutional disorders and cancer.

78

Individual gene disruptions from balanced chromosomal rearrangements define novel neurodevelopmental loci and genomic disorders. *H. Brand[1,4], V. Pillalamarri[1], I. Blumenthal[1], M. Stone[1], S. Pereira[2], C. Morton[2,3,4], J. Gusella[1,4,5], M. Talkowski[1,4,5].* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Departments of Obstetrics, Gynecology, and Reproductive Biology , Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 3) Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 4) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 5) Departments of Psychiatry, Neurology, and Genetics, Harvard Medical School, Boston, MA.

Cytogenetically defined balanced chromosomal aberrations (BCAs) represent substantial contributors to congenital anomalies but have been only routinely detected at extremely low resolution by karyotyping as deep, high resolution whole-genome sequencing (WGS) is cost prohibitive. Recent studies from our laboratory have delineated BCAs at high resolution using a large-insert WGS approach to derive nucleotide resolution of BCA breakpoints at a cost comparable to conventional cytogenetic methods. We previously reported sequencing of 38 subjects with a neurodevelopment disorder (NDD) and found several disruptions in novel genes with a strong effect on neurodevelopment (Talkowski et al., 2012, Cell). Herein we report sequencing 56 independent subjects with congenital anomalies harboring a karyotypically identified BCA. We discovered complex rearrangements (>3 breakpoints) in 15 subjects, representing ~10-fold increase over cytogenetic estimates. Of the 41 cases with canonical (non-complex) BCAs we found disruption of 27 genes among 26 subjects, many of which represent likely pathogenic loci. In DGAP055, sequencing identified disruption of CDK6, a locus within a 7q21 microdeletion syndrome associated with mental retardation, microcephaly, dysmorphism, and short stature; no causative gene has been identified. This disorder includes many symptoms overlapping with DGAP055's mental retardation, microcephaly, and short stature. Furthermore, CDK6 is a strong candidate for a causal role in short stature based on compelling evidence from GWAS association of CDK6 variants with height (rs2282978; p=7.8E-23). Two other genes (CACNA1C, CTNND2) demonstrate the variable expressivity of NDD even in known 'causal' loci, as mutations in these genes cause Timothy syndrome and cri-du-chat syndrome, respectively, but neither of these subjects presented with hallmark symptoms of the respective disorder. Another intriguing class of loci involves genes disrupted in subjects with a complex NDD that are also associated with other psychiatric disorders: TRANK1 (bipolar disorder), DOCK9 (bipolar), TCF4 (schizophrenia), and ZNF804A (schizophrenia). This study emphasizes the significance of cytologically visible chromosomal abnormalities as a unique resource for defining strong effect mutations with a significant impact in human developmental disorders, and argues for widespread adoption of methods to delineate these events in routine genetic studies.

79

A novel treatable disorder of protein glycosylation; phosphoglucomutase 1 deficiency. *E. Morava[1], T. Marquardt[2], J. Cegielska[3], K. Raymond[4], C. Stanley[5], D. Lefeber[6].* 1) Dept. of Pediatrics; Human Genetics Center, Tulane University, New Orleans, LA, US; 2) University of Munster, Department of Pediatrics, Munster, Germany; 3) University of Warsaw, Department of Pediatrics, Division of Metabolic and Endocrine Disorders, Poland; 4) Mayo Clinics, Mayo Laboratories, Rochester, NY, US; 5) The Childrens' Hospital of Philadelphia, Philadelphia, US; 6) Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands.

We discovered a novel inborn error of glycosylation; phosphoglucomutase (PGM1) deficiency. Patients presented with cleft palate or uvula, hypoglycemia, endocrine abnormalities, cardiomyopathy and normal intelligence. The transferrin isoform pattern was pathognomic with decreased galactosylation, confirmed by mass spectrometry. PGM1 is a key enzyme between glycolysis and glycogenesis catalyzing the bidirectional transfer of phosphate from position 1 to 6 of glucose, leading to the production of Gluc-6-P. Gluc-1-P is connected to galactose metabolism, while UDP-glucose and UDP-galactose are essential activated sugars for normal protein glycosylation. The biochemical results of decreased glycosylation led us to evaluate the galactose intake in 8 patients with PGM1 deficiency. The amount of dietary galactose intake varied amongst patients, as did the degree of hypoglycosylation, determined by transferrin analysis. Patients with low galactose intake (less then 0.4g/kg/day) had increased tri-, di-, mono- and asialotransferrin at the time of diagnosis and had endocrine dysfunction. Variability in disease severity was hypothesized due to an altered balance between Gluc-1-P and Gal-1-P concentrations. Six patients consecutively underwent dietary intervention receiving either milk, or in case of milk aversion, oral lactose or galactose powder, according to WHO recommendations. After increasing galactose intake to the age appropriate range in patients, the transferrin pattern improved and galactosylation increased. In two cases, using an oral galactose powder supplement, liver function tests and endocrine dysfunction normalized in a year period. Patients on age appropriate milk intake and/or oral lactose supplement showed a significant improvement of secretory protein glycosylation, however no full restitution. One patient on lactose-supplemented diet needed diazoxide treatment due to recurrent hypoketotic hypoglycemia. We propose that PGM1 deficiency is a treatable subtype of the congenital disorders of glycosylation. We suggest that patients with galactose supplementation have a better clinical and biochemical outcome, compared to patients using dietary lactose. Since abnormal glycosylation in PGM1 defect is most likely the consequence of a decrease in the relative availability of the galactose building-units, necessary for the oligosaccharide chains of glycans, we propose age appropriate galactose supplementation in PGM1 deficient patients.

80

Mutations in *HCFC1* a transcriptional coregulator causes a novel X-linked cobalamin disorder (*cblX*) with a severe neurological phenotype. *T.H. Shaikh[1,2], H.C. Yu[1], J.L. Sloan[3], G. Scharer[1], A. Brebner[4], A. Quintana[1], N.P. Achilly[3], I. Manoli[3], C.R. Coughlin[1], E.A. Geiger[1], U. Schneck[1], D. Watkins[4], J.L. VanHove[1], B. Fowler[5], M.R. Baumgartner[5,6], D.S. Rosenblatt[4], C.P. Venditti[3].* 1) Department of Pediatrics, Section of Genetics University of Colorado School of Medicine, Aurora, CO 80045, USA; 2) Colorado Intellectual and Developmental Disabilities Research Center (IDDRC), University of Colorado School of Medicine, Aurora, CO 80045, USA; 3) Genetics and Molecular Biology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA; 4) Department of Human Genetics, McGill University, Montreal, Quebec H3A 1B1, Canada; 5) Division of Metabolism, Chidren's Research Center, University Children's Hospital, Zürich 8032 , Switzerland,; 6) Zürich Center for Integrative Human Physiology, University of Zürich, Zürich 8057, Switzerland.

Derivatives of vitamin B12 (cobalamin) are essential cofactors for enzymes required in intermediary metabolism. Defects in cobalamin metabolism lead to disorders characterized by the accumulation of methylmalonic acid and/or homocysteine in blood and urine. Combined methylmalonic acidemia and hyperhomocysteinemia cblC type (*cblC*) is one of the most common of the inborn errors of cobalamin metabolism, caused by mutations in *MMACHC*. However, several patients with *cblC* confirmed by complementation analysis lack mutations in *MMACHC*, suggesting genetic heterogeneity within this complementation group. We used exome sequencing to identify the genetic basis of a novel, X-linked form of combined methylmalonic acidemia and hyperhomocysteinemia, designated *cblX*. A missense mutation in a global transcriptional coregulator, *HCFC1*, was identified in the index patient. Additional male subjects were ascertained through two international diagnostic laboratories and 14/18 had one of 5 distinct missense mutations affecting three highly conserved amino acids (Gln68, Ala73, Ala115) within HCFC1 Kelch domains. A common phenotype of severe neurological symptoms (intractable epilepsy, profound neurocognitive impairment) and mild biochemical manifestations compared to early onset cblC patients was observed in all affected subjects. In two patient fibroblast lines, *MMACHC* mRNA and protein expression was severely reduced while HCFC1 protein levels remained intact. Furthermore, siRNA knockdown of the *HCFC1* resulted in the coordinate down-regulation of *MMACHC*. This distinct X-linked disorder highlights a novel disease mechanism by which transcriptional dysregulation leads to an inborn error of metabolism with a complex clinical phenotype.

81

Prediction of phenotypes and tetrahydrobiopterin-responsiveness in phenylketonuria using data from the genotypes and locus-specific databases. *N. Blau[1], S. Wettstein[2], W.W. Yue[3], J. Underhaug[3], B.D. Marsden[4], A. Martinez[4], A. Honegger[5], B. Perez[6].* 1) Div Metabolic Disorders, Univ Children's Hosp, Heidelberg, Germany; 2) Div Metabolism, Univ Children's Hospital, Zürich, Switzerland; 3) Structural Genomics Consortium, Univ Oxford, Oxford, UK; 4) Dept of Biomedicine, Univ Bergen, Bergen, Norway; 5) Univ Zürich, Dept Biochemistry, Zürich, Switzerland; 6) Dept Mol Biology, CSICUAM, Univ Autonoma, Madrid, Spain.

Background: Management of phenylketonuria (PKU) patients depends on the individual phenotype. The variability in the metabolic phenotypes in PKU is caused by different mutations within the PAH gene and thus residual phenylalanine hydroxylase (PAH) activity. In addition, it has been shown that genotypes are useful in predicting cofactor tetrahydrobiopterin (BH4; sapropterin) responsiveness in PKU. **Objectives:** To analyze data from available PKU-associated databases (locus-specific *PAH*vdb and genotypes BIOPKU) and to establish algorithms for genotype-phenotype correlation and BH4-responsiveness prediction. **Methods:** First, the relative frequencies of mutations, genotypes, affected gene regions and protein domains were calculated. Subsequently, PAH mutations and genotypes were scored using data from FoldX (protein damage algorithm that uses an empirical force fields), SIFT (protein function prediction based on the degree of conservation of AA residues), Polyphen2 (impact of an AA substitution on the protein structure and function), SNPs3D (molecular functional effects of non-synonymous SNPs based on structure and sequence analysis), and Rosetta ddG (impact of a sequence change on a protein's stability) prediction tools. The 3D atomic environment of each mutation was visualized using the interactive iSee concept. The *PAH*vdb database (833 variations; www.biopku.org) and BIOPKU database (4181 PKU patients with full genotype; www.biopku.org) were used. **Results/Discussion:** Amongst the 4181 patients (15,1% HPA, 24,4% mild PKU, 41,3% classic PKU, 19,2% no information) we observed 463 different mutations. The most frequently affected sites were exon 7 (22,9%) and intron 10 (9,4%), and the most affected PAH region was the catalytic domain (60,8%), followed by the regulatory (14,0%) and tetramerization domain (4,9%). c.1066-11G>A / c.1066-11G>A was the most frequent genotype (3,3%). BH4-responsiveness data were available from 2128 patients (44,4% responders). Using genotype scoring both the phenotype and BH4-responsiveness was estimated, offering a robust method for patients' characterization and management.

82

Three apparent pseudo-deficiency alleles in the *IDUA* gene identified by newborn screening. *L.M. Pollard[1], S.R. Braddock[2], K.M. Christensen[2], D.J. Boylan[2], L.D. Smith[3], B.A. Heese[3], A.M. Atherton[3], C.E. Lawson[3], M.E. Strenk[3], M. Willing[4], L. Manwaring[4], T.C. Wood[1].* 1) Biochemical Genetics Laboratory, Greenwood Genetic Center, Greenwood, SC; 2) Division of Medical Genetics, SSM Cardinal Glennon Children's Medical Center, St. Louis, MO; 3) Division of Genetics, Children's Mercy Hospitals and Clinics, Kansas City, MO; 4) Department of Pediatrics, Washington University School of Medicine, St. Louis, MO.

Pseudo-deficiency has been reported for several lysosomal enzymes in which a phenotypically unaffected patient's cells demonstrate deficient activity for a specific enzyme using an artificial substrate in vitro. Examples of this phenomenon are rare because they are typically identified in family members being evaluated for carrier status. The recent initiation of newborn screening for several lysosomal storage disorders has resulted in testing large numbers of unaffected individuals for these conditions. Our laboratory has received samples from 14 infants with an abnormal newborn screen for alpha-iduronidase, the enzyme responsible for Hurler syndrome (MPS I), for diagnostic enzyme analysis using a 4-methylumbelliferone (4-MU) substrate. Of these 14 cases, five had normal alpha-iduronidase activity (> 6 nmol/hr/mg) and one is affected with MPS I. However, the remaining eight infants had alpha-iduronidase activities below normal, but near or above the affected cut-off (2 nmol/hr/mg). Molecular analysis of these patients revealed three common sequence alterations in IDUA: c.235G>A (p.A79T), c.246C>G (p.H82Q) and c.965T>A (p.V322E). Three African American infants are homozygous for p.A79T, which has only been observed in African Americans, with an allele frequency of 2.8%. Two African American infants are heterozygous for both p.A79T and p.V322E; the latter is predicted to be damaging by both PolyPhen and SIFT, and has an allele frequency < 1% in both African and European Americans. Additionally, one African American infant is heterozygous for p.A79T and a p.D223N change, which is only detected in African Americans, with an allele frequency <1%. Finally, p.H82Q, with an allele frequency <1% in both European and African Americans, was detected in two infants: homozygous in a Caucasian infant, and heterozygous in compound with p.V322E in a biracial infant. It is too early to rule out MPS I in these patients based on clinical features; however, of the six who have had urine studies performed, all had normal chromatography/electrophoresis. We will perform enzyme analysis for these eight patients using a tandem mass spectrometry substrate to determine if the proposed pseudo-deficiency is specific to the 4-MU substrate used in our diagnostic assay. These findings indicate that pseudo-deficiency is a common phenomenon for alpha-iduronidase, which could have a negative impact on newborn screening for MPS I.

83

Atherosclerosis and glycosaminoglycan metabolism defects: Frequent association of endothelial dysfunction in patients with mucopolysaccharidosis. *S. Yano[1], K. Moleley[1], L. Wong[1], C. Castelnovi[1], C. Azen[2], Z. Pavlova[3].* 1) Genetics/Pediatrics, University of Southern California, Los Angeles, CA; 2) Clinical Trials Unit Children's Hospital Los Angeles, University of Southern California, Los Angeles, CA; 3) Pathology, Children's Hospital Los Angeles, University of Southern California, Los Angeles, CA.

Cardiovascular lesions including coronary stenosis are frequently associated in patients with genetic defects of glycosaminoglycan (GAG) metabolism. Atherosclerosis is the result of a complex interaction between lipoproteins, extracellular matrix consisting of proteoglycans, and cells of the vessel wall. Inflammatory response of T cells to GAGs is involved in atherogenesis that is preceded by endothelial dysfunction (ED) as a key event. Decreased nitric oxide (NO) bioavailability is considered the hallmark of ED. Reduced endothelial NO synthase (eNOS) has been reported in atherosclerotic arteries. Histopathologic similarity between the atherosclerotic changes in adults and in patients with GAG metabolism defects has been known although reduced eNOS has not been reported in GAG metabolism defects. Although it is well recognized that early diagnosis of coronary artery (CA) involvement in patients with Mucopolysaccharidosis (MPS) is of extreme importance for proper clinical management, diagnosis of early CA changes is difficult. Even CA angiography has limited value since it can underestimate the extent of the stenotic vascular lesions. Hence, the prevalence of CA lesions in patients with MPS is not well known. Impairment in reactive hyperemia-digital peripheral arterial tonometry (RH-PAT) with EndoPAT has been validated to correlate coronary microvascular function in patients with atherosclerosis. RH-PAT is thought to reflect endothelial NO production. Objective: (1) To evaluate eNOS in the CA specimens from a patient with GAG metabolism defect to confirm presence of the pathophysiology similar to the atherosclerotic changes characterized by ED. (2) To evaluate endothelial function in patients with various GAG metabolism defects. Methods and Results: (1) Immunohistological staining revealed decreased eNOS in the CA in a 3-year-old patient with MPS-I who died due to ischemic heart attack. (2) Evaluation of endothelial function with EndoPAT in 30 patients with GAG metabolism defects revealed a significantly high incidence of ED compared with 12 controls. Conclusions: Decreased eNOS in the MPS-I patient suggested the common pathology in GAG metabolism defects and in atherosclerosis, and a high incidence of ED in patients with GAG metabolism defects diagnosed by RH-PAT with EndoPAT may suggests that this method can provide vital information of CA lesions in the clinical management of patients with GAG metabolism defects.

84

First results of a 6 month, open label, phase I/II clinical trial of intrathecal (IT) enzyme replacement therapy (ERT) and its extension in mucopolysaccharidosis IIIA (MPSIIIA, Sanfilippo syndrome) patients. *C. Breen[1], P. Haslett[2], F.A. Wijburg[3], J. de Ruijter[3], J.P. Marchal[3], F. Heap[4], S. Rust[4], K. Baez[2], N. Nair[2], S.A. Jones[1].* 1) Manchester Centre for Genomic Medicine, University of Manchester and Central Manchester University Hospitals, Manchester, United Kingdom; 2) Shire, Lexington, MA, USA; 3) Academic Medical Center, Amsterdam, The Netherlands; 4) Royal Manchester Children's Hospital, Manchester, United Kingdom.

Introduction: MPSIIIA is a lysosomal storage disease caused by a deficiency of heparan *N*-sulfatase, which results in accumulation of heparan sulfate (HS) leading to progressive neurodegeneration. HGT-1410 is an investigational recombinant heparan N-sulfatase that is under study as ERT for the treatment of MPSIIIA. We report results from the first-in-human trial (NCT01155778) of HGT-1410 and preliminary data from its extension (NCT01299727). Methods: HGT-1410 (10 mg, 45 mg, 90 mg) was administered via IT drug delivery device (IDDD) every 4 weeks in patients aged ≥3 years (developmental age ≥1 year). Primary objective: assessment of safety and tolerability of IT-administered HGT-1410. Secondary objectives included: effect of therapy on HS levels in CSF, cognitive status (developmental quotient; DQ) and cortical gray matter volume (MRI) assessed every 6 months. Results: Twelve patients were enrolled, 6 in the Netherlands and 6 in the UK. Median age was 5.7 years (range 3.1 to 23.6) and 11 continue in the ongoing extension study. The patients were heterogeneous with respect to age, disease stage and phenotype. In the initial trial, there were 10 serious adverse events (SAEs) among 7 patients; 9 of them related to the IDDD. No SAEs were considered related to HGT-1410. No patients discontinued from the study. Among the study patients, CSF levels of HS were increased at baseline relative to age-matched non-MPS controls. CSF levels of HS exhibited marked and persistent declines following the first dose of IT HGT-1410. Response profiles among the 45-mg and 90-mg treatment groups appeared superior to the 10-mg group. HGT-1410 had no discernible effect on decline in DQ or on cortical grey matter volume, observed over 6 to 24 months, when compared to the data of an ongoing natural history study in MPSIIIA patients (NCT01047306). Conclusions: This initial, relatively short-term observation suggests that HGT-1410 was biologically active and generally well tolerated after IT delivery to children with MPSIIIA. Mechanical failures of the IDDD present challenges that may be particular to this patient population. These preliminary data do not demonstrate an effect of IT-administered HGT-1410 on cognitive decline or loss of cortical gray-matter volume in children with the severe form of MPSIIIA. However, this phase I/II trial was not designed to test the efficacy of the therapy. Further clinical studies of this new therapeutic approach for MPSIIIA are warranted.

85

Development of AAV8-mediated gene therapy clinical trial for Crigler-Najjar syndrome type I: optimization of liver-specific expression cassette. *N. Pastore[1], E. Nusco[1], A. Auricchio[1,2], N. Brunetti-Pierri[1,2].* 1) Telethon Institute of Genetics and Medicine, Naples, Italy; 2) Department of Translational Medicine, Federico II University of Naples, Italy.

Crigler-Najjar syndrome type I is a severe inborn error of bilirubin metabolism due to deficiency of the liver-specific UDP-glucuronosyltransferase encoded by the UGT1A1 gene. Current therapy relies on phototherapy to prevent life-threatening elevations of serum bilirubin but liver transplantation is the only permanent treatment. Gene therapy has the potential to provide a safer and a definitive cure for this severe disease. Adeno associated viral (AAV) vectors derived from serotype 8 have recently shown very promising results for liver-directed gene therapy of patients with hemophilia B. These encouraging results may pave the way towards AAV8-mediated gene therapy for inborn errors of liver metabolism, such as Crigler-Najjar syndrome type I. As a first step towards the development of a clinical trial for this disorder, we focused on the design and optimization of the AAV8 expression cassette. Towards this goal, we injected intravenously the Gunn rats, the animal model of Crigler-Najjar syndrome type I, with an AAV8 vector expressing the human UGT1A1 under the control of the liver-specific thyroxine-binding globulin (TBG) promoter. A significant and sustained 60% reduction of baseline serum bilirubin levels was detected only at the high dose of 7.4×10e13 genome copies (gc)/kg. Next, we achieved a 50% reduction of baseline serum bilirubin with 2.7×10e13 gc/kg of a vector encoding the UGT1A1 under the control of the synthetic, liver-specific LP1 promoter. Interestingly, the injection of another vector bearing the codon optimized UGT1A1 cDNA (cohUGT1A1) under the control of the LP1 promoter resulted in complete normalization of serum bilirubin levels at the further lower dose of 1.1×10e13 gc/kg. This vector also resulted in higher levels of hepatic UGT1A1 protein normalized for the vector genome copy number of injected animals. In summary, our study shows that the AAV8-LP1-cohUGT1A1 results in higher hepatic transgene levels and sustained correction of hyperbilirubinemia in the Gunn rats. These pre-clinical data may form the basis for the development of a gene therapy trial for liver-directed gene therapy of Crigler-Najjar syndrome type I.

86

Identification of chemical and pharmacological chaperones to treat Zellweger Spectrum patients with the common allele, PEX1-Gly483Asp. *N.E. Braverman[1], S.J. Steinberg[2], S. Heibler[2], G.E. Maclean[1].* 1) Human Genetics, McGill University, Montreal, Quebec, Canada; 2) Neurogenetics, Kennedy Krieger Institute, Baltimore, Maryland, USA.

Zellweger spectrum disorder (ZSD) results from recessive defects in any one of 13 PEX proteins, which are required for peroxisome biogenesis. ZSD causes progressive multisystem dysfunction. One mutation in PEX1, a AAA ATPase required for peroxisome matrix enzyme import, Gly843Asp (G843D), represents up to 40% of all ZSD alleles. Thus, identifying treatments for this allele will benefit many patients. We recently showed that PEX1-G843D behaves as a misfolded protein whose function can be improved by chemical and pharmacologic chaperones. In a small molecule screen, we reported recovery of peroxisome matrix enzyme import by flavonoid compounds. Flavonoids are known to bind ATPases, and thus could bind specific regions of PEX1-G843D, such as the ATP binding domain, to improve conformation. In order to develop a lead, we evaluated 70 flavonoids using our cell based phenotype assay. We used a PEX1-G843D/null patient fibroblast cell line, expressing a GFP-PTS1 reporter. Functional recovery of peroxisome matrix enzyme import was determined visually by scoring the number of cells with peroxisomal vs. cytosolic GFP fluorescence. We used a second, independent assay to confirm these results. SAR was determined for the compounds, and diosmetin was found to be the most effective, reaching 75-100% recovery of the GFP reporter into the peroxisome at 10 μM. Interestingly, our group of effective drugs included 5 related flavones and 4 of their corresponding flavonols. Structure comparisons showed that methoxylation or hydroxylation at the 4' position and hydroxylation at the 5 and 7 positions increased efficacy, flavones were more effective than the flavonols, and flavanones, isoflavones, and chalcones were ineffective. Additional comparisons showed that a set of modifications were uniquely best for our specific target, possibly allowing the drug to bind PEX1 reversibly. Finally, combination therapy with a chemical chaperone, betaine, allowed us to achieve equivalent recovery of import using 5 μM diosmetin. In summary, we have identified a target for pharmacological chaperone therapy that captures a large group of children with a rare genetic disease. We have also identified a group of potential pharmacological chaperones that could be uniquely optimized once their mechanism of action is validated. As our assay measures downstream recovery of peroxisome matrix enzyme import, it has the potential to identify additional compounds with a variety of mechanisms of action.

87

Efficacy of hematopoietic cell therapy in X-linked adrenoleukodystrophy: a multinstitutional study (ALD-101). *G. Raymond[1], P. Orchard[2], P. Aubourg[3], M. Escolar[4], J. Kurtzberg[5], S. Paadre[6], J. Balser[6].* 1) Department of Neurology, University of Minnesota, Minneapolis, MN; 2) Department of Pediatrics, University of Minnesota, Minneapolis, MN; 3) University Paris Descartes, Paris France; 4) Program for the Study of Neurodevelopment in Rare Diseases, Children's Hospital of Pittsburgh, Pittsburgh, PA; 5) Pediatric Blood & Marrow Transplant Program, Duke University Medical Center, Durham, N.C; 6) Veristat, Holliston, MA.

Objective: The only therapy for childhood cerebral adrenoleukodystrophy (CCALD) is hematopoietic cell therapy (HCT), but there is limited outcome information compared to untreated boys. Methods: We conducted a retrospective study (ALD-101) to characterize subjects with untreated CCALD and collect efficacy and safety data from HCT treated boys. Data was collected on 136 cases (72 untreated/ 65 HCT) from diagnosis till either 2 years post-diagnosis or death from 5 centers, 4 in the US and 1 in France. Established measures of neurologic function (NFS) and MRI (Loes) were used in all cases. Results: In the untreated, 70 of 72 (97%) had at least one NFS score and an MRI; 30 (42%) with gadolinium (23 Gad+/7Gad-). Enhancement was highly predictive of rapid progression. Of the Gad+, 19 had more than one NFS score recorded and the majority showed significant decline in 6-18 months and no resolution of enhancement in the untreated group. In the 65 HCT-treated boys, all were evaluated with NFS and MRI with contrast. In the treated cohort there was resolution of enhancement (median 3.4 months) and stabilization of MRI and NFS. Engraftment failure occurred in 18.5% and the rate of severe acute and chronic GVHD was 11% and 5% respectively. The highest incidence of death occurred in those boys undergoing an HLA mismatched , non related transplant (12/32; 37.5%). Conclusions: We report here the largest retrospective, multi-institutional study of untreated and treated CCALD, demonstrating that MRI enhancement is predictive of progression, and that it rapidly resolves following HCT. Successful HCT improved all measures of disease.

88

Rare mutations in RINT1 predispose carriers to early-onset breast cancer. *D.J. Park[1], K. Tao[2], F. Le Calvez-Kelm[3], T. Nguyen-Dumont[1], N. Robinot[3], F. Hammet[1], F. Odefrey[1], H. Tsimiklis[1], Z.L. Teo[1], L.B. Thingholm[1], C. Voegele[3], A. Lonie[4], B.J. Pope[4], E.M. John[5,6], I.L. Andrulis[7], M.B. Terry[8], M. Daly[9], S. Buys[10], G.G. Giles[11], J.L. Hopper[12], D.E. Goldgar[13], F. Lesueur[3], S.V. Tavtigian[2], M.C. Southey[1], Breast Cancer Family Registry, kConFab.* 1) Genetic Epidemiology Laboratory, The University of Melbourne, Victoria 3010, Australia; 2) Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT 84112, USA; 3) Genetic Cancer Susceptibility Group, International Agency for Research on Cancer, 69372 Lyon, France; 4) Victorian Life Sciences Computation Initiative, Carlton, Victoria 3010, Australia; 5) Cancer Prevention Institute of California, Fremont, CA 94538, USA; 6) Department of Health Research and Policy, Stanford Cancer Center Institute, Stanford, CA 94305, USA; 7) Department of Molecular Genetics, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada; 8) Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY 10032, USA; 9) Fox Chase Cancer Center, Philadelphia, PA 19111, USA; 10) Huntsman Cancer Institute, University of Utah Health Sciences Center, Salt Lake City, UT 84112, USA; 11) Centre for Cancer Epidemiology, The Cancer Council Victoria, Carlton, Victoria 3052, Australia; 12) Centre for Molecular, Environmental, Genetic, and Analytical Epidemiology, School of Population Health, The University of Melbourne, Victoria 3010, Australia; 13) Department of Dermatology, University of Utah School of Medicine, Salt Lake City, UT 84132, USA.

Whole-exome sequencing analysis of germline DNA from 93 people from 49 early-onset, multiple-case breast cancer families identified three mutations in *RINT1* (NM_021930), each occurring in a separate family, and not present in the public sequencing databases: p.Q115X, p.M378del and p.D403Y. The latter two variants occurred at protein positions that are evolutionarily conserved. In the family harbouring p.Q115X, the mutation was carried by all of the four women who were affected with unilateral breast cancer and by a male relative affected by bladder and lung cancers.

RINT1 was originally identified as a RAD50-interacting protein. It has since been demonstrated that RINT1 functions as a tumour suppressor that maintains Golgi dynamics and centrosome integrity. In addition, overexpression of truncated RINT1 has been shown to result in a defective radiation-induced G2/M checkpoint. Mice carrying only one functional copy of *Rint1* spontaneously develop a variety of tumours including mammary adenocarcinomas at a combined rate of 81% (Lin *et al*., 2007), which is higher than the rate at which *Brca1*$^{+/-}$ mice spontaneously develop tumours (Jeng *et al*., 2007).

Screening an additional 798 multiple-case breast cancer families identified six occurrences (each in different families) of very rare genetic variants (minor allele frequency (MAF) <0.5% in all public sequencing database populations): p.A138V (2), p.R256Q (1), p.P419R (1) and p.S462L (1), all predicted to be possibly pathogenic by PolyPhen2 or Align-GVGD *in silico* analysis; and c.1333+1G>A (1), which affects a splice donor core dinucleotide.

To further assess the association of rare *RINT1* genetic variants with breast cancer predisposition, we screened a population-based sample of 1313 women with early-onset breast cancer and 1123 age and ethnicity-matched controls. A total of 23 breast cancer cases were found to be carriers of rare (MAF < 0.5%), possibly pathogenic variants (in-frame insertion/deletion (2) or Align-GVGD grade >C0 missense variants (21)), compared with 6 controls found to carry variants that fulfilled these criteria (OR=3.32, 95%CI (1.31, 10.00); Fischer's Exact Test P-value for the null hypothesis= 0.0040). These data indicate that rare, evolutionarily unlikely genetic variants in *RINT1* are associated with increased risk for early-onset breast cancer.

89

The BER glycosylase *NEIL1* is a risk gene for familial breast cancer. *M.R. Dufault[1], E. Hahnen*[2,3], H. Hellebrand[1], J. Hauke[2,3], G. Neidhardt[2,3], S. Engert[1], S. Preissler-Adams[5], N. Arnold[6], T.M. Strom[7], K. Rhiem[2,3], B. Wappenschmidt[2,3], N. Ditsch[4], R.K. Schmutzler#[2,3], A. Meindl#[1].* 1) Department of Gynaecology and Obstetrics, Klinikum rechts der Isar at the Technical University, Munich, Germany; 2) Center of Familial Breast and Ovarian Cancer, University Hospital of Cologne, Cologne, Germany; 3) Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany; 4) Klinikum Großhadern, Ludwig-Maximilians-Universität, Munich, Germany; 5) Institut of Human Genetics at the University of Münster, Germany; 6) Department of Obstetrics and Gynaecology at the University of Kiel, Germany; 7) Institute of Human Genetics, Klinikum rechts der Isar, Technical University, Munich, Germany.

* contributed equally as first author; # contributed equally as corresponding authors

Introduction: Approximately 70% of familial breast cancer susceptibility remains unexplained. Over the past decade, several genes apart from *BRCA1* & *BRCA2* have been identified as breast cancer susceptibility genes including the genes *RAD51C*, *BRIP1*, *PALB2, ATM*, CHEK2. While these genes were identified using a candidate gene approach, more recently, *XRCC2*, *FANCC* & *BLM* have all been identified using an exome sequencing approach. A common characteristic of all of these genes is their involvement in the Homologous Recombination (HR) DNA repair pathway. **Methods and Results:** We performed exome sequencing of families with a strong history of breast cancer in at least two affected family members and identified a deleterious frameshift mutation in the *NEIL1* gene (c.572insC; p.P192Afs*51). By screening familial cases & population matched controls we identified three distinct frameshift mutations in 10 out of 950 families compared with eight out of 3165 controls (OR = 4.16, 95%1 CI = 1.47 - 12.17, P = 0.0030). Furthermore, we found five putative pathogenic variants in 13 out of 950 families which were only present in eight out of 3165 controls (OR = 5.41, 95%1 CI = 2.07 - 15.11, P = 0.0002). Overall, we identified *NEIL1* mutations in 23 out of 950 families (2.42 %) compared to sixteen out of 3165 controls (0.51 %). This 4.79 -fold increase in risk clearly points to *NEIL1* being a novel moderately penetrant gene for familial breast cancer. Segregation analysis was performed for the truncating mutations and in most instances, the mutation tracked with the phenotype. Similarly, segregation could be shown for the missense mutation G326A. **Discussion:** NEIL1 is a Base Excision Repair (BER) protein that responds early to DNA damage caused by oxidative stress. In addition, it also responds quickly to interstrand crosslinks (ICL). Inactivating mutations in the *NEIL1* gene have been linked to human gastric and colorectal cancer and promoter methylation in head and neck squamous cell carcinoma. Consistent with its role in colorectal cancer, two cases of colon cancer were present in pedigrees that harbored *NEIL1* mutations. In other families, pancreatic cancer was present. Our data identified *NEIL1* as the first gene of the BER pathway associated with a risk of familial breast cancer.

90

More than 25% of breast cancer families with wild-type results from commercial genetic testing of BRCA1 and BRCA2 are resolved by BROCA sequencing of all known breast cancer genes. *T. Walsh[1], S. Casadei[1], M.K. Lee[1], A.M. Thornton[1], G. Bernier[1], C. Spurrell[1], J. Mandell[1], T. Lajus[2], E. Swisher[1], M-C. King[1].* 1) Medical Genetics, University of Washington, Seattle, WA; 2) Coordenadora de Pesquisa Translacional Liga Contra o Câncer, Brazil.

A challenge to the present practice of genetic testing for inherited risk of breast cancer is how to explain the illness in breast cancer patients with severe family history, but negative (wild-type) results from commercial testing for *BRCA1* and *BRCA2*. To address this problem, we applied BROCA, a targeted capture sequencing approach, to identify all single base pair substitutions, insertion-deletions, and copy number variants in all known breast cancer genes. Our subjects were all living affected persons in 800 families with at least three cases of breast and/or ovarian cancer. Probands of all families had received wild-type results of commercial genetic testing of *BRCA1* and *BRCA2*. Of the 800 families, 206 (26%) were resolved by BROCA. The 206 resolved families harbored 166 different damaging germline mutations in 21 different genes. Of the resolved families, 39% (80/206) carried mutations in *BRCA1* or *BRCA2* that were not detected by commercial sequencing of the proband, either because the patient did not have comprehensive commercial large rearrangement testing in addition to commercial Sanger sequencing; or the proband was wild-type for *BRCA1* and *BRCA2*, but other cases in the family carried a *BRCA1* or *BRCA2* mutation; or the family carried a *BRCA1* or *BRCA2* mutation not reported by commercial testing (e.g. mutation of a splice enhancer, a small damaging in-frame indel, or a missense with good experimental evidence of damaging effect.) Of the resolved families, 37% (77/206) carried mutations in other breast cancer genes whose role in inherited predisposition is well characterized: 39 in *CHEK2*, 28 in *PALB2*, and 10 in *TP53*. Finally, of the resolved families, 20% (41/206) carried mutations in breast cancer genes that have been published but whose role in inherited predisposition has been less fully characterized: 15 in *ATM*, 4 in *BARD1*, 5 in *BRIP1*, 1 in *CDH1*, 2 in *ABRAXAS*, 2 in *NBN*, 3 in *RAD51C*, 5 in *RAD51D*, 1 in *STK11*, 1 in *XRCC2*. Our results indicate that families severely affected by breast cancer but with wild-type results from commercial genetic testing are well served by complete genomic sequencing of all known breast cancer genes. As such comprehensive testing becomes more widespread, it will be important to determine more precisely the risks associated with damaging mutations in each of these genes so as to incorporate them most effectively into clinical care.

91

Nine genes for inherited predisposition to breast cancer among African-American women. *O.I. Olopade[1], T. Walsh[2], Y. Zheng[1], S. Casadei[2], A.M. Thornton[2], M.K. Lee[2], M. Churpek[1], D. Huo[1], C. Zvosec[1], F. Liu[1], Q. Niu[1], J. Zhang[1], J. Fackenthal[1], M-C. King[2], J.E. Churpek[1].* 1) Dept Medicine, Univ Chicago, Chicago, IL; 2) Division of Medical Genetics, Univ Washington, Seattle, WA.

African American (AA) women are more likely to develop early onset and triple negative breast cancer than are white women. One potential contributor to these disparities may be more inherited mutations among AA women in breast cancer susceptibility genes. However, the magnitude of this effect is not known because the frequency and spectrum of inherited mutations in *BRCA1* and *BRCA2* among AA women is not well characterized. Even less is known among about the frequency and spectrum of inherited mutations in other breast cancer genes among AA women. We addressed this question using BROCA, a targeted genomic capture assay, to screen all 26 known breast cancer genes in DNA from AA women with breast cancer. BROCA detects all classes of mutations: point mutations, small insertions and deletions, and large genomic rearrangements (CNVs). For this project, we counted only unambiguously damaging mutations: truncations, complete gene deletions, splice mutations shown experimentally to lead to a mutant message, and missenses shown experimentally to lead to loss of protein function. All candidate variants were validated by Sanger sequencing. Participants in the project were 395 unrelated AA women with breast cancer ascertained through the Cancer Risk Clinic at The University of Chicago. Of the participants, 18% (71/395) carried one or more damaging mutations in nine different breast cancer genes. Mutation frequencies were: 28 in *BRCA1*, 30 in *BRCA2*, 3 in *CHEK2*, 3 in *PALB2*, 4 in *ATM*, 2 in *BARD1*, 2 in *PTEN*, 1 in *TP53*, and 1 in *BRIP1*. Three patients carried 2 mutations each. The 74 events involved 57 different mutations, 3 of which were large genomic deletions. Preliminary analysis indicates that the prevalence of damaging mutations was more than 25% among patients diagnosed at age 45 or younger, or with a close relative with either breast or ovarian cancer, or with triple negative breast cancer, or with two breast cancer primaries. Given that participants expressed interest in genetic testing, the prevalence of damaging mutations in this series is likely higher than among AA patients generally. Nonetheless, the unexpectedly high frequency in this population of damaging mutations in *BRCA1, BRCA2*, and other genes indicate that comprehensive genetic testing should be recommended and supported by health insurance for all breast cancer patients meeting National Comprehensive Cancer Network (NCCN) guidelines and for relatives of patients with damaging mutations.

92

Germline loss-of-function mutations in 15 different DNA repair genes are present in 22% of 1412 patients with ovarian, peritoneal or fallopian tube cancers not selected for age at diagnosis or family history. *M.I. Harrell[1], B. Norquist[1], T. Walsh[2], M.K. Lee[2], S. Casadei[2], K.P. Pennington[1], K.J. Agnew[1], A. Thornton[1], M.C. King[2], M.J. Birrer[3], E.M. Swisher[1].* 1) Division of Gynecologic Oncology, University of Washington, Seattle, WA; 2) Division of Medical Genetics, University of Washington, Seattle, WA; 3) Department of Medicine, Dana Farber Cancer Center, Boston, MA.

Germline mutations in *BRCA1* and *BRCA2*, and to a lesser extent in the mismatch repair genes, have heretofore been thought to explain nearly all of hereditary ovarian cancer. Recently, more ovarian cancer susceptibility genes have been identified. In order to assess the frequency and spectrum of inherited mutations in many genes, we evaluated women with ovarian, peritoneal, and fallopian tube cancers unselected for age at diagnosis or family history. Participants consisted of two groups: 577 cases prospectively enrolled at diagnosis at UW; and 835 cases enrolled in the GOG218 (n= 775) or GOG213 (n=60), randomized phase III treatment trials for advanced stage ovarian cancers. Germline DNA from all 1412 patients was sequenced using BROCA, a targeted capture and massively parallel genomic sequencing approach that detects all mutation classes. Cohort 1 was sequenced with the clinical version of BROCA, while cohort 2 was sequenced with BROCA-HR, an extended version of BROCA modified to include additional genes in the Fanconi anemia-BRCA (FA-BRCA) and related pathways. Only truncating and functionally validated missense mutations were counted. Mutation rates were similar for the two cohorts. Of all patients, 22% (304/1412) carried a damaging mutation in a gene known to be implicated in inherited ovarian or breast cancer. Of these 304 patients, 74% (226/304) carried mutations in *BRCA1* (142) or *BRCA2* (84) and only 1.6% (5/304) carried mutations in a mismatch repair gene (2 in *MSH2*, 3 in *MSH6*, 1 in *MLH1*). Germline damaging mutations in genes not generally sequenced for inherited ovarian cancer were present in 24% (73/304) of patients: 21 in *BRIP1*, 13 in *CHEK2*, 10 in *RAD51C*, 9 in *RAD51D*, 7 in *PALB2*, 4 in *TP53*, 3 in *NBN*, 3 in *BARD1*, 2 in *MRE11A*, and 1 in *XRCC2*. In addition, germline damaging mutations were detected in 15 other DNA repair candidate genes, which will be evaluated for loss of heterozygosity in tumors and, for familial cases, for co-segregation with ovarian and related cancers. In conclusion, the distribution of germline mutations in patients with ovarian carcinomas is wider than previously appreciated, and FA-BRCA genes other than *BRCA1/2* explain a large number of cases. Effective identification of hereditary risk will require comprehensive testing of many genes for all women with ovarian, peritoneal or tubal carcinoma regardless of age or family history. Cancer gene panels could do so now in a cost effective manner.

93

The Mainstreaming Cancer Genetics Programme - Integrating genetic testing into routine clinical practice in the United Kingdom. *N. Rahman[1, 2], S. Mahamdallie[1], E. Ruark[1], H. Hanson[2], I. Slade[3], A. George[1, 2], K. Snape[2], R. Sultan[1], A. Rimmer[4], M. Munz[4], G. Lunter[4], S. Banerjee[2], C. Turnbull[1, 2], Mainstreaming Cancer Genetics Consortium.* 1) Division of Genetics & Epidemiology, The Institute of Cancer Research, Sutton, United Kingdom; 2) Royal Marsden NHS Foundation Trust, London, UK; 3) Dept of Public Health, Oxford, UK; 4) Wellcome Trust Centre for Human Genetics, Oxford, UK.

Genetic testing of cancer predisposition genes is one of the major activities of clinical genetics. Currently, nearly 100 genes associated with predisposition to cancer are known, but in most countries testing is very restricted with respect to the number of genes and the number of people tested. The value of genetic testing in individuals with cancer is underappreciated; it provides important information with respect to the cause and optimal treatment of the current cancer, and the risk and optimal management of future cancer. Moreover, testing cancer patients followed by cascade testing of mutation carriers is an effective and efficient way of identifying unaffected mutation carriers in whom screening and risk-reducing strategies can be deployed. The Mainstreaming Cancer Genetics (MCG) Programme is a UK national cross-disciplinary initiative to develop the NGS assays, analytical and interpretive pipelines, clinical infrastructure, training, ethical and evaluation processes required for routine genetic testing to be integrated into cancer patient care. In collaboration with Illumina we have developed a NGS panel targeting 97 cancer predisposition genes with >1500 probes (called the TruSight cancer panel) and have completed detailed evaluation of performance, sensitivity and specificity under different levels of multiplexing, coverage and throughput. We are using bespoke analytical pipelines developed for high-throughput clinical diagnostic data analysis (called GAMA) and clinical interpretation (called CIGMA). The design principles, performance and validation data for these assays and pipelines will be presented. We are implementing a new mixed-model of cancer gene testing whereby consent for medical testing (i.e. in cancer patients) is undertaken by oncologists, with only the mutation carriers seen by geneticists. Predictive testing (i.e. in unaffected individuals) will continue to be undertaken in genetics. We have developed protocols and e-learning modules to deliver the required training for oncologists. We hope this model will allow throughput of testing to greatly increase, whilst retaining input and support from Genetics, where it is required. We are piloting the new model in ovarian and breast cancer patients, and the results will be presented. A key first aspiration of the MCG programme is to make genetic testing available to all ovarian cancer patients, 15% of whom have a germline cancer predisposition gene mutation. www.mcgprogramme.com.

94

Identification of a second major locus predisposing to an autosomal dominant inherited disorder of multiple schwannomas. *L. Messiaen[1], J. Xie[1], A.P. Poplawski[1], Y.F. Liu[1], A.R. Gomes[1], P. Madanecki[2], C. Fu[1], M.R. Crowley[3], D.K. Crossman[3], L. Armstrong[4], D. Babovic-Vuksanovic[5], A. Bergner[6], J.O. Blakeley[6], A. Blumenthal[7], M.S. Daniels[8], H. Feit[9], K. Gardner[10], S. Hurst[7], C. Kobelka[11], C. Lee[12], R. Nagy[13], K.A. Rauen[12], J.M. Slopis[8], P. Suwannarat[11], J.A. Westman[13], A. Zanko[12], B.R. Korf[1, 3], A. Piotrowski[1, 2].* 1) Dept Genetic, Univ Alabama, Birmingham, Birmingham, AL; 2) Faculty of Pharmacy, Medical University of Gdansk, Gdansk, Poland; 3) Heflin Center for Genomic Sciences, University of Alabama at Birmingham, Birmingham, Alabama, USA; 4) Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada; 5) Department of Medical Genetics, Mayo Clinic College of Medicine, Rochester, Minnesota, USA; 6) Johns Hopkins Comprehensive Neurofibromatosis Center, Baltimore, Maryland, USA; 7) Lakeridge Health Corporation, Oshawa, Ontario, Canada; 8) Clinical Cancer Genetics Program, MD Anderson Center, University of Texas, Houston, Texas, USA; 9) Henry Ford Hospital, Department of Neurology, Detroit, Michigan, USA; 10) Department of Neurology, Veteran's Administration Hospital of Pittsburgh and University of Pittsburgh, Pittsburgh, Philadelphia, USA; 11) Kaiser Permanente Genetics Northern California, San Francisco, California, USA; 12) Department of Pediatrics, Division of Medical genetics, University of California at San Francisco, San Francisco, California, USA; 13) Division of Human Genetics, Department of Internal Medicine, Ohio State University Wexner Medical Center, Columbus, Ohio, USA.

Schwannomatosis, the third major form of neurofibromatosis, is characterized by development of multiple schwannomas without bilateral vestibular schwannomas, congenital cataracts or ependymomas, manifestations typically found in neurofibromatosis type 2. Constitutional mutations in the SWI/SNF chromatin-remodeling complex gene *SMARCB1* on chromosome 22q have been found only in ~50% of familial and <10% of sporadic schwannomatosis patients, suggesting additional predisposing genes exist. Genetic analysis of schwannomas in patients with a *SMARCB1* germline mutation reveals a complex mechanism. The mutated germline *SMARCB1* allele is retained in the schwannomas, whereas part of the other chromosome 22q copy containing the wild-type *SMARCB1* and the nearby *NF2* gene is lost, followed by mutation of the remaining wild-type *NF2* gene, *in cis* with the germline mutated *SMARCB1* first-hit. We studied 20 probands (6 familial, 11 sporadic, and 3 cases with unknown family history), all without a 1st hit *SMARCB1* mutation detected in blood or schwannomas, but with a 22q deletion and a different *NF2* mutation in every schwannoma. We hypothesized that an alternative gene on chromosome 22 might carry a first hit predisposing to schwannomatosis in this group of patients. We selectively enriched for 3.72 Mb of evolutionary conserved chromosome 22 sequences and initially performed massive parallel sequencing in 8 of these unrelated *SMARCB1*-negative schwannomatosis patients. As pathogenic germline mutations in one specific gene on 22q were found in 7/8 patients, targeted sequencing of this gene was pursued in the remaining 12 patients, resulting in detection of mutations in 16/20 (80%) of patients, but not in controls. Mutations were retained in all 25 schwannomas studied and segregated with the disorder within the families. The predicted effect of the mutations on protein function adds robust evidence they are disease causing. Our findings identify this gene as a second major locus predisposing to an autosomal dominant inherited disorder of multiple schwannomas.

95

Identification of putative driver mutations in neurofibromatosis type 1 (NF1)-associated plexiform neurofibromas. *A. Pemov[1], H. Li[2], N.F. Hansen[3], J.C. Mullikin[3,4], M. Wallace[2], D.R. Stewart[1].* 1) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 2) University of Florida, Gainesville, FL; 3) Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 4) NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

BACKGROUND. NF1 is an autosomal dominant tumor pre-disposition genetic disorder, caused by constitutive inactivation of one of two copies of the tumor suppressor *NF1*. Individuals with NF1 are prone to the development of both benign and malignant tumors. Plexiform neurofibromas (PNF) are benign tumors; however, the tumors are associated with high morbidity. In addition, 8-12% of people with NF1 develop malignant peripheral nerve sheath tumor (most originating from PNFs). Although it is accepted that somatic inactivation of *NF1* is an early necessary step in development of PNF, little is known regarding which other genes are affected in the tumor. In this study, we performed whole-exome sequencing (WES) of 24 tumor samples matched with germline DNA obtained from 22 unrelated NF1 patients. **METHODS**. Germline DNA was extracted from peripheral white blood cells and tumor DNA was obtained from primary Schwann cell (SC) cultures established from dissected plexiform neurofibromas. All cell cultures were of low passage and contained at least 70% SC. We used the Illumina TruSeq Exome Enrichment Kit that targets 62 Mb of exonic sequence, including 5' UTR, 3' UTR, microRNA, and other non-coding RNA. After WES, aligning, genotype calling and filtering, we analyzed the data using statistical software (Youn & Simon, Bioinformatics, 2011, v.11, p.175) designed to identify driver genes from passengers. In addition to WES, the DNA samples were analyzed on Illumina 2.5M SNP-arrays to detect LOH regions. **RESULTS**. First, we analyzed *NF1* locus in both germline and tumor DNA. We identified pathogenic mutations in the *NF1* gene in all germline samples. In addition, we identified somatic inactivation of *NF1* in 15 out of 24 tumors. Second, we estimated the mutation burden in individual samples. There were on average 10 (range 1-55) somatic mutations per tumor. Finally, we analyzed the data to distinguish driver mutations. After correcting P-values for multiple testing we identified 9 frequently-mutated statistically significant genes. Evaluation of biological functions of the genes revealed that such processes as chromosome partition, nonsense-mediated mRNA decay and RNA-mediated silencing might play a role in PNF progression. **CONCLUSIONS**. To our knowledge, this is the first attempt to identify somatic mutations in NF1-associated PNF via WES. PNFs accumulate modest number of mutations. We identified 9 putative driver genes. The majority of the tumors contained 2-3 driver mutations.

96

Somatic Structural and Rare Germline Variation in Childhood Cancers. *D.I. Ritter[1,2], B. Powell[2,3], H. Cheung[2], R. Gibbs[1,2], D.A. Wheeler[1,2], S. Plon[2,3].* 1) Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX; 2) Baylor College of Medicine, Department of Molecular and Human Genetics, Houston, TX; 3) Baylor College of Medicine, Texas Children's Cancer Center, Department of Pediatrics, Houston, TX.

We have taken a comprehensive approach with whole genome and whole exome sequencing (WES) to understand the genetic basis of cancer susceptibility in childhood cancer patients with either a family history of cancer, second primary malignancies, congenital anomalies or learning disabilities. The sample set includes multiple expected inheritance patterns: de novo, autosomal dominant, autosomal recessive and X-linked, and patients with the following cancers: Wilms tumor, rhabdomyosarcoma, acute lymphoblastic leukemia, lymphoma, heptaoblastoma and neuroblastoma. Analysis of both somatic structural variation (SV) and loss of heterozygosity (LOH) in concert with rare germline single nucleotide variants (SNVs) and insertions/deletions (INDELS) are performed. We apply an empirical annotative filtering approach to prioritize SV in whole genome sequencing of tumor/normal pairs, combined with somatic LOH and SNVs/INDELs from WES data. To identify SV, we use complementary algorithms of insert-size (Breakdancer) and soft-clip stacking (CREST) first applied to SV seen by karyotype or arrayCGH (somatic t(9;22) and germline t(6;12), and chr17 microdeletion) as well as previously validated events from the published Lupski genome. From this, we derived an empirical filter, combined with downstream annotative filtering and clustering to identify both germline and somatic SV events. Our annotation includes: Refseq genes and gene events (intron/utr/cds), cytobands, centromere/telomere, repeats, transcription factor binding site clusters, frequency in Database of Genomic Variants and noncoding experimentally tested enhancers. Additionally, we use LOHcate, a program for tumor/normal whole-exome loss of heterozygosity, CARNAC, a somatic variant calling pipeline, and Pindel to further analyze the sequencing datasets. We pinpoint breakpoints for somatic SV, such as a complex translocation in POLR1A/UPS22/RPS9 t(2;17;19) from a Wilms tumor sample, TMEM120B/RET t(12;10) translocation in rhabdomyosarcoma with amplifications in 7q31 and 10q11.21, and a translocation in CDK12/noncoding (~25kb upstream of DHRS7) t(17;14) in neuroblastoma. We have identified multiple regions of somatic LOH and are currently preparing rare germline events within these regions for validation. With this analysis, we aim to uncover germline events associated with childhood cancer susceptibility and create a more comprehensive understanding of the etiology of cancer.

97

Frequency Uniqueness Score: Predicting the Disease Risk of Coding Variants. *A.C. Alexander[1], B.E. Engelhardt[2,3,4].* 1) Department of Computer Science, Duke University, Durham, NC; 2) Department of Biostatistics & Bioinformatics, Duke University, Durham, NC; 3) Department of Statistical Science, Duke University, Durham, NC; 4) Institute for Genome Science & Policy, Duke University, Durham, NC.

In both clinical and research applications there is an acute need for a rapid assessment of the disease risk of non-synonymous amino acid variants from whole-genome or exome sequencing data, with classification of each variant based on whether it is pathogenic or functionally neutral. For example, improved classification accuracy that leads to early identification of a Mendelian disorder can have a meaningful impact on patient prognosis, and can expedite discoveries into possible therapies by focusing efforts research efforts on smaller sets of candidate variants. Existing computational approaches to variant classification all suffer from low overall accuracy rates, with recent performance comparisons showing accuracies of less than 70% for popular tools such as Polyphen 2 and SIFT, and accuracies of approximately 80% for the best-performing tools. Their poor performance limits their utility in the determination of disease causes. Here we present an approach that represents a major departure from previous methods that have relied primarily on cross-species conservation metrics and predicted protein structure impact. Leveraging recent large-scale population studies including the 1000 Genomes Project and the NHLBI Exome Sequencing Project, we use three simple human-specific classes of features including gene variation metrics, locus variational frequency, and a metric for prior gene-disease association. We combine these metrics to predict the probability of a variant being pathogenic using a random forest classifier, which allows us to model feature interactions and provides a measure of the importance of each feature in prediction. We demonstrate that our approach substantially outperforms existing state-of-the-art methods on a variety of performance measures, with overall accuracy rates in excess of 90% on all tested data sets, and accuracy of 98% using cross-validation on the Uniprot humsavar 2011_12 data set used to validate other methods. Unlike other approaches, our method can be applied to all coding variants including indel and splice site variants across all genes, and will naturally improve over time as more comprehensive estimates of human genetic variation become available. These convincing results open the door to automated pathogenicity risk assessment and context-dependent variant classification in the clinical setting.

98

Exome-based linkage mapping and variant prioritization for inherited retinal disorders. *D.C. Koboldt[1], D.E. Larson[1], L.S. Sullivan[2], S.J. Bowne[2], R.S. Fulton[1], E. Sodergren[1], S.H. Blanton[3], K. Meltz Steinberg[1], S.P. Daiger[2], R.K. Wilson[1], G.W. Weinstock[1].* 1) The Genome Institute at Washington University, St. Louis, MO; 2) Human Genetics Center, Univ. of Texas Health Science Ctr, Houston, TX; 3) Hussman Institute of Human Genomics, Univ. of Miami, Miami, FL.

Exome sequencing in families with rare genetic disorders has the potential to rapidly identify new disease genes. Yet individual exomes harbor thousands of coding variants, and the identification of a single causal mutation among them remains a significant challenge. We developed a scoring algorithm to prioritize potential causal variants within a family according to segregation with the phenotype, population frequency, predicted effect, and gene expression in the tissue(s) of interest. We also describe two complementary approaches to exome-based linkage analysis that help narrow the search space in families with multiple affected individuals: (1) Shared IBD analysis of common variants identifies segments of maximum identity-by-descent among affected individuals; (2) Rare heterozygote rule out nominates regions based on shared rare variants and the absence of homozygous differences between affected individuals. We showcase our methods using exome sequence data from families with autosomal dominant retinitis pigmentosa (adRP), a rare disorder characterized by progressive vision loss. We performed exome capture and sequencing on 75 samples representing 24 families with probable adRP but lacking common disease-causing mutations. A subset of these families also had regions from traditional linkage mapping of extended pedigrees, the results of which were highly concordant with our linkage analyses. Seven of 24 families (29.17%) were revealed to have unrecognized mutations in known RP genes that were both high-scoring by our scoring algorithm and deemed likely pathogenic by clinical assessment. Analysis of the remaining 17 families has identified candidate variants in a number of interesting genes, some of which have withstood further segregation testing in extended pedigrees. Family-based exome sequencing is a powerful strategy for the identification of novel disease genes, yet these studies often identify thousands of potential causative variants. Here, we demonstrate that comprehensive scoring of individual variants coupled with two genetic linkage approaches can substantially refine the search for disease-causing mutations.

99

Integrative annotation of variants from 1,092 humans: application to cancer genomics. *E. Khurana, M. Gerstein, Functional Interpretation Group of the 1000 Genomes Consortium.* Yale University, New Haven, CT.

Plummeting sequencing costs have led to a great increase in the number of personal genomes. Interpreting the large number of variants in them, particularly in non-coding regions, is a central challenge for genomics. We investigate patterns of selection in DNA elements from the ENCODE project using the full spectrum of sequence variants from 1,092 individuals in the 1000 Genomes Project Phase 1, including single-nucleotide variants (SNVs), short insertions and deletions (indels) and structural variants (SVs). We analyze both coding and non-coding regions, with the former corroborating the latter. We identify a specific sub-group of non-coding categories that exhibit very strong selection constraint, comparable to coding genes: “ultra-sensitive” regions. We also find variants that are disruptive due to mechanistic effects on transcription-factor binding (i.e. "motif-breakers"). Using connectivity information between elements from protein-protein interaction and regulatory networks, we find that variants in regions with higher network centrality tend to be deleterious. Indels and SVs follow a similar pattern as SNVs, with some notable exceptions (e.g. certain deletions and enhancers). Using these results, we develop a scheme and a practical tool to prioritize non-coding variants based on their potential deleterious impact. As a proof of principle, we experimentally validate and characterize a small number of candidate variants prioritized by the tool. Application of the tool to ~90 cancer genomes (breast, prostate and medulloblastoma) reveals ~100 candidate non-coding cancer drivers. This approach can be readily used in precision medicine to prioritize variants.

100

Efficiency of whole exome/genome sequencing for achieving a diagnosis in rare presentations. *M.C. Towne[1], A.H. Beggs[1, 2], P.B. Agrawal[1, 2].* 1) The Manton Center for Orphan Disease Research, Division of Genetics, Children's Hospital Boston, Boston, MA; 2) Harvard Medical School, Boston, MA.

To measure the cost effectiveness and efficiency of whole exome/genome sequencing (WES/WGS) compared to targeted gene analysis in a pediatric population with an undiagnosed potentially genetic disease. We analyzed the molecular diagnostic process of individuals enrolled in The Manton Center for Orphan Disease Research Gene Discovery Core, a repository for individuals with undiagnosed and rare conditions. DNA from 103 families was sent for WES or WGS analysis. Of the 99 datasets returned, we have completed analysis for 45 families. Ten of these patients received a genetic diagnosis after enrollment through either research WES/WGS or clinical targeted gene analysis. By tabulating the number of target genes/gene panels sequenced in each of these patients prior to obtaining a genetic diagnosis, associated time and costs, we can estimate the relative efficiencies of traditional genetic diagnostic and WES/WGS approaches. Here, we present preliminary data on 5 of these 10 families. Two patients were diagnosed by targeted sequencing while 3 were diagnosed by WES/WGS. We saw shorter turnaround times between initial presentation and diagnosis for the participants who had a diagnosis found by research WES/WGS (210 days) versus those diagnosed by targeted analysis (605.5 days). The average number of genetics clinic visits to obtain a diagnosis was 6, and the number of targeted molecular tests performed per patient prior to a genetic diagnosis was 9.5. We expect the cost for WES/WGS, even when sending trios for analysis (average of $3,700 USD for WES per family) to be less than the targeted gene analysis approach. Further analysis of the cost per test is underway. Of significance, in the 3 year period examined, the cost of exome sequencing has decreased by nearly 50% ($1,999 to $1,025). Analysis did not include costs associated with additional duration of stay, other tests, administrative/visit fees and parental stress due to delays in diagnosis. In one patient, WGS data did not confirm the mutation identified by targeted sequencing. Our preliminary data suggests that WES/WGS is an efficient and likely cost-effective way of diagnosing the genetic basis of disease compared to targeted sequencing. These findings are reflective of research testing, and WES/WGS would likely have a shorter turnaround time when handled in a more automated clinical laboratory. In our cohort, genetic diagnoses were identified in 22% of the 45 families, demonstrating limitations of WES/WGS.

101

Computational prediction and in vivo validation of suppressors of human disease mutations. *D.M. Jordan[1,2], E.E. Davis[3], N. Katsanis[3], S.R. Sunyaev[2].* 1) Biophysics Program, Harvard University, Cambridge, MA; 2) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 3) Center for Human Disease Modeling, Department of Cell Biology, Duke University Medical Center, Durham, NC.

Predicting the phenotypic effects of genetic variation is an important and widespread problem in modern genetics, with applications in gene and variant discovery studies, population genetics studies, and clinical genetic diagnostics. There are a number of computational tools that exist for this task, such as PolyPhen and SIFT. These methods rely on a comparative genomics approach, building a multiple sequence alignment and using it to assess the behavior of the variant over evolutionary time. This approach assumes that a variant observed in another species will only very rarely cause disease in humans. We use the training data for PolyPhen to evaluate how faulty this assumption is, by assessing how many variants known to cause disease in humans are found in the reference sequences of other species. We find that nearly 10% of variants annotated as pathogenic in humans appear in the reference sequence of at least one other vertebrate species, indicating that the assumption that human disease mutations are not found in the genomes of other species is significantly violated. This result holds even after filtering disease annotations to a very high level of confidence. The pattern of variation in genes where these variants occur is consistent with a model where a single effectively neutral change in the same gene can act as a compensation for the disease-causing variant, which we show through simulation. Based on this model, we use a combination of comparative genomics data and computational prediction of structural stability to generate candidate pairs of disease-causing variants and compensatory changes, which can then be directly tested by experiment. We report on this dataset of candidate variant pairs and the preliminary results of in vivo validation with a zebrafish morpholino rescue model. The ability to identify the interacting partners of specific variants allows us both to explore the shortcomings of the comparative genomics approach to variant assessment, and to uncover new biology relating to known human disease mutations.

102

The Empowered Whole Genome Cohort: Shareable Joint Genome Interpretation for Research and Personal Insight. *N.M. Pearson, J. Deschenes, C. Palm, D. Richards, D. Bassett.* Ingenuity Systems, Redwood City, CA.

A single human genome, studied alone, rarely sheds clear light on its carrier's health. But comparing many genomes, in the context of phenotypic variation, promises to ultimately refine diagnostic and prognostic care worldwide. Thus healthy people's genomes, especially, are for now most informative when pooled for mutual and broader benefit. Well-sequenced people may thus wish to share their genomes, in order to make them personally and societally useful -- yet few such people can even access their own data, or that of others, with (or even without) the computational resources to make sense of it. To help remedy this, we invited whole-genome sequencees from Illumina's Understand Your Genome program to establish free accounts on a HIPAA- and Safe Harbor-compliant platform for interpreting comprehensively sequenced human genomes (Ingenuity Variant Analysis), where each sequencee can upload, annotate, and query her/his own genome about any phenotype, via a well-curated biological knowledge base that augments Illumina's 344-gene clinically focused annotation. More importantly, the Variant Analysis platform lets sharing-minded sequencees pool and sensibly compare their genomes, using functionally-informed small- or large-sample-appropriate tests of rare or/and common variant association, to better understand phenotypic differences among cohort members (as well as statistical methods and challenges of such study). Together, sequencees can jointly publish relevant analyses for open review, and -- most usefully -- can offer their genomes (with ancillary data, as willing) to researchers who need well-phenotyped controls (or cases) for disease studies. Here we report early feedback from, and findings by, this pioneering participatory whole genome cohort. Relevant insights will refine further collaborative efforts among citizen-scientist data donors and professional researchers -- and may, more strikingly, spark direct biomedical discoveries in their own right.

103

Understanding molecular mechanisms of human disease mutations and coding variants through 3D protein networks. *H. Yu[1,2], J. Das[1,2], Y. Guo[1,2,3], X. Wei[2,4], X. Wang[2,3], B. Thijssen[5], A. Grimson[3], S.M. Lipkin[4], A.G. Clark[1,3].* 1) Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 2) Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca, NY; 3) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY; 4) Department of Medicine, Weill Cornell College of Medicine, New York, NY; 5) Department of Bioinformatics, Maastricht University, 6200 MD Maastricht, The Netherlands.

To better understand the molecular mechanisms and genetic basis of human disease, we systematically examine relationships between 3,949 genes, 62,663 mutations, and 3,453 associated disorders by generating a 3D structurally resolved human interactome. This network consists of 4,222 high-quality binary protein-protein interactions with their atomic-resolution interfaces. We find that in-frame mutations (missense mutations and in-frame insertions/deletions) are enriched on the interaction interfaces of proteins associated with the corresponding disorders, and that the disease specificity for different mutations of the same gene can be explained by their location within an interface. We also predict 292 candidate genes for 694 unknown disease-gene associations when a known disease protein interacts with our newly predicted candidate at the interface where corresponding disease-specific mutations are highly enriched. By considering the dominance/recessiveness of the disease mutations, we further find that although recessive mutations on the interaction interface of two interacting proteins tend to cause the same disease, this widely accepted "guilt-by-association" principle does not apply to dominant mutations. Furthermore, recessive truncating mutations (nonsense mutations and frameshift insertions/deletions) on the same interface are much more likely to cause the same disease, even if they are close to the N-terminus of the protein; whereas dominant truncating mutations tend to be enriched between interfaces. These results suggest that a significant fraction of truncating mutations can generate functional protein products, contrary to the common belief that truncating mutations most often cause complete loss of function. Finally, we find that rare non-synonymous coding variants are significantly enriched at the interaction interface, compared to common ones, indicating that our approach could be particularly effective in assessing the functional relevance of thousands of coding variants on a genomic scale.

104

Evaluation of Power of the Illumina HumanOmni5M-4v1 BeadChip to Detect Risk Variants for Human Complex Diseases. *C. Xing[1], J. Huang[2], Y-H. Hsu[3,4,5], A.L. DeStefano[1,6,7], N.L. Heard-Costa[6,7], P.A. Wolf[6,7], S. Seshadri[6,7], D.P. Kiel[3,4], L.A. Cupples[1,7], J. Dupuis[1,7].* 1) Biostatistics, Boston University, Boston, MA, USA; 2) Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 3) Hebrew SeniorLife, Institute for Aging Research, Boston, MA, USA; 4) Harvard Medical School, Boston, MA, USA; 5) Molecular and Integrative Physiological Sciences Program, Harvard School of Public Health, Boston, MA, USA; 6) Neurology, Boston University School of Medicine, Boston, MA, USA; 7) Framingham Heart Study, National Heart, Lung, and Blood Institute, Framingham, MA, USA.

Although genome-wide association studies (GWAS) have successfully identified thousands of disease-risk loci that harbor common variants associated with complex disease traits, a large portion of heritability is not explained. Emerging sequencing technologies can characterize all the variants. However, the cost is still high. Illumina recently released the HumanOmni5M-4v1 (Omni5) genotype array with ~ 4.3 million assayed SNPs, a much denser array compared to other arrays. Omni5 balances both cost and array density. In this article, we investigated the power of Omni5 to detect genetic associations. The Omni5 included variants in a wide range of minor allele frequencies (MAF) down to less than 1%. Theoretical power calculations indicated that Omni5 has increased power compared to other arrays with lower density when evaluating associations with some known loci, although there are some exceptions. We further evaluated the genetic associations between known loci and several traits in the Framingham Heart Study (FHS): femoral neck bone mineral density (FNBMD), lumbar spine bone mineral density (LSBMD), and hippocampal volume (HV). Finally, we searched genome-wide for novel associations using the Omni5 genotypes. We compared our associations with the ones obtained on the same participants from Affymetrix 500K + MIPS 50K arrays and two imputed datasets based on Affymetrix 500K + MIPS 50K arrays: (1) HapMap Phase II and (2) 1000 Genomes as reference panels. We observed increased evidence for genotype-phenotype associations with smaller p-values for known loci using the Omni5 genotypes. With limited sample sizes, we also identified novel variants with small p-values close to or at genome-wide significant levels. Our observations support the notion that dense genotyping using the Omni5 can be powerful in detecting novel variants. Comparison with imputed data with higher density also suggests that imputation helps but can not replace genotyping especially when imputation ratio is low.

105

Integrated analysis of protein-coding variation in over 50,000 individuals. *M. Lek[1,2], D.G. MacArthur[1,2], A. Levy Moonshine[2], M. Rivas[3], S. Purcell[1,4], P. Sullivan[5], S. Kathiresan[1,2], M.I. McCarthy[3], M. Boehnke[6], S. Gabriel[2], D.M. Altshuler[2], G. Getz[1,2], M.J. Daly[1,2], M.A. DePristo[2], Exome Aggregation Consortium.* 1) Massachusetts General Hospital, Boston, MA; 2) Broad Institute of Harvard and MIT, Cambridge, MA; 3) University of Oxford, Oxford, UK; 4) Mt Sinai School of Medicine, New York, NY; 5) University of North Carolina, Chapel Hill, NC; 6) University of Michigan, Ann Arbor, MI.

The increasing availability of DNA sequencing data has empowered variant discovery in studies of both common and rare diseases. However, for these data to provide maximum utility it will be critical to generate consistent variant calls across tens of thousands of samples.

We have assembled and jointly analyzed exome sequencing data from a collection of over 55,000 individuals sequenced as part of a variety of population genetic and disease-specific studies, an approach enabled by the development of new compressed file formats and variant-calling algorithms. We demonstrate that joint calling substantially improves the accuracy, sensitivity and consistency of variant detection. In particular we highlight the benefits of the creation of very large joint-called sets of cases and controls for detecting rare causal variants in both complex and Mendelian diseases.

Our results provide a view of the spectrum of human functional genetic variation extending down to extremely low population frequencies. We describe the frequency and genomic distribution of human protein-coding genetic variation, and show that the frequency spectrum of rare variants can be used to assess the accuracy of functional annotation approaches and to identify genes more likely to harbor severe disease-causing mutations. We also report the distribution of predicted loss-of-function (LoF) variants across human genes, their validation with independent RNA sequencing data, and their application in candidate gene prioritization for severe disease.

Finally, we present new genotyping arrays containing the majority of protein-coding LoF variants and reported disease-causing mutations at an appreciable frequency in our cohort, empowering cost-effective association studies of rare, likely functional genetic variation and direct estimates of the penetrance of reported disease mutations in large, phenotyped cohorts.

106

The effect of CAD/MI SNPs on other vascular domains and the relation with recurrent vascular events. *V. Tragante[1,2], P.A.F.M. Doevendans[1], H.M. Nathoe[1], Y. van der Graaf[3], W. Spiering[4], A. Algra[3,5], G.J. de Borst[6], P.I.W. de Bakker[2,3,7,8], F.W. Asselbergs[1,3,9].* 1) Division Heart & Lungs, University Medical Center Utrecht, Utrecht, Netherlands; 2) Division Biomedical Genetics, University Medical Center Utrecht, Utrecht, Netherlands; 3) Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, The Netherlands; 4) Department of Vascular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands; 5) Department of Neurology, Rudolph Magnus Institute for Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands; 6) Vascular Surgery, UMC University Medical Center Utrecht, Utrecht, The Netherlands; 7) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachussetts; 8) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts; 9) Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, The Netherlands.

Genome-wide association studies (GWAS) have identified many genetic loci related to coronary artery disease (CAD) and myocardial infarction (MI). However, the extent to which these loci are related to other vascular diseases is not clear. We calculated a multi-locus genetic risk score (GRS) in 8446 participants of the SMART (Second Manifestations of ARTerial disease) study based on the lead SNPs at 30 CAD/MI loci, and tested this GRS for cross-sectional association to CAD/MI, ischemic stroke (IS), abdominal aortic aneurysm (AAA) and peripheral artery disease (PAD), adjusting for age and sex. We also investigated whether this GRS was associated with recurrent vascular events using Cox regression, adjusting for age, sex, BMI, type 2 diabetes, LDL cholesterol, smoking and hypertension. We found that the GRS was significantly associated with CAD ($p=1.31 \times 10\text{-}9$), IS ($p=0.030$) and PAD ($p=6.93 \times 10\text{-}04$), but not with AAA ($p=0.057$). The lead SNP at the 9p21 locus (rs4977574) was associated with all 4 vascular diseases ($p < 4 \times 10\text{-}3$), illustrating the functional pleiotropy of this locus. We observed a nominally significant association for rs964184 (gene complex APOA5-A4-C3-A1 on chromosome 11) where the T allele confers increased risk for IS (OR = 1.32 (95% CI 1.14 - 1.52, $p = 2.33 \times 10\text{-}04$)). The GRS was associated with recurrent risk of MI ($p=0.026$), with a HR of 1.13 (95% CI 1.00 - 1.28) for individuals in the top quartile of the GRS distribution (N=30 recurrent events) compared to those in the bottom quartile (N=8 recurrent events). Finally, we found a significant positive relationship between the GRS and the number of vascular events ($p=3.26 \times 10\text{-}05$). These findings suggest that CAD/MI associated risk alleles play an etiological role in different types of atherosclerotic disease.

107

Multi-trait meta-analysis of genome-wide association studies (GWAS) of lipid levels and BMI reveals pleiotropy. *V. Lagou[1,2], R. Mägi[3], I. Surakka[4,5], A.-P. Sarin[4,5], M. Horikoshi[1,2], L. Marullo[6], T. Ferreira[1], G. Thorleifsson[7], S. Hägg[8,9], M. Beekman[10,11], C. Ladenvall[12], A. Mahajan[1], J.-J. Hottenga[13], J. S. Ried[14], T. W. Winkler[15], C. Willenborg[16], M. I. McCarthy[1,2], A. P. Morris[1], S. Ripatti[4,5,17], I. Prokopenko[1,2,18], ENGAGE (European Network for Genetic and Genomic Epidemiology) consortium.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom; 3) Estonian Genome Center, University of Tartu, Tartu, Estonia; 4) Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland; 5) National Institute for Health and Welfare, Helsinki, Finland; 6) Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy; 7) deCODE Genetics, 101 Reykjavik, Iceland; 8) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 9) Department of Medical Sciences, Molecular Epidemiology, Uppsala University Hospital, Uppsala, Sweden; 10) Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; 11) Netherlands Consortium for Healthy Ageing, Leiden, The Netherlands; 12) Department of Clinical Sciences, Diabetes and Endocrinology, Lund University and Lund University Diabetes Centre, CRC at Skåne University Hospital, Malmö, Sweden; 13) Netherlands Twin Register, Dept Biological Psychology, VU Univ Amsterdam, The Netherlands; 14) Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 15) Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany; 16) AG Kardiovaskuläre Genomik, Medizinische Klinik II, Universität zu Lübeck, Lübeck, Germany; 17) Wellcome Trust Sanger Institute, United Kingdom; 18) Department of Genomics of Common Disease, School of Public Health, Imperial College London, Hammersmith Hospital, London, United Kingdom.

Serum lipid levels, fat storage and obesity share biochemical pathways and can be influenced by common genetic factors. Analysis of the genetic effects on multiple traits simultaneously allows dissection of variable patterns of multi-trait associations. Within the ENGAGE consortium, we assessed multi-trait genetic effects on four blood lipids (high-/low-density lipoprotein and total cholesterol, triglycerides [HDL/LDL/TC/TG]) and body-mass index (BMI). The 1000 Genomes reference panel (06/2011) was used for imputation in up to 50,539 individuals from 19 European GWAS. Each study carried out multi-trait analysis by fitting a multiple logistic regression on SNP genotypes allowing for joint effects of four lipid traits and BMI and combining evidence across study-specific likelihoods. Individual trait effects from each study multi-trait model were estimated by fixed effects inverse-variance meta-analyses. Single-trait meta-analyses, conditional on remaining traits, were used to verify the independence of trait-specific genetic effects. Joint analysis enabled identification of 26 signals with genome-wide significant (P_{LRT}<5.0×10^{-8}) multi-trait effects, including 11 loci with associations driven by the individual trait effects: a) *TRIB1* on BMI, b) *GCKR*, *FADS1*, *PLTP* on TG, c) *CELSR2* on HDL, d) *CEPT* on BMI/HDL, e) *MLXIPL*, *LPL*, *APOA1* on BMI/TG, f) *LIPC* on HDL/TG and g) *APOE* on HDL/LDL/TG. At three loci previously associated with adiponectin (*TRIB1*), liver enzymes (*TRIB1*, *MLXIPL*) and lipodystrophy (*LPL*), genome-wide significant effects on obesity were observed for the first time, with higher BMI being related to lower TG indicating complex relationships between obesity and regulation of lipid levels. At the remaining 15 pleiotropic loci, multiple traits contributed to the signal. Pleiotropy was also observed at two loci [*IRS1* (P_{LRT}=5.0×10^{-4}) and *SH2B1* (P_{LRT}=5.0×10^{-3})] previously associated with adiposity in GWAS with higher BMI being related to higher HDL at *IRS1* and higher BMI with lower TG at *SH2B1*. We detected a substantial proportion of metabolic trait loci with complex patterns of genetic effects, some of which may not follow epidemiological correlations. This study highlights the emerging need of systematic investigation of multi-trait effects across the genome.

108

Genome-Wide Association with Fasting Glucose and Insulin in 20,200 African Americans Suggests New Quantitative Trait Loci and Allelic Heterogeneity at Known Loci: the African American Glucose and Insulin Genetic Epidemiology (AAGILE) Consortium. *J. Meigs[1], M-F. Hivert[2], A. Morris[3], M. Li[4], M. Ng[5], J. Liu[6], R. Jensen[7], X. Guo[8], L. Yanek[9], M. Nalls[9], L. Bielak[10], M. Irvin[11], W-M. Chen[12], P. An[13], E. Kabagambe[14], B. Cade[15], J. Wilson[16], the. MAGIC Investigators[17], J. Hong[18], D. Rybin[18], C-T. Liu[18], the. AAGILE Investigators[19].* 1) Gen Med Div, Massachusetts Gen Hosp, Boston, MA; 2) University of Sherbrooke, Sherbrooke, Quebec, Ca; 3) University of Oxford, Oxford, UK; 4) Johns Hopkins University, Baltimore, MD; 5) Wake Forest University, Winston-Salem, NC; 6) Womens Health Initiative, Seattle, WA; 7) University of Washington, Seattle, WA; 8) Cedars Sinai Medical Center, Los Angeles, CA; 9) NIH, Bethesda, MD; 10) University of Michigan, Ann Arbor, MI; 11) University of Alabama, Birmingham, AL; 12) University of Virginia, Charlottesville, VA; 13) Washington University, St. Louis, MO; 14) Vanderbilt University Medical Center, Nashville, TN; 15) Brigham and Women's Hospital, Boston, MA; 16) University of Mississippi Medical Center, Jackson, MS; 17) USA and Europe; 18) Boston University School of Public Health, Boston, MA; 19) USA.

High fasting glucose (FG) and insulin (FI) disproportionately affecting African Americans (AA) may have a genetic basis. We used meta-analyses (m-a) of genome-wide (g-w) association studies (GWAS) of FG and FI in AA to test whether loci identified in European (EU) ancestry individuals also are associated in AA, and to find new AA quantitative trait (QT) loci. We performed GWAS in 16 cohorts of 20,209 (FG) and 17,871 (FI, adjusted for BMI) non-diabetic AA (mean age 56 yr) using additive genetic models to test associations with 3.3M single nucleotide polymorphisms (SNPs), and combined results in METAL using inverse-variance weighted m-a. To leverage possible AA-EU heterogeneity at each SNP, we combined AA METAL results with MAGIC published results (Manning 2012, N= up to 96,496 EU in 29 cohorts) and m-a the two results files using MANTRA, a Bayesian method accounting for allelic heterogeneity among population clusters that returns a Bayes Factor, with (logBF) >6 suggesting g-w SNP-trait association. We evaluated associations for 23 known FG and 8 FI loci by testing reported EU Index SNPs (Dupuis 2010, Manning 2012) in AA and also identifying the "Best" SNP within +/- 250 kb of the EU Index SNP. We sought new loci for replication based on low METAL P values in AA and high MANTRA logBF in AA+EU. For 23 known FG loci, 1 EU Index (MTNR1B) and 1 AA Best SNP (GCK) were g-w significant (P<2.5×10-8) in AA, 9 Index and 23 Best SNPs were nominally significant (P<0.05), with 13 Best SNPs remaining significant after P value correction for the number of SNPs tested at the locus. At 14/23 FG loci the r2 (YRI HapMap 2) for Index vs Best SNP was <0.2, suggesting potentially independent signals. For 8 known FI loci, no Index or Best SNPs were g-w significant in AA, but 3 Index and 8 Best SNPs had P<0.05, with 5 Best SNPs remaining significant after P value correction. At 4/8 FI loci the r2 Index vs. Best SNP was <0.2. For new discovery, 14 novel AA FG and 8 novel AA FI loci had a SNP P<10-5 (in AA) and logBF >4 (in AA+EU), giving 22 high-interest SNPs now being tested for replication. We conclude that all 31 FG and FI loci known in EU show at least nominal association in AA, suggesting some genetic determinants of FG and FI are similar across AA and EU, but with apparent allelic heterogeneity at many loci. Combining fixed effects m-a in an AA sample with trans-ethnic m-a in an AA-EU sample has identified a wealth of new AA diabetes-related QT SNPs to test for replication.

109

Harnessing Web 2.0 Social Networks for Genetic Epidemiology Studies with Millions of People. *Y. Erlich[1], J. Kaplanis[1], M. Gershovits[1], P. Nagaraj[1,2], D. MacArthur[3,4], A. Price[5].* 1) Whitehead Inst Biomedical Research, Cambridge, MA; 2) Massachusetts institute of technology; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 4) Broad Institute of Harvard and MIT, Cambridge, MA; 5) Harvard School of Public Health, Boston, MA.

Understanding the genetic architecture of complex traits is one of the top missions of human genetics. Emerging lines of studies have highlighted the entangled etiologies of these traits, which can include epistasis, parent-of-origin effects, sex and age interactions, and environmental risk factors. To conduct robust genetic epidemiological analysis, statistical models require sampling substantial amount of data from large families. However, the recruitment of large cohorts of extended kinships is both logistically challenging and cost-prohibitive. Here, we present a Big Data strategy to address this challenge: harnessing existing, free, and massive Web 2.0 social network resources to trace the aggregation of complex traits in extremely large families. We collected millions of public profiles from Geni.com, the world's largest genealogy-driven social network. Using this information, we constructed a single pedigree of 13 million individuals spanning many generations up to the 15th century and validated its quality using unilineal Y chromosome and mitochondira markers. In addition, Natural Language Processing was used to convert genealogical information into birth and death locations to obtain a proxy for environmental factors. We obtained multiple of phenotypes from this resource including longevity, fertility, migration patterns, and facial morphologies phenotyped form digital photos in Geni.com. This dataset provides a wide range of kinships for familial aggregation studies. We will present the dataset, which we aim to release as a community resource and show heritability estimates across distant relatives to disentangle analysis of epistasis, parent-of-origin, and shared environments.

110

Integrated model of multiple types of rare variants and prior information improves the power of detecting risk genes for autism. *X. He[1], S.J. Sanders[2], L. Liu[3], S.D. De Rubeis[4,5], E.T. Lim[6,7], J.S. Sutcliffe[8], G.D. Schellenberg[9], R.A. Gibbs[10], M.J. Daly[6,7], J.B. Buxbaum[4,5,11,12], M.W. State[2], B. Devlin[13], K. Roeder[1,3].* 1) Lane Center of Computational Biology, Carnegie Mellon University, Pittsburgh, PA; 2) Departments of Psychiatry and Genetics, Yale University School of Medicine, New Haven, CT 06520, USA; 3) Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA; 4) Seaver Autism Center for Research and Treatment, Mount Sinai School of Medicine, New York, New York 10029, USA; 5) Department of Psychiatry, Mount Sinai School of Medicine, New York, New York 10029, USA; 6) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA; 7) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA; 8) Vanderbilt Brain Institute, Departments of Molecular Physiology & Biophysics and Psychiatry, Vanderbilt University, Nashville, Tennessee, 37232, USA; 9) Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104; 10) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA; 11) Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York 10029; 12) Friedman Brain Institute, Mount Sinai School of Medicine, New York, New York 10029, USA; 13) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15213, USA.

While whole exome sequencing (WES) greatly facilitates the study of rare genetic variation, it is widely believed that very large sample sizes are required to identify risk genes for complex disease from such data. In contrast, a surprising degree of progress has been made for early-onset disorders, like autism spectrum disorders (ASD), by identifying recurrent de novo mutations in moderate sized samples. In this work we propose statistical strategies to build on these promising results by using multiple types of data in a unified Bayesian framework. With our initial efforts, we develop methods that can incorporate WES data regarding de novo mutations, inherited rare variants, and rare variants identified from cases and controls. TADA, for Transmitted And De novo Association, integrates these data by a gene-based likelihood model involving parameters for mutation rates, allele frequencies and gene-specific penetrances. Inference is based on an Hierarchical Bayes strategy that borrows information across all genes to improve parameter estimation. We validated TADA using realistic simulations mimicking rare, large-effect mutations affecting risk for ASD and show it has much better power than other common methods of analysis. Thus TADA's integration of various kinds of WES data can be an effective means of identifying novel risk genes. Indeed by applying TADA to all published WES data from subjects with ASD and their families, as well as from a case-control study of ASD, we identified several novel and promising ASD candidate genes with strong statistical support. Moreover, based on published comparisons of the rate of de novo mutations in ASD probands versus their siblings, it has been conjectured that half of the 116 genes that sustained exactly one severe de novo mutation in probands are ASD risk genes. TADA successfully identifies approximately half of these genes as promising candidates deserving further investigation. We are pursuing several refinements of the TADA framework, in particular we investigate how to use the prior information of rare variants to improve TADA. We are studying three types of external information: effects of mutations on protein function; the allele frequencies of variants in large independent samples; and the selective constraints on variants in the population. Our simulations and analysis on a collection of known ASD risk genes demonstrate that such information boosts the power of association studies.

111

Tests of aggregate rare variant association applied to a multiethnic sequencing study. *A.D. Ablorh[1], S. Lindstrom[1], C.A. Haiman[2], B.E. Henderson[2], L. Le Marchand[3], S. Lee[4], D.O. Stram[2], P. Kraft[1,4].* 1) Epidemiology, Harvard School of Public Health, Boston, MA; 2) Department of Preventive Medicine, Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, California; 3) Epidemiology Program, University of Hawaii Cancer Research Center, Honolulu, HI; 4) Biostatistics, Harvard School of Public Health, Boston, MA.

For many complex diseases, the genetic basis of susceptibility has not been fully explained and it is possible that rare variants with a minor allele frequency (MAF)≤0.005, contribute to the remaining genetic heritability. Several association methods have been proposed for aggregating information from individual rare variants in order to maximize power. However, these have been proposed in the context of studies conducted in a single ethnicity. It remains an open question how to best combine rare variants in association testing across ethnicities in a multi-ethnic study. Since causal variants unique to one population may be harder to detect but causal variants that are common to all may reveal themselves, performance of rare variant association methods may differ in multi-ethnic compared to single-ethnic study populations. We selected four statistical approaches for comparison. Three involve first applying the Combined Multivariate Collapsing Method (CMC) within each population separately. The first tests whether the common collapsed effect is non-zero using fixed effects meta-analysis; the second tests whether any of the unconstrained collapsed effects is non-zero; the third, a modification of MANTRA, constrains the collapsed effects according to population genetic similarity. The last approach is an extension of the Sequence Kernel Algorithm Test (SKAT) that combines results across populations, MetaSKAT. We applied all four methods to a targeted sequencing study where we sequenced twelve breast cancer susceptibility loci in 1560 women (771 cases and 789 controls) from the Nurses' Health Study (NHS) and 2107 women (1542 cases and 1565 controls) from the Multiethnic Cohort (MEC). Sequencing allows us to characterize variants across a spectrum of allele frequencies much lower than those previously explored. Our study population includes women from four ethnic groups (33% European-, 20% African-, 19% Japanese-, and 27% Latin-American). 87.3% of sequenced variants were rare (MAF≤0.005), and 70.2% were population private. We present the results of meta-analysis of rare genetic variation using the four statistical approaches in a multi-ethnic population, and discuss the implications for design and analysis of future studies.

112

Whole-genome sequence based association studies of complex traits: the UK10K project. *N. Timpson, UK10K Consortium: Cohorts.* MRC/UoB IEU/SSCM, Bristol Univ, Bristol, United Kingdom.

The UK10K project (www.uk10k.org) is a collaborative effort between multiple research centres mainly in the UK which has four core aims: uncover genetic variants of a broad frequency range, generate comprehensive reference panels for imputation of genetic variation out into larger collections, perform association analyses in an extensive list of phenotypes and lastly to provide a sequence and phenotype based resource for the research community. As part of Cohorts component of the UK10K project, 4,000 individuals from two deeply phenotyped cohorts - TwinsUK and the Avon Longitudinal Study of Parents and Children (ALSPAC) - have been sequenced to 6x coverage using next-generation sequencing technology. The data generated has been used to explore phenotypic associations in over 50 cardiometabolic and health related traits using both single point common variant and burden based rare variant tests. These have indicated novel trait associations in a series of contexts; new variants in known gene regions, new associations in novel gene regions, association in previously poorly imputed regions and rarer variation based associations. Initial results indicate that a number of these appear to replicate in the full ALSPAC data set and further analyses are allowing us to address a range of related questions related to the performance and utility of whole-genome sequencing. Examples of these questions are: (i) To what extent rare variants contribute to phenotypic variance in complex traits; (ii) To what extent does the presence of structure within phenotype and genotype data complicate the interpretation of association signals at differing allele frequencies; (iii) To what extent is there an over-representation of functionally annotated genomic regions within loci harboring association signal. A panel of approximately 50 cardiometabolic quantitative traits is analyzed in conjunction with whole-genome sequence data to address these questions, and initial association results will also be presented. In summary, the data provide insights about what large-scale whole-genome sequencing efforts are likely to reveal for the genetic architecture of complex traits. All data, both phenotypic and genetic, generated by the project are being made available to the scientific community under managed access.

113

Genetic variation associated with the susceptibility to herpes zoster in the eMERGE Network. *D. Crosslin[1,2], D. Carrell[3], E. Baldwin[3], M. de Andrade[4], I. Kullo[5], G. Tromp[6], H. Kuivaniemi[6], K. Doheny[7], E. Pugh[7], A. Kho[8], M. Hayes[9], M. Ritchie[10], S. Verma[10], G. Armstrong[10], A. Saip[11], J. Denny[11], D. Crawford[12,13], P. Crane[14], S. Mukherjee[14], E. Bottinger[15], T. Manolio[16], R. Li[16], A. Burt[1], D. Kim[1,2], B. Keating[17], D. Mirel[18], E. Larson[3], C. Carlson[19], G. Jarvik[1,2], The electronic Medical Records and Genomics (eMERGE) Network.* 1) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 2) Department of Genome Sciences, University of Washington, Seattle, WA; 3) Group Health Research Institute, Center for Health Studies, Seattle, WA; 4) Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN; 5) Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN; 6) Sigfried and Janet Weis Center for Research, Geisinger Health System, Danville, PA; 7) Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD; 8) Divisions of General Internal Medicine and Preventive Medicine, Northwestern University, Chicago, IL; 9) Division of Endocrinology, Metabolism, and Molecular Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL; 10) Center for Systems Genomics, Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA; 11) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 12) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 13) Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN; 14) Division of General Internal Medicine, University of Washington, Seattle, WA; 15) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine Mount Sinai, New York, NY; 16) Office of Population Genomics, National Human Genome Research Institute, Bethesda, MD; 17) Center for Applied Genomics. Children's Hospital of Philadelphia, Philadelphia, PA; 18) Program in Medical & Population Genetics, Broad Institute of Harvard & MIT, Cambridge, MA; 19) Fred Hutchinson Cancer Research Center, Public Health Sciences Division, Seattle, WA.

Herpes zoster, commonly referred to as shingles, is caused by the varicella zoster virus (VZV). VZV initially manifests as chicken pox in youth, but can remain latent in nerve tissues for many years and reemerge as shingles for unknown reasons. Two-thirds of shingles cases are older than 60 years. Shingles presents as painful, usual unilateral vesicular skin infection leaving 10%-18% of those affected with chronic post-herpetic neuralgia. Annual U.S. costs of incident infections are $1.1 billion. Assessment of genetic variation has the potential to inform inter-individual variation in susceptibility to VZV reemergence. While a vaccine is now available, it is only 50% effective and is not widely used. Polymorphisms of the immune system that predicts risk of VZV re-activation could suggest the basis for latency and aid clinical decision-making as well as informing insight into the immune biological process of infection and clearance. We performed a joint and site-stratified genome-wide association analyses to identify variants associated with infection susceptibility to VZV in subjects from the electronic Medical Records and Genomics (eMERGE) Network. The eMERGE Network comprises a multi-ethnic cohort of roughly 50,000 individuals linked to electronic medical records for phenotype mining from nine participating sites in the US. Early joint association results using logistic regression from four eMERGE sites (1,684 cases, 12,088 controls) suggest two genomic regions of interest reaching genome-wide significance (P-value = 1.0×10^{-8}). One region on chromosome 14 represents variants found in the gene encoding signal recognition particle 54kDa (*SRP54*). This signal recognition particle binds to the signal sequence of presecretory proteins as they emerge from the ribosome. On chromosome 18, there are variants found in the region representing the gene zinc finger and BTB domain containing 7C (*ZBTB7C*). The relationship of its function as a tumor suppressor gene with VZV infection susceptibility is not clear, but is currently being assessed. More data are pending from the remaining eMERGE sites, which will enhance power. We will also assess genomic association using survival analysis with endpoints of loss to follow-up, date of death, or first zoster reactivation for unvaccinated subjects. We aim to robustly identify variants associated with infection susceptibility to VZV.

114

Whole-Genome Detection of Disease-Associated Deletions or Excess Homozygosity in a Case-Control Study of Rheumatoid Arthritis. *C.C. Wu[1], S. Shete[2], E.J. Jo[3], Y. Xu[4], E.Y. Lu[5], W.V. Chen[5], C.I. Amos[5, 6].* 1) Environmental and Occupational Health, College of Medicine, National Cheng Kung University, Tainan, Taiwan; 2) Biostatistics, MD Anderson Cancer Center, Houston, Texas; 3) Duncan Cancer Center, Baylor College of Medicine, Houston, Texas; 4) Biostatistics, School of Public Health, Yale University, New Haven, Connecticut; 5) Genetics, MD Anderson Cancer Center, Houston, Texas; 6) Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Hanover, New Hampshire.

Copy number variations are abundant in humans and represent one of the least well studied classes of genetic variants. Few comprehensive studies have been performed of copy number variation's contribution to complex human disease susceptibility. Because known rheumatoid arthritis susceptibility loci explain only a portion of familial clustering, we performed a genome-wide study of association between deletions and rheumatoid arthritis using high-density 550K SNP genotype data. We recently developed a statistical method for detecting deletions or excess homozygosity associated with complex disease in case-control studies, using SNPs in genome-wide association studies. We used this method and tested each contiguous SNP locus between the 868 cases and 1194 controls to detect deletion variants or excess homozygosity that influence susceptibility. Our method is designed to detect statistically significant evidence of deletions or homozygosity at individual SNPs for SNP-by-SNP analyses and to combine the information among neighboring SNPs for cluster analyses. In addition to successfully detecting known deleterious deletion variants on HLA-DRB1 and C4 genes on MHC, we identified additional 4.3-kb and 28-kb clusters on chromosomes 10p (5,316,846-5,321,159) and 13q (20,783,404-20,811,429), respectively, which were significant at a corrected 0.05 nominal significance level, adjusted for multiple comparison procedures. Independently, we performed analyses using the PennCNV method and identified cases and controls that had chromosomal segments with copy number < 2. PennCNV is an algorithm for identifying and cataloging copy numbers for individuals on the basis of a hidden Markov model. Using Fisher's exact test to compare the numbers of cases and controls per SNP, we identified 26 neighboring significant SNPs (protective; more controls than cases) that jointly showed evidence of deletions on chromosome 14 with p-values < 10-8. We further extended our method to logistic regression frame work, which allows us to adjust for the population structure using eigen vectors. This approach supported our previous findings.

115

Linkage analysis of hypertriglyceridemia in a single large family identifies 3 novel potentially pathogenic variants. *E.A. Rosenthal[1], J. Ranchalis[1], J.D. Brunzell[2], A.G. Motulsky[3], D.A. Nickerson[3,4], E.M. Wijsman[1,5], G.P. Jarvik[1,3].* 1) Dept Med Gen, Univ Washington School of Medicine, Seattle, WA; 2) Div. of Metabolism, Endocrinology, and Nutrition, Univ Washington School of Medicine, Seattle, WA; 3) Dept Genome Sci, Univ Washington, Seattle, WA; 4) Dept of Bioengineering, Univ Washington, Seattle, WA; 5) Dept of Biostatistics, Univ Washington, Seattle, WA.

Hypertriglyceridemia (HTG) is a risk factor for cardiovascular disease, a leading cause of death in the U. S. There are few pharmacological treatments, with non-compliance rates >5% due to side effects. Investigating the genetics of HTG may lead to new drug targets. There are ~35 known SNPs that explain only ~10% of variation in triglycerides (TG). Due to the genetic heterogeneity of HTG, a family study design is optimal for identification of novel loci with large effect size as the same mutation can be seen in many relatives and co-segregation with TG observed. We considered HTG in a large family of European American descent (N=121, 85 with TG, 82 with marker genotypes on ~5cM map), ascertained for familial combined hyperlipidemia, a diagnosis defined by variable atherogenic lipoprotein levels in multiple relatives. This family has 4 relatives with HTG > top ½% of U.S. adults, suggesting a strong genetic component. We log transformed TG after adjusting for sex and age using population data (log-adjTG). Known pathogenic SNPs do not explain log-adjTG in this family. Bayesian segregation analysis supports ≥1 quantitative trait locus (QTL) (posterior probability = 0.71). We analyzed the data with Bayesian Markov chain Monte Carlo joint oligogenic linkage and association using Loki2.4.7, as it does not need a fixed trait model, handles large pedigrees, and imputes missing genotypes. To evaluate linkage evidence, we calculated the Bayes factor (BF) for 2 cM intervals, which compares the posterior and prior odds that a QTL is located in a region. We detected linkage to chrs. 7, 14, 17 and 18 (maxBF=25, 44, 183 and 29, respectively). Whole exome data on 16 individuals revealed 3 highly conserved novel missense single nucleotide variants (SNVs) at *SLC25A40* on chr. 7, *PRKCH* on chr.14, and *PLD2* on chr. 17. Jointly, these SNVs explained 49% of the genetic variance (Vg) in log-adjTG; only the *SLC25A40* SNV was significantly associated with log-adjTG (p=0.0005), adjusting for pedigree structure, and explained 28% Vg. The SLC25A40 protein is involved in inner mitochondrial membrane transport. The SNV causes a *TYR125CYS* substitution just outside the 2nd helical transmembrane region. Cysteines stabilize folded proteins through disulfide bonds; an extra cysteine may allow for a protein configuration change that disrupts protein function. Whole gene testing in subjects from the Exome Sequencing Project confirmed the association between TG and *SLC25A40* (p=0.03).

116

Rare APOC3 loss-of-function variants lower plasma triglycerides and protect against clinical coronary heart disease. *J. Crosby[1,2], G.M. Peloso[3,4], P. Auer[5,6], D.R. Crosslin[7,8], G. Jarvik[7], L.A. Cupples[9,10], A. Reiner[11], E. Boerwinkle[1,2,12], S. Kathiresan[3,4,13] On Behalf of the Exome Sequencing Project HDL/TG Working Group.* 1) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA; 2) Department of Biostatistics, Bioinformatics, and Systems Biology, The University of Texas Graduate School of Biomedical Sciences at Houston, Houston, TX, USA; 3) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 4) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 5) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; 6) Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, USA; 7) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA, USA; 8) Department of Biostatistics, University of Washington, Seattle, WA, USA; 9) Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; 10) Framingham Heart Study, Framingham, MA, USA; 11) Department of Epidemiology, University of Washington, Seattle, WA, USA; 12) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; 13) Cardiology Division, Massachusetts General Hospital, Boston, MA, USA.

Plasma concentration of triglycerides is heritable and correlated with risk for coronary heart disease (CHD). Whole exome sequencing has the potential to discover rare mutations with large effects on a phenotype of interest. We sequenced the protein-coding regions of 18,666 genes in each of 3,734 participants of European and African-American ancestries from the U.S. National Heart, Lung, and Blood Institute's Exome Sequencing Project. Using linear regression and SCORE-Seq burden analyses, we tested whether rare mutations, individually or when aggregated within a gene, associated with plasma triglycerides. For mutations significantly associated with plasma triglycerides and replicated in an independent sample, we subsequently evaluated association with risk for CHD in 110,970 individuals. A burden of rare variants in the apolipoprotein C-III (APOC3) gene was associated with lower plasma triglyceride levels. This result could be attributed to four functional variants; three were annotated as loss-of-function (R19X, splice site mutation IVS2+1 G>A, and splice site mutation IVS3+1 G>T) and a fourth as missense (A43T). Carriers of R19X and splice site IVS2+1 G>A had 46% lower concentration of plasma apoC-III protein, consistent with loss of normal gene function. About 1 in 200 individuals were heterozygous carriers for at least one of the four variants and carriers had 38% lower plasma triglycerides ($P < 1 \times 10^{-20}$). APOC3 rare mutation carriers were at 40% reduced risk for CHD (OR 0.60, 95% CI 0.47 - 0.75, $P=4 \times 10^{-6}$ among 110,472 non-carriers and 498 carriers). In conclusion, rare APOC3 variants that disrupt gene function lead to lower circulating apoC-III protein level, reduced plasma concentration of triglycerides, and a decreased risk for clinical CHD. These observations based on human "knockout" alleles highlight pharmacologic inhibition of APOC3 function as a promising strategy to prevent CHD.

117

Novel rare and low frequency coding variants associated with LDL cholesterol levels. *C. Willer[1,2,3], L.A. Lange[4], Y. Hu[5], H. Zhang[1], C. Xue[2], E.M. Schmidt[2], M. Boehnke[5], L. Groop[6,7], M. McCarthy[8], T. Meitinger[9], M. Fornage[10,11], C. Ballantyne[12], E. Boerwinkle[11], D. Altshuler[13,14,15,16], D-y. Lin[17], G. Jarvik[18], L.A. Cupples[19,20], C. Kooperberg[21], J.G. Wilson[22], D.A. Nickerson[23], G.R. Abecasis[5], S.S. Rich[24], R.P. Tracy[25,26], NHLBI Exome Sequencing Project.* 1) Department of Internal Medicine, Univ Michigan, Ann Arbor, MI; 2) Department of Computational Medicine and Bioinformtics, U of Michigan, Ann Arbor, MI; 3) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 4) Department of Genetics, University of North Carolina, Chapel Hill, NC; 5) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 6) Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, University Hospital Skåne, Malmö, Sweden; 7) Glostrup Research Institute, Glostrup University Hospital, Glostrup, Denmark; 8) Oxford Centre for Diabetes, Endocrinology and Metabolism and Oxford NIHR Biomedical Research Centre, University of Oxford, Churchill Hospital, Oxford, UK; 9) Institute of Human Genetics, Hemholtz Center Munich, German Research Center for Environmental Health and Institute of Human Genetics, Technical University Munich, Neuherberg, Germany; 10) Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX, USA; 11) Center for Human Genetics, University of Texas, Health Science Center at Houston, Houston, TX, USA; 12) Baylor College of Medicine, and Methodist DeBakey Heart and Vascular Center, Houston, TX, USA; 13) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 14) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 15) Department of Molecular Biology, Massachusetts General Hospital, Boston, MA; 16) Department of Genetics, Harvard Medical School, Boston, MA; 17) Department of Biostatistics, University of North Carolina, Chapel Hill, NC; 18) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA, USA; 19) National Heart, Lung and Blood Institute's Framingham Heart Study, Framingham, MA; 20) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 21) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 22) Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS; 23) Department of Genome Sciences, University of Washington, Seattle, WA; 24) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 25) Department of Biochemistry, University of Vermont, Colchester, VT; 26) Department of Pathology, University of Vermont, Colchester, VT.

Elevated low-density lipoprotein cholesterol (LDL-C) is a treatable, heritable risk factor for cardiovascular disease. Genome-wide association studies have identified 157 variants associated with lipid levels but are not well suited to assess the impact of rare and low frequency variants. To determine whether rare or low frequency coding variants are associated with LDL-C, we exome sequenced 2,005 individuals, including 554 individuals selected for extreme LDL-C (> 98th or < 2nd percentile). Follow-up analyses included sequencing 1,302 additional individuals and genotype-based analysis of 52,222 individuals. We observed significant association between LDL-C and the burden of rare or low frequency variants for four genes: a new association at *PNPLA5*, a phospholipase domain containing gene, and novel as well as known variants in three known lipid genes: *PCSK9, LDLR* and *APOB*. We replicated the novel *PNPLA5* signal in an independent large-scale sequencing study of 2,084 individuals. The frequency, impact on protein structure, and effect sizes of associated rare variants differed among loci - from extremely rare loss-of-function variants in *APOB* to low frequency variants with more modest effect sizes in *PCSK9*. We present clear evidence that uncommon and rare variants contribute to variation of LDL-C levels in the general population. Our major specific findings are: 1) the identification of a novel LDL-C associated gene, *PNPLA5*; 2) the identification of known and novel variants in three previously identified genes *LDLR, PCSK9*, and *APOB*; and 3) the observation that associated variants have a range of minor allele frequency and putative functional importance which necessitates a variety of analytic approaches to optimize gene discovery.

118

Association Analysis of C-Reactive Protein Levels in European Americans and African Americans Sequenced Through the NHLBI Exome-Sequencing Project. *U.M. Schick[1,2], P.L. Auer[1,2,3], D.S. Kim[4], E.J. Benjamin[5,6,7,8,9], J.C. Bis[10], E. Boerwinkle[11], C.S. Carlson[1], J. Dupuis[5,12], M. Fornage[13], L. Hsu[1,14], R.D. Jackson[15], C. Kooperberg[1], L. Lange[16], H. Lin[17], S.M. Leal[18], A.C. Morrison[11], N. Pankratz[19], U. Peters[1], B. Psaty[10,14,20,21], S.S. Rich[22], R. Tracy[23], J.G. Wilson[24], M.D. Gross[2,19], A.P. Reiner[1,2,25] on behalf of the NHLBI GO Exome Sequencing Project.* 1) Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; 2) Equal Contribution; 3) Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, USA; 4) Department of Genome Sciences, University of Washington, Seattle WA, USA; 5) National Heart, Lung, and Blood Institute's and Boston University's Framingham Heart Study, Framingham, MA, USA; 6) Department of Medicine and Department of Epidemiology, Boston University School of Medicine, Boston, MA, USA; 7) Section of Cardiovascular Medicine and Section of Preventive Medicine, Boston University School of Medicine, Boston, MA, USA; 8) Cardiology Division, Massachusetts General Hospital, Boston, MA, USA; 9) Boston University School of Medicine's Whitaker Cardiovascular Institute, Evans Memorial Medicine Department, Boston, MA, USA; 10) Cardiovascular Health Research Unit and Department of Medicine, University of Washington, Seattle, WA, USA; 11) Human Genetics Center, University of Texas Health Science Center, School of Public Health, Houston, TX, USA; 12) Department of Biostatistics, Boston University School of Medicine, Boston, MA, USA,; 13) Brown Foundation Institute of Molecular Medicine and Human Genetics Center, Division of Epidemiology, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA; 14) Department of Biostatistics, University of Washington, Seattle, WA , USA; 15) Division of Endocrinology, Diabetes, and Metabolism, Ohio State University, Columbus, OH, USA; 16) Department of Genetics, University of North Carolina School of Medicine, Chapel Hill, NC, USA; 17) Department of Medicine, Boston University School of Medicine, Boston, MA, USA; 18) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 19) Department of Lab Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA; 20) Department of Epidemiology and Department of Health Services, University of Washington, Seattle, WA, USA; 21) Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA; 22) Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA; 23) Departments of Biochemistry and Pathology, University of Vermont, Burlington, VT, USA; 24) Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA; 25) Department of Epidemiology, University of Washington, Seattle, WA, USA.

Introduction: C-reactive protein (CRP) is a systemic marker of inflammation with predictive value in assessing cardiovascular risk. Genome-wide association studies (GWAS) have identified ~25 loci associated with CRP levels, however taken together these common variants explain only a small proportion of the phenotypic variance of CRP levels. **Methods:** We performed exome sequencing in 3,379 unrelated Americans of European (EA) and African (AA) ancestry through the NHLBI Exome Sequencing Project (ESP). The aims of this study were to 1) assess whether these 25 known CRP GWAS loci also contain rare coding variants associated with CRP levels, and 2) perform an exploratory exome-wide search for rare variants associated with CRP levels. Validation data sets included independent samples genotyped in ESP on the exome chip and exome sequence data from the Cohorts for Heart and Aging Research in Genetic Epidemiology consortium (CHARGE). **Results:** In single variant tests within CRP GWAS loci, novel rare coding variants in *CRP* (rs77832441 (T59M)) and *TOMM40* (rs112849259 (synonymous)) were associated with decreased CRP levels. In the combined ESP discovery sample and validation samples, the *CRP* T59M variant was associated with more than a 60% reduction in CRP levels (meta-analysis $P<1.65\times10^{-15}$). At the gene level, rare coding variation in candidate genes *PPP1R3B* ($P=2.61\times10^{-4}$) and *RORA* ($P=1.31\times10^{-4}$) was associated with CRP in the ESP AA discovery sample with suggestive evidence of replication of *RORA* in ESP exome chip samples (P=0.062). Gene burden tests of *PIGW* reached exome-wide significance in race-combined ESP gene-based tests ($P=2.10\times10^{-6}$), but failed to replicate in independent CHARGE sequencing and ESP exome chip samples. **Conclusion:** Overall, our results suggest that variants distributed across the allele-frequency spectrum within candidate genes identified by GWAS contribute to variation in CRP levels.

119

Genetic association studies illuminate the role of low frequency and rare variation in explaining the variation of blood pressure traits. *A. Manning[1], X. Sim[2], H.M. Highland[3], M.A. Rivas[4], H.K. Im[5], A. Mahajan[4], A.E. Locke[2], N. Grarup[6], P. Fontanillas[1], A.P. Morris[4], T.M. Teslovich[2], J. Flannick[1], C. Fuchsberger[2], K. Gaulton[4], H.M. Kang[2], J.B. Meigs[7], C.M. Lindgren[1,4] for T2D-GENES and GoT2D.* 1) Medical and Population Genetics Program, Broad Institute, Cambridge, MA; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 3) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX; 4) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 5) Department of Health Studies, University of Chicago, Chicago, USA; 6) Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, DK; 7) Massachusetts General Hospital, Boston, Massachusetts, USA.

Introduction: A goal of the T2D-GENES and GoT2D consortia is to illuminate the role that low frequency (LF) and rare genetic variants play in diabetes-related metabolic traits. To that end we performed genetic association studies of systolic and diastolic blood pressure (SBP and DBP, respectively) in ethnically diverse T2D case-control samples sequenced and genotyped on a variety of platforms. **Methods:** Analyses were performed on (1) whole genome sequencing in European samples (N=2300), (2) whole exome sequencing in 10K individuals from 5 ethnicities (African American, East Asian, European, Hispanic and South Asian), and (3) exome chip genotyping in European samples (N=39K). SBP and DBP values were adjusted and inverse normalized by cohort and separately in T2D cases and controls. The statistical analyses were single variant and gene-based tests accounting for relatedness and population structure. Three gene-based tests (SKAT, burden tests and a Bayesian analysis) were applied to sets of rare variants: non-synonymous, loss of function (LOF) and pathogenic mutations. Statistical significance was defined as P<0.05/number of tests performed. **Results:** For the single variant analysis, no LF or rare variants reached statistical significance, although common variants at several known loci were observed with genome-wide significance. Among the gene based results for SBP are genes (listed below) that are expressed in blood plasma, platelets and liver, making them plausible biological candidates. In the European sample with N=2300, two genes show suggestive results: *MFGE8* (SKAT P=0.0002), *LTK* (SKAT P=0.004). One gene shows suggestive results in the Bayesian analysis: *CD300LB* (log10BF=2.2 with empirical P=0.0003). In the 10K sample, no genes were statistically significant with SBP using SKAT or burden tests, but suggestive results were found with the gene *SNUPN* in the Bayesian LOF test (log10BF=3.1). In the exome chip analysis, a statistically significant association with SBP was observed with LF variants in *PIK3R3* ($P=1.5\times10^{-6}$, N=36668), which according to Gene Ontology is related to biological processes such as blood coagulation, platelet activity and the insulin receptor signaling pathway. **Conclusion:** Our results suggest that LF and rare coding variants contribute to variability in blood pressure and that larger sample sizes may be needed to detect genome-wide significant LF and rare coding variants that contribute to the genetic architecture of blood pressure.

120

Contribution of coding variation to type 2 diabetes-related quantitative traits in 13,000 exomes from multiple ancestries. *X. Sim[1], H.M. Highland[2], M.A. Rivas[3], H.K. Im[4], A.K. Manning[5], A. Mahajan[3], A.E. Locke[1], N. Grarup[6], P. Fontanillas[5], A.P. Morris[3], T.M. Teslovich[1], J. Flannick[5], C. Fuchsberger[1], K. Gaulton[3], H.M. Kang[1], J.B. Meigs[7], C.M. Lindgren[3], T2D-GENES and GoT2D Consortia.* 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 2) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 4) Department of Health Studies, University of Chicago, Chicago, USA; 5) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 6) Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, DK; 7) Massachusetts General Hospital, Boston, MA, USA.

Within the T2D-GENES and GoT2D consortia, we performed deep exome sequencing of 18,162 genes in ~13,000 type 2 diabetes case-control individuals from five major ancestry groups (African-American, East Asian, European, Hispanic, and South Asian). We are carrying out quantitative trait (QT) association analysis to assess the role of coding variation in type 2 diabetes-related lipid, anthropometric, and glycemic traits (the latter only analyzed in non-diabetic controls). Current analysis of exome sequence data from ~10,000 multi-ethnic individuals (2,000 from each ancestry group) identified ~2.5 million variants, 40% of which are non-synonymous. Within the non-synonymous variants, only 4% are polymorphic across all ancestry groups while 76% are ancestry-specific. Similarly, 5% of synonymous variants are polymorphic across all ancestry groups and 78% are ancestry-specific. Applying a combination of single variant and gene-level association tests, some interesting results from analyses to date identified (1) coding variants associated with diabetes-related traits specific to different ancestries, and (2) multiple non-synonymous variants exhibiting allelic heterogeneity across different ancestries. Gene-based association identified four loss-of-function variants significantly associated with LDL-cholesterol at *APOB* ($P = 4\times10^{-7}$), present in four individuals (1 African-American, 1 Hispanic, and 2 South Asians), who exhibited extremely low LDL-cholesterol levels (< 60mg/dl). We observed exome-wide significant gene-based association signals for body mass index (BMI) at *EIF3G* ($P = 6.9\times10^{-7}$), specific to African-Americans. The gene-based association was driven primarily by a single missense variant ($P = 1.2\times10^{-7}$) observed in five individuals, all of whom had BMI < 25kg/m^2. Finally, we detected evidence of association with decreased glycated hemoglobin (HbA1c) at the *G6PD* locus in African American (minor allele frequency [MAF] = 0.12), East Asian (MAF = 0.02), and South Asian (MAF = 0.02) ancestries, at different index missense variants, suggesting allelic heterogeneity at this locus. This diverse catalog of coding variation across wide allelic spectrum will further facilitate characterization of coding variants, and larger sample sizes may be needed to elucidate the role of low frequency and rare exonic variant associations in diabetes-related quantitative traits.

121

A Meta-Analysis of Genome-Wide Association Studies Identifies a Novel Locus Associated with Thrombin Generation Potential. *A. Rocañín-Arjó[1], L. Carcaillon[2], W. Cohen[3], N. Saut[3], M. Germain[1], L. Letenneur[4], M. Alhenc-Gelas[5], A.M. Dupuy[6], M. Bertrand[7], P. Amouyel[8], P.Y. Scarabin[2], D.A. Trégouët[1], P.E. Morange[3].* 1) INSERM UMR_S 937; ICAN Institute, Université Pierre et Marie Curie, Paris, Paris, France; 2) Inserm, U1018, Universit e Paris Sud 11, Villejuif, France Villejuif, France; 3) INSERM, UMR_S 626, Marseille, France; Université de la Méditerranée, Marseille, France; 4) INSERM, U897, Bordeaux, France; Université Victor Segalen, Bordeaux, France; 5) Assistance Publique des Hopitaux de Paris, Laboratoire d'hématologie, Hôpital Européen Georges Pompidou, Paris, France; 6) INSERM U888, Hôpital La Colombière, Montpellier, France; 7) INSERM UMR_S 708, Université Pierre et Marie Curie (UPMC, Paris 6), Paris, France; 8) INSERM U744, Lille, France; Institut Pasteur de Lille, Lille, France Université de Lille Nord de France, CHRU de Lille, Lille, France.

High thrombin generation levels have been associated with the risk of venous thrombosis and ischemic stroke. In order to investigate the genetic architecture underlying the inter-individual variability of thrombin generation, we measured three quantitative biomarkers for thrombin generation, the Endogeneous Thrombin Potential (ETP), the Peak height and the Lag-time in two independent samples, the Three City Study and MARTHA, with genome-wide genotype data. By applying an imputation strategy based on the 1000 Genomes project to these two samples totalling 2,100 individuals inferred for ~6,6M markers, genome-wide significant associations were observed for ETP and Peak at the F2 gene (p = 4.62 10-22 and p = 1.05 10-8, respectively), a well-established susceptibility locus for thrombin generation. Further conditional analysis on the F2 signal revealed suggestive evidence (p < 10-6) at 6 additional independent single nucleotide polymorphisms (SNPs) for at least one of the three phenotypes. These 6 SNPs were tested for replication in a third independent sample of 796 individuals part of the MARTHA12 study. For one SNP located on chromosome 9, the association with Lag-time was highly significant (p = 3.26 10-7). We further replicated this association (p = 4.8 10-3) in another independent sample of 530 healthy patients from the FITE-NAT study. When the results of the four data sets were combined, the overall evidence for association at the chr. 9 locus with Lag-time reached p = 1.71 10-14. This locus has never been reported associated to thrombin generation, paving the way for novel mechanistic pathways in the aetiology of thrombotic disorders.

122

Genome-wide association analysis of blood pressure traits in nearly 30,000 African ancestry individuals reveals a common set of associated genes in African and non-African populations. *N. Franceschini[1], E. Fox[2], Z. Zhang[3], T. Edwards[4], M. Nalls[5], Y. Sung[6], B. Tayo[7], Y. Sun[8], O. Gottesman[9], A. Adeyemo[10], A. Johnson[11], J. Young[12], K. Rice[13], H. Tang[14], J. Smith[15], G. Ehret[16], A. Morrison[17], E. Boerwinkle[17], B. Psaty[18], D. Arnett[19], S. Redline[20], R. Cooper[7], N. Risch[21], D. Rao[6], J. Rotter[22], A. Chakravarti[16], A. Reiner[23], D. Levy[11], B. Keating[24], X. Zhu[3], the AGEN Consortium.* 1) Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Medicine, University of Mississippi Medical Center, Jackson, MS; 3) Department of Epidemiology & Biostatistics, School of Medicine, Case Western Reserve University, Cleveland, OH; 4) Center for Human Genetics Research, Vanderbilt Epidemiology Center, Department of Medicine, Vanderbilt University, Nashville, TN; 5) Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD; 6) Division of Biostatistics, Washington University School of Medicine, St. Louis, MO; 7) Department of Preventive Medicine and Epidemiology, Loyola University Chicago Stritch School of Medicine, Maywood, IL; 8) Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA; 9) The Charles Bronfman Institute for Personalized Medicine, Mount Sinai School of Medicine, New York, NY; 10) Center for Research on Genomics and Global Health, National Human Genome Research Institute, Bethesda, MD; 11) Center for Population Studies, National Heart, Lung, and Blood Institute, Framingham, MA; 12) Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 13) Department of Biostatistics, University of Washington, Seattle, WA; 14) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 15) Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI; 16) Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 17) Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston Houston TX; 18) Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA; 19) Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL; 20) Department of Medicine, Harvard Medical School, Boston, MA; 21) Institute for Human Genetics, Department of Epidemiology and Biostatistics, University of California, San Francisco CA; 22) Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, CA; 23) Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 24) Department of Pediatrics, University of Pennsylvania, Philadelphia, PA.

High blood pressure (BP) is more prevalent and contributes to more severe manifestations of cardiovascular disease among African Americans compared with other U.S. ethnic/racial groups. Thus far, African ancestry BP genome-wide association studies (GWAS) have identified few validated variants for BP. We report on a large GWAS meta-analysis of BP in individuals of African ancestry that includes 29,378 individuals from 19 discovery cohorts, followed by replication in additional samples of individuals of African (n=10,386), European (n=69,395), and East Asian (n=19,601) ancestry. In the discovery African ancestry meta-analyses, we selected 45 independent single nucleotide polymorphisms (SNPs) at a pre-specified threshold (P<1.0×10-5) that were carried forward for replication. SNPs in five loci replicated in African, European or East Asians when adjusting for multiple testing, reaching genome-wide significance for either systolic or diastolic BP in trans-ethnic meta-analysis (EVX1-HOXA, P=2.1×10-12, ULK4, P= 2.1×10-13, RSPO3, P=2.43×10-11, PLEKHG1, P=1.9×10-11, SOX6, P= 5.12×10-10). Three of these BP loci (EVX1-HOXA, RSPO3 and PLEKHG1) have not been previously reported for BP traits. We also identified allelic heterogeneity at the SOX6 locus, and provided evidence for fine mapping in four validated BP loci. Our findings suggest that BP loci may have significant effects across populations of different ancestry and also demonstrate that multi-ethnic samples are a key component in identification, fine mapping and understanding trait variability.

123

Gene pathway burden test application to cardiovascular disease using whole genome sequencing data. *M.AA. Almeida[1], J.P. Peralta[2], J.W. Kent[1], T.M. Teslovich[3], G. Jun[3], C. Fuchsberger[3], A.R. Wood[4], A. Manning[5], T.M. Frayling[4], P. Cingolani[6], D.M. Lehman[7], T.D. Dyer[1], G. Abecasis[3], L. Almasy[1], R. Duggirala[1], J. Blangero[1].* 1) Texas Biomedical Research Institute, Genetics Department, 7620 NW Loop 410, San Antonio, TX., USA; 2) Centre for Genetic Epidemiology and Biostatistics, University of Western Australia, 35 Stirling Highway Crawley, WA, Australia; 3) University of Michigan, Ann Arbor, MI, USA; 4) University of Exeter, Exeter, The Queen's Dr Exeter, United Kingdom; 5) Broad Institute, 7 Cambridge Center, Cambridge, MA 02142, USA; 6) McGill University, 845 Sherbrook Street West Montréal, Canadá; 7) University of Texas Health Center at San Antonio, 7703 Floyd Curl Dr , San Antonio, TX, USA.

The advent of whole genome sequencing provides many opportunities for understanding the source of causal variation underlying complex diseases. As part of the T2D-GENES Consortium, we have directly sequenced 590 individuals (and accurately imputed another 448 members) from 20 large Mexican American pedigrees to try to understand the role of rare and private variants in type 2 diabetes (T2D). Those individuals are part of the SAFS (San Antonio Family Study) and many phenotypes have been measured on these individuals. We have observed ~22 million single nucleotide variants (SNVs) in this sample. Such a large amount of data imposes statistical and analytical barriers that require the development of alternative approaches that allow a fast and sensible screening of potential causal genes and pathways. While gene-centric testing is now common, less effort has been placed to date on formal tests of the contribution of sequence variation in gene pathways to complex disease risk. We have developed a single degree-of-freedom test using a random effect model based on an empirical pathway-specific genetic relationship matrix (GRM) as the focal covariance kernel. The empirical pathway-specific GRM (the PSGRM) utilizes all variants (or a chosen likely functional subset) identified in gene members of a given biological pathway and is tested by the use of a LRT (Likelihood Ratio Test). Gene pathway definitions were obtained from the latest KEGG database release and a PSGRM was estimated for each pathway. Those pair-wise relationship definitions were tested using cardiovascular disease (CVD, defined as ECG-derived evidence of myocardial infarction, history of surgery related to atherosclerosis, and CVD-related mortality) as our focal trait. The glycerolipid metabolism pathway exhibited a significant association (p = 0.00087) and “absorbed” most of the observed CVD’s heritability in this sample. This gene pathway is composed by 50 genes and a set of 43,819 SNVs were employed in the PSGRM calculation. Non-synonymous variants in this pathway that were predicted to be highly deleterious (PolyPhen-2 score > 0.8) were independently tested and a promising association with CVD was detected in the gene ALDH7A1, a gene that has previously been associated with lipid variation in another independent study. Our results suggest that our simple pathway-based test may be useful for reducing the search space for specific functional variants influencing complex phenotypes.

124

Interpreting eQTLs by linking enhancers to target genes. *J. Wang[1,2], A. Kundaje[1,2,3,4], L.D Ward[1,2], M. Kellis[1,2], GTEx Consortium and Roadmap Epigenomics Program.* 1) Computer Science Dept, Massachusetts Institute of Technology, Cambridge, MA 02139; 2) Broad Institute of MIT and Harvard, Cambridge, MA 02139; 3) Computer Science Dept., Stanford University, Stanford, CA, 94305; 4) Dept. of Genetics, Stanford University, Stanford, CA, 94305.

It is a challenge to interpret the downstream effects of genetic variants located in non-coding regulatory regions, because the target genes of those regulatory elements may not be the most proximal gene and the regulatory relations can be cell-type specific. Thus, a more complete and accurate linking map between distal regulatory elements and their specific target genes is necessary to better understand the eQTL relations. Enhancers represent an important family of distal regulatory elements. By exploring the dynamics of enhancer activities, along with cell-type specific gene expression patterns, we can statistically link enhancers to their target genes, which provide a basis to understand eQTLs. We have developed a generative model to probabilistically assign enhancers and genes into modules, and estimating the linking probabilities between enhancers and genes jointly. Applying the model on the histone modification and gene expression datasets of 26 cell-types from NIH Roadmap Epigenomics, we discovered 21 enhancer modules, 12 gene modules and enhancer-gene linking probabilities from this dataset. As performance validation, the cell-type specific linking is compared with ChIA-PET datasets and show significant overlaps in matched cell-types. Furthermore, the predicted linking relations are verified by checking whether they can accurately quantify gene expressions based on enhancer activities. We observe clear improvements compared to the results of correlation-based methods. The genes that are most affected by linked enhancers are enriched in cell-type specific pathways. We compared the predicted linking between enhancers and genes to the eQTL linking between SNPs and gene expressions from GTEx project. We observe a number of overlapped linking relations which provide direct interpretations to those eQTLs. Interestingly, among those overlapped linking relations, we observe examples of enhancers linked to the genes which are not the most proximal. For example, six enhancers linked to CD52 overlapped with eQTLs from whole blood cells. For two of those enhancers, CD52 is not the most proximal gene, supporting the value of long distance enhancer linking. Similar overlapped long distance linking examples are also observed for CD48, CD37 and LCK, which are all related to immune functions. We believe that, by integrating our predicted enhancer-gene linking structure, people can better interpret and prioritize eQTLs involved in long distance regulations.

125

Genetic architecture of regulatory variation influencing response to human rhinovirus infection. *M. Caliskan[1], Y. Gilad[1], C. Ober[1,2].* 1) Human Gen, Univ Chicago, Chicago, IL; 2) Obstetrics and Gynecology, Univ Chicago, Chicago, IL.

Human rhinovirus (HRV) is the most prevalent human respiratory virus. Each year, HRV infects billions of people and is responsible for at least half of all common colds, the most common illness of human. In addition to common colds, HRV infection affects the morbidity of a range of respiratory illnesses, including bronchiolitis, pneumonia, asthma, chronic obstructive pulmonary disease, and cystic fibrosis. Despite its biological importance, little is known about the genetic architecture of variation in response to HRV. To address this, we obtained genome wide genotype and gene expression data in paired peripheral blood mononuclear cell (PBMC) samples from 98 individuals; PBMCs were cultured for 24 hours with and without HRV. Among the 10,868 genes detected as expressed, 2,093 were significantly up-regulated and 3,464 were down-regulated in HRV-infected compared to uninfected PBMCs (Bonferroni corrected significance threshold). Among the genes differentially expressed, the top enriched canonical pathways included activation of IRF by cytosolic pattern recognition receptors, interferon signaling, role of RIGI-like receptors in antiviral innate immunity, and IL-12 signaling and production in macrophages. We next mapped local (cis) expression quantitative trait loci (eQTLs) in uninfected and HRV-infected PBMCs, and identified 590 genes with cis-eQTLs in uninfected and 530 genes with cis-eQTLs in HRV-infected PBMCs (FDR 1% threshold). More importantly, 32 genes had significantly different effects in untreated and treated cells, providing evidence for genotype x HRV interactions. These HRV-interacting regulatory variations were significantly enriched for lung function (FEV1/FVC ratio) GWAS P-values <0.10 in the Hutterites (Permutation P=0.0041), and pulmonary function genes reported in the NHGRI GWAS catalogue were enriched for genotype x HRV interaction P-values <0.10 (Permutation P=0.045). In summary, this study represents the first integrated genome-wide study of genetic variation and gene expression response to HRV infection and suggests that host genotypes that are associated with gene expression response to HRV have long-term and persistent effects on lung function phenotypes.

126

Genome-wide association of expression response of primary immune cells identifies novel cis and trans loci specific to distinct pathogen responses. *C. Ye[1], M. Lee[1,2,5,7], A.C. Villani[1], T. Raj[1,2,3], W. Li[1,5], T.M. Eisenhaure[1,5], S.H. Imboywa[3], P.I. Chipendo[3], K. Rothamel[4], K. Raddassi[3], M.H. Lee[3], I. Wood[3], C. McCabe[1], B.E. Stranger[6,10], C.O. Benoist[4], P.L. De Jager[1,2,3], A. Regev[1,8,9], N. Hacohen[1,2,5], Immunological Variation Consortium.* 1) Broad Institute, Cambridge, MA; 2) Harvard Medical School, Boston, MA; 3) Department of Neurology, Brigham and Women's Hospital, Boston, MA; 4) Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA; 5) e, Massachusetts General Hospital, Boston, MA; 6) Section of Genetic Medicine, University of Chicago, Chicago, IL; 7) Harvard-MIT Healthy Sciences and Technology, Boston, MA; 8) Howard Hughes Medical Institute, Chevy Chase, MD; 9) Department of Biology, Massachusetts Institute of Technology, Cambridge, MA; 10) Institute of Genomics and Systems Biology, University of Chicago.

The rising prevalence of many chronic autoimmune diseases suggests that the interaction between genetic variations in innate immune genes and their environment may play a role in defining disease pathogenesis. While much heterogeneity exists in environmental components in natura, which is often hard to assess systematically, we can study variant-environment interactions by stimulating immune cells with well-defined and controlled pathogen input, and subsequently measure the response as the output. To leverage this approach, we performed a systematic study aimed at identifying expression response quantitative trait loci (reQTLs) in anti-bacterial and anti-viral immune responses. Using a 415-gene signature, we profiled the response of primary monocyte-derived dendritic cells (MoDCs) activated by lipopolysaccharide (LPS), influenza and IFNβ stimulation (a shared downstream cytokine of the anti-viral and anti-bacterial pathways) in 560 individuals of 3 ethnicities. We identified 119/415 (29%) genes to have significant (FDR < 0.05) cis-reQTLs, most of which are shared between stimuli due to a common IFN response. Using imputation, multiethnic meta-analysis and ENCODE annotations, we identified potential causal variants that perturb the canonical interferon-stimulated response element (ISRE) in three immune response genes. These reQTLs were validated by measuring the ability of each allele to drive reporter gene expression in IFNβ-stimulated cells. For one locus we show differential IFNβ-stimulated binding of a critical transcription factor to the major and minor variants of the ISRE in vitro. We identified GWAS variants in 32 unique genes to be cis-reQTLs, including associations to Crohn's and SLE variants, under influenza and IFNβ stimulation. These SLE variants were also associated in trans with known downstream targets (interferon stimulated genes) under influenza infection, which we successfully validated through functional experimentations. Taken together, we demonstrate widespread expression response QTLs in core innate immune response programs to bacteria and viruses. By studying stimulated cells with defined environmental input and leveraging naturally occurring genetic variations, we identified known and new biological relationships in innate immune pathways that have important implications in defining key mechanisms involved in immune disease.

127

Expression QTL analysis from primary immune cells of a multi-ethnic cohort identifies novel disease-causing regulatory effects. *B.E. Stranger[1,2,3,4,5], T. Raj[3,4,5,6], C. Ye[3], S. Mostafavi[7], K.L. Rothamel[8], M. Lee[3,5], J.M. Replogle[3,4,6], T. Feng[8], S.H. Imboywa[6], M. Lee[6], C. McCabe[3,6], D. Koller[7], A. Regev[3], N. Hacohen[3,9], C.O. Benoist[3,5,8], P.L. De Jager[3,4,5,6], The Immunological Variation Consortium.* 1) Section of Genetic Medicine, University of Chicago, Chicago, IL; 2) Institute for Genomics and Systems Biology, University of Chicago, Chicago, IL; 3) The Broad Institute of Harvard and MIT, Cambridge, MA; 4) Division of Genetics, Brigham and Women's Hospital, Boston, MA; 5) Harvard Medical School, Boston, MA; 6) Department of Neurology, Brigham and Women's Hospital, Boston, MA; 7) Department of Computer Science, Stanford University, Stanford, CA; 8) Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA; 9) Department of Medicine, Massachusetts General Hospital, Boston, MA.

The extent to which human genetic variation drives variation in immune function is critical to the systematic dissection of altered immune function underlying complex immune-mediated disease. As part of The Immunological Variation Consortium, we have performed an expression quantitative trait locus (eQTL) study in two primary human immune cell-types representing adaptive and innate immune function. Gene expression levels were quantified in highly purified CD4+ T cells and CD14+CD16- monocytes from 162 individuals of African American, 155 East Asian and 377 European ancestry, that were genotyped and imputed to approximately 10M SNPs.

We show that the small differences (6-10%) in *cis*-eQTL found among populations are driven primarily by population divergence in allele frequencies. Similarly, after conducting a multi-ethnic meta-analysis for both T cells and monocytes, we find that most (63%) *cis*-eQTLs are shared between the two cell types. Combining data across populations and cell types increases the total number of *cis*-associations (6,546 and 5,632 genes at FDR 0.05 in monocytes and T cells, respectively), as well as enhances our ability to refine the location of the associations due to differences in LD among populations.

We identified 892 trait-associated SNPs that are also significant eQTLs, where 33% of the eQTLs have not been previously reported. We find that susceptibility alleles for inflammatory diseases display a preponderance of regulatory effects in T cells, the adaptive arm of the immune system. In exception to this trend, Alzheimer's and Parkinson's disease (PD) susceptibility alleles show a striking enrichment of effects in monocytes, the innate arm of the immune system. We report evidence for several disease-associated variants affecting RNA expression in *trans*: for example, the PD-associated MAPT H1 haplotype drives the expression of MAPK8IP1 ($p=3.48 \times 10^{-32}$), and a shRNA knockdown experiment suggests the effect is mediated by MAPT expression and not one of the other 5 genes found in this haplotype. Our multi-ethnic exploration of both arms of the immune system highlights a remarkable level of *cis*-eQTL sharing across populations and cell types, and provides important new insights into the role of immune cell populations in the onset of neurodegenerative, inflammatory, and other diseases.

128

Allele Specific Expression Analysis Using Transcriptome Sequencing in Three Tissues of a Twin Cohort Reveals Large Effect of Gene-by-Gene and Gene-by-Environment Interactions. *A. Buil[1,2,3], A.A. Brown[4], A. Viñuela[5], M.N. Davies[5], H.F. Zheng[6], J.B. Richards[5,6], K.S. Small[5], R. Durbin[4], T.D. Spector[5], E.T. Dermitzakis[1,2,3].* 1) Genetics and Development, University of Geneva, Geneva, Switzerland; 2) Institute of Genetics and Genomics in Geneva, University of Geneva, Geneva, Switzerland; 3) Swiss Institute of Bioinformatics, Switzerland; 4) Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 5) Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom; 6) Departments of Medicine, Human Genetics, Epidemiology and Biostatistics McGill University, Canada.

While allele specific expression is expected to result from genetic regulatory variants, a proper estimation and dissection of the causes has not been performed to date. In this study we used RNA-seq data from fat, LCLs and skin from ~400 female MZ and DZ twin pairs (2330 RNA-seq samples in total) to quantify allelic specific expression and to dissect its underlying causes . First, we performed eQTL analysis and we found 9861 genes with at least a significant cis eQTLs for fat, 10015 for LCL and 9243 for skin (FDR=1%). Then, we estimated ASE at every heterozygous site for every individual. At a 10% FDR, we observed a significant ASE effect in 9.5% of the transcript heterozygous sites in fat, 9.3% in LCL and 9.1% in skin. ASE may be caused by genetic or epigenetic /environmental factors. To measure the relative contribution of the underlying causes of allelic expression we estimated the variance components of the ASE ratios using the identity-by-descended status (IBD) of the twin pairs at the ASE site and the identity-by-state status (IBS) at the best eQTL. We found that about 35% of the variance in ASE is due to the effect of the best eQTL , 14% to the additive effect of the other genetic variants in cis, 33% to the interaction between cis and trans variants and 18% to the individual environment. The additive trans and the shared environmental effects were negligible. There were small differences among tissues. The sum of all the genetic effects gives an average heritability estimate of 73% for fat, 86% for LCL and 83% for skin. Our results show a complex genetic architecture for allelic expression that identifies GxG and putative GxE effects. We took advantage of the twin structure of our sample to look for examples of GxE interactions. Since MZ twins are genetically identical, differences in allelic expression in a MZ pair have to be caused by non genetic effects. For every site, we calculated the association between allelic expression differences within MZ pairs and SNPs around the site and found examples of potential GxE interactions. One example in fat tissue was found for ADIPOQ, a gene that codifies for adiponectin, whose expression has been observed to be affected by environmental factors such as diet and physical exercise. We are exploring further to find putative GxG interactions affecting allelic expression.

129

Epigenomic variation between species, tissues, populations and individuals. *A. Kundaje[1,2], W. Meuleman[1,2], J. Wang[1,2], N. Kumar[1], S. Kyriazopoulou-Panagiotopoulou[3], M. Kasowski[4], M. Snyder[4], M. Kellis[1,2].* 1) Computer Science Dept., Massachusetts Institute of Technology, Cambridge, MA. 02139; 2) Broad Institute of MIT and Harvard, Cambridge, MA 02139; 3) Computer Science Dept., Stanford University, Stanford, CA, 94305; 4) Dept. of Genetics, Stanford University, Stanford, CA, 94305.

In multicellular organisms, epigenetic information is a key enabler of dynamic regulatory regions shaping the identity of each cell. This information is encoded in distinct combinations of epigenetic modifications defining 'chromatin states' specific to different types of functional elements such as promoters, enhancers, transcribed elements and repressed domains. Here, we provide the first systematic analysis of chromatin state dynamics across 3 axes of variation: inter-species variation, inter-cell type variation in the same species, and inter-individual variation for the same cell type. First, we used genome-wide maps of histone modifications from the mod/ENCODE consortia across human, fly, and worm to learn unified chromatin state models and study the similarities and differences in the specific mark combinations that define different types of functional elements. We found remarkable conservation of the chromatin code across the 3 species, for active enhancers, promoter, and transcribed regions, illustrating a universality of chromatin states across the animal kingdom. However, we also find a notable exception in the marking of repressed regions, where repressed Polycomb and heterochromatin domains that are distinct in fly and human appear to co-associate in worm, indicating potentially different mechanisms of repression. Second, we learned robust chromatin state maps in 90 human cell types using data from the Roadmap Epigenomics consortium that enabled us to construct detailed lineage trees relating the different cell-types. We developed novel predictive models to link distal enhancers to their target genes that revealed extensive cell-type specific activity of both enhancer and promoter regions associated with distinct cell-type specific regulatory programs. Finally, we learned chromatin state maps jointly across lymphoblastoid cell lines from 19 individuals spanning diverse ancestry. We find that enhancer regions are the most variable functional elements across individuals, showing significant switching between active, weak, bivalent/poised, and repressed states. In contrast, promoter and transcribed states showed negligible variation across individuals and populations, suggesting that the observed enhancer variation is inconsequential or potentially buffered by enhancer redundancy. Together, these analyses provide key insights into the conservation and variation of chromatin regulation and dynamics across organisms, cell-types and individuals.

130

Predicting genome-wide DNA methylation using methylation marks, genomic position and DNA regulatory elements. *W. Zhang[1], TD. Spector[2], P. Deloukas[3], JT. Bell[2], BE. Engelhardt[4,5,6].* 1) Department of Molecular Genetics and Microbiology, Duke University, Durham, NC; 2) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 3) Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK; 4) Department of Biostatistics & Bioinformatics, Duke University, Durham, North Carolina, USA; 5) Department of Statistical Science, Duke University, Durham, North Carolina, USA; 6) Institute for Genome Sciences & Policy, Duke University, Durham, North Carolina, USA.

DNA methylation is one of the most studied epigenetic modifications of DNA, and is known to have a role in cellular processes and complex traits and disease, including cancer. Recent assays for individual-specific fine-scale DNA methylation profiles across genome-wide CpG sites have enabled epigenome-wide association studies to identify specific CpG sites associated with a phenotype. To expand these studies, computational prediction of site-specific methylation status is of great interest, but approaches to date predominantly tackle methylation within a genomic locus using DNA sequence content as features and are often limited to specific genomic regions. Using data from the Illumina 450K methylation array for whole blood samples from 100 individuals, we identify striking correlation patterns of DNA methylation specific to CpG islands (CGIs), CGI shores, and non-CGIs. For example, we see what appears to be a circular pattern of correlation across the CGI shore and shelf regions. As compared to single nucleotide polymorphisms (SNPs), where linkage disequilibrium induces correlation between SNPs, correlations between neighboring CpG sites decays rapidly with genomic distance, making CpG sites less predictive of their neighboring sites, especially in regions of sparse coverage on the array. Based on these findings, we predict CpG site methylation levels using a random forest classifier, using as features neighboring CpG site methylation levels and genomic distance, and co-localization with coding regions, CGIs, and regulatory elements from the ENCODE project, among others. Our approach achieves 91%-94% prediction accuracy of genome-wide methylation levels at single CpG site precision with higher accuracy when restricting the genomic distance of neighboring CpG sites. The accuracy increases to 98% when restricted to CpG sites within CGIs. Our classifier outperforms state-of-the-art methylation classifiers and is interpretable by identifying features that contribute to prediction accuracy. Neighboring CpG site methylation status, CpG island status, co-localized DNase I hypersensitive sites, and transcription factor binding sites including Elf1(ETS-related transcription factor 1), MAZ(Myc-associated zinc finger protein), Mxi1(MAX-interacting protein 1) and Runx3(Runt-related transcription factor 3) were found to be the most predictive features of methylation levels, suggesting an interacting role for these elements in epigenetic modification and regulation.

131

An ENU mutagenesis screen identifies the first mouse mutants of a novel epigenetic modifier, Rearranged L-Myc Fusion (Rlf). *S.K. Harten[1], L. Bourke[1], V. Bharti[1], H. Oey[2], N. Whitelaw[1], L. Daxinger[2], E. Whitelaw[1,2].* 1) Genetics & Population Health, QIMR, Brisbane, QLD, Australia; 2) La Trobe Institute of Molecular Science, Melbourne, VIC, Australia.

An ENU mutagenesis screen was established to identify novel genes involved in epigenetic reprogramming. The screen utilizes a transgenic GFP reporter gene, which is expressed in a variegated manner in erythrocytes and is highly sensitive to epigenetic changes. A gene discovery pipeline involving SNP arrays and whole exome sequencing has enabled identification of causative mutations in ~40 mutant lines. To date the screen has produced mouse mutants of both known modifiers of epigenetic state, such as *Dnmt1* and *Smarca5*, and genes not previously implicated as epigenetic modifiers, such as Re-arranged L-Myc Fusion (*Rlf*). Here we report three independent lines with mutations in *Rlf*. Each line shows a reduced percentage of GFP expressing cells compared to wild-types. Homozygous *Rlf* mutants show increased methylation at the transgene and haploinsufficiency for Rlf alters transcriptional silencing at Agouti Viable Yellow (A^{vy}), an independent endogenous epiallele. Taken together, these findings suggest that *Rlf* is a modifier of epigenetic state. *Rlf* is conserved between human and mouse and contains multiple widely-spaced zinc finger domains, however little else is known. Our studies revealed reduced body-weight and postnatal lethality in *Rlf* homozygous null mutants, indicating that Rlf is critical for proper development. Histological analysis of mid-gestation *Rlf* mutant embryos revealed the presence of a heart defect. Analysis of RNA-Seq data, comparing RNA from wild-type and Rlf mutant fetal livers, showed differential expression of genes involved in metabolism. Levels of 4-hydroxyphenylpyruvic acid dioxygenase (HPD), an enzyme involved in tyrosine catabolism, were markedly reduced. Genetic deficiency of *HPD* underlies Tyrosinemia type 3 in humans and is associated with mental retardation and ataxia. Genome-wide bisulphite sequencing studies are underway to examine the role of Rlf in the control of DNA methylation both within and outside of CpG islands. The functional effects of non-CpG island methylation, both on gene expression and development are also being investigated.

132

Zebrafish Mutation Project: Functional Genomics of Disease. *E.M. Busch-Nentwich, J.E. Collins, I. Sealy, N. Wali, R.J. White, C.M. Dooley, C. Scahill, S. Carruthers, Z. Pusztai, C. Herd, A. Hall, R.N.W. Kettleborough, J. Morris, J. Barrett, D.L. Stemple.* Wellcome Trust Sanger Institute, Hinxton, United Kingdom.

The advent of high-throughput sequencing has greatly accelerated the identification of inherited and *de novo* disease causing mutations. Following discovery the analysis of the developmental and cellular pathways of the affected genes is a crucial step on the path towards therapy with model organisms as the central tool. Traditionally, vertebrate model organisms such as mouse and zebrafish have been used on a gene-by-gene basis, however, in order to keep pace with the increasing speed of discovery, new approaches are needed. Owing to a high quality genome reference sequence and its genetic and embryological tractability the zebrafish is a vertebrate model especially suited for large scale studies. Previously having established methods to generate and identify disruptive zebrafish point mutations on a genome wide scale the Zebrafish Mutation Project is now assigning biological function to every protein-coding gene in the zebrafish genome. We submit alleles to a high-throughput assessment of morphological phenotypes which is followed by the quantitative analysis of genome wide transcriptional changes in response to the loss of function. For this transcriptome analysis we have developed a new sequence based method, the differential expression transcript counting technique (DETCT). Using this, we find a wealth of genes displaying alterations in transcript levels reflecting the observed morphological changes. Ontology term enrichment analysis on gene ontology (GO) annotations combined with the zebrafish anatomical and development (ZFA) ontology has led to surprisingly detailed insights into phenotypes. Thus far two general trends are emerging. Firstly, transcript profiles for previously uncharacterised mutants confirm predicted cellular function and show tissue-specific effects on transcript abundance, thus providing mechanistic evidence. Secondly, we are beginning to build pathway-specific gene networks. Transcriptome analysis of mutants has revealed novel candidate genes which, when mutated, lead to a phenotype affecting the same developmental pathway. Our approach and the results will be discussed especially in the context of gene families implicated in human disease.

133

Unraveling the genetic architecture of Multiple Sclerosis and the underlying implicated pathways. *NA. Patsopoulos[1,2,3,4] for the International Multiple Sclerosis Genetics Consortium (IMSGC).* 1) Department of Neurology, Brigham & Women's Hospital, Boston, MA; 2) Division of Genetics, Department of Medicine, Brigham & Women's Hospital, Boston, MA; 3) Harvard Medical School, Boston, MA; 4) The Broad Institute, Cambridge, MA.

Multiple sclerosis (MS) is a neurodegenerative disease with a genetic background. So far efforts to identify the genetic component using genome-wide association studies (GWAS) has a yielded ~50 loci. Here we present results of the International Multiple Sclerosis Genetics Consortium's (IMSGC) analysis of all available published GWAS enriched with new unpublished data and imputed in the 1000 genomes European panel. The final analysis was performed in 14,802 multiple sclerosis (MS) cases and 26,703 controls and ~8 million SNPs. We applied an exhaustive strategy to identify primary and secondary signals, identifying 78 primary and another 10 secondary genome-wide effects (p-value < 5×10-8). At the less conservative level of p-value < 10-5, we found another 96 primary effects and 30 secondary ones. By applying extended stepwise models we identified overall ~4.700 statistically independent effects at the nominal level (p-value<0.05). We replicated these in a custom designed iSelect chip (~80K SNPs) genotyping 18,000 cases and 18,000 controls and we present detailed results on the replication success rates. This large sample size allows us to quantify the exact variance explained and heritability by each susceptibility allele using multivariate models, and eventually estimate the overall contribution of common genetic variability to MS susceptibility. Furthermore, we describe the underlying mechanisms and pathways implicated by the newly identified loci and how these enrich and complete the current knowledge about the role of innate and humoral immunity. Finally, we leverage eQTL studies and the ENCODE and NIH Epigenomics Roadmap data to characterize the functional implications of the associated loci and nominate the most likely candidate causal variant in each locus. By performing the largest experiment to date in the genetic analysis of common variation in MS, we describe a holistic approach to identify and functionally annotate the most complete genetic map of MS susceptibility.

134

Autosomal dominant congenital spinal muscular atrophy is caused by mutations in BICD2, a golgin and important motor adaptor. *B. Wirth[1], L.A. Martinez-Carrera[1], I. Hölker[1], A. Heister[2], A. Verrips[3], S.M. Hosseini-Barkooie[1], C. Gilissen[3,4,5,6], S. Vermeer[3,4], M. Pennings[3], R. Meijer[3], M. te Riele[7], C.J.M. Frijns[8], O. Suchowersky[9], L. MacLaren[10], S. Rudnik-Schöneborn[11], R.J. Sinke[12], K. Zerres[11], R.B. Lowry[10], H.H. Lemmink[12], L. Garbes[1], M. Synofzik[13], J.A. Veltman[3,4,5], H.J. Schelhaas[7], H. Scheffer[3,4], K. Neveling[3,4].* 1) Institute of Human Genetics, Institute of Genetics and Center for Molecular Medicine Cologne, University of Cologne, Cologne, Germany; 2) Dept. of Human Genetics, Radboud University Medical Centre, Nijmegen, The Netherlands; 3) Dept. of Pediatric Neurology, Canisius-Wilhelmina Hospital, Nijmegen, The Netherlands; 4) Inst. for Genetic and Metabolic Disease, Radboud University Medical Centre, Nijmegen, The Netherlands; 5) Nijmegen Centre for Molecular Life Sciences, Radboud University Medical Centre, Nijmegen, The Netherlands; 6) Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway; 7) Department of Neurology, Radboud University Medical Centre, Nijmegen, The Netherlands; 8) Dept. of Neurology, University Medical Centre Utrecht, Utrecht, the Netherlands; 9) Dept. of Medicine (Neurology), Medical Genetics, and Psychiatry, University of Alberta, Edmonton Alberta, Canada; 10) Dept. of Medical Genetics and Pediatrics, Alberta Childrens Hospital, University of Calgary, Calgary, Canada; 11) Institute of Human Genetics, University Aachen, Aachen, Germany; 12) Dept of Genetics, University of Groningen and University Medical Center Groningen, Groningen, The Netherlands; 13) Hertie-Institute for Clinical brain Research and Center for Neurology, Department of Neurodegenerative Diseases, University of Tübingen, Germany.

Spinal muscular atrophy (SMA) is a heterogeneous group of neuromuscular disorders caused by degeneration of lower motor neurons. While functional loss of SMN1 is associated with autosomal recessive childhood SMA, the genetic cause for most families with a dominant inheritance of SMA is unknown. Here, we identified pathogenic variants in the bicaudal D2 (*BICD2*) in families with autosomal dominant SMA. Affected individuals present with congenital slowly-progressive muscle weakness mainly of lower limbs and congenital contractures. In a large Dutch family, linkage analysis identified a locus on chromosome 9q22.3, in which exome sequencing uncovered the c.320C>T (p.Ser107Leu) mutation in *BICD2*. In 3/24 families with dominant SMA additional pathogenic variants c.2108C>T (p.Thr703Met), c.563A>C, (p.Asn188Thr) and c.2239C>T (p747R>C) were identified. Complete sequencing of the coding region of *BICD1* revealed no pathogenic variant, excluding this *BICD2* paralog as a further candidate gene. BICD2 is a golgin and motor adaptor protein involved in Golgi dynamics and in vesicular and mRNA transport. Transient transfection of HeLa cells with all three mutant *BICD2* cDNAs caused massive Golgi fragmentation. This observation was even more prominent in primary fibroblasts, especially in the patient carrying the mutation p.Thr703Met localized in C-terminal coiled-coil domain. Furthermore, BICD2 was reduced in affected individuals and trapped within the fragmented Golgi. Previous studies have shown that Drosophila mutant BicD causes reduced larvae locomotion by impaired clathrin-mediated synaptic endocytosis in neuromuscular junctions. *In vivo* studies in zebrafish overexpressing each of the *BICD2* mutants or wt RNA are in progress. *BICD2* is the first gene described to be involved in an autosomal dominant inherited SMA and its interacting partners are further excellent candidates for the many unsolved SMA cases.

135

Rare variants in restless legs syndrome. *E.C. Schulte[1,2], M. Kousi[3], B. Schormair[2,4], F. Knauf[2], P. Lichtner[2], C. Trenkwalder[5,6], B. Högl[7], B. Frauscher[7], K. Berger[8], I. Fietze[9], N. Gross[1], M. Hornyak[1,10,11], K. Stiasny-Kolster[12,13], W. Oertel[13], C.G. Bachmann[14], W. Paulus[15], A. Zimprich[16], A. Peters[17], C. Gieger[18], T. Meitinger[2,4], B. Müller-Myhsok[19], N. Katsanis[3], J. Winkelmann[1,2,4].* 1) Neurologische Klinik und Poliklinik, Technische Universität München, Munich, Germany; 2) Institut für Humangenetik, Helmholtz Zentrum München, Munich, Germany; 3) Center for Human Disease Modelling, Department of Cell Biology, Duke University, Durham, NC, USA; 4) Institut für Humangenetik, Technische Universität München, Munich, Germany; 5) Paracelsus Elena Klinik, Kassel, Germany; 6) Klinik für Neurochirurgie, Georg August Universität, Göttingen, Germany; 7) Department of Neurology, Medizinische Universität Innsbruck, Innsbruck, Austria; 8) Institut für Epidemiologie und Schlafmedizin, Westfälische Wilhelms Universität Münster, Münster, Germany; 9) Zentrum für Schlafmedizin, Charité Universitätsmedizin, Berlin, Germany; 10) Interdisziplinäres Schmerzzentrum, Albert-Ludwigs Universität Freiburg, Freiburg, Germany; 11) Algesiologikum München, Munich, Germany; 12) Schlaflabor Marburg, Marburg, Germany; 13) Klinik für Neurologie, Philipps Universität Marburg, Marburg, Germany; 14) Abteilung für Neurologie, Paracelsus Klinikum Osnabrück, Osnabrück, Germany; 15) Abteilung für Neurophysiologie, Georg August Universität, Göttingen, Germany; 16) Department of Neurology, Medizinische Universität Wien, Vienna, Austria; 17) Institute of Epidemiology II, Helmholtz Zentrum München, Munich, Germany; 18) Institute of Genetic Epidemiology, Helmholtz Zentrum München, Munich, Germany; 19) Max-Planck Institut für Psychiatrie München, Munich, Germany.

Background: Restless legs syndrome (RLS) is a common and genetically complex neurologic disorder characterized by nightly dysesthesias in the legs at rest leading to sleep disturbances. Although GWAS have identified genetic risk factors, these explain only about 6.8% of the total heritability. A portion of the remaining heritability could potentially lie in low-frequency and rare variants (MAF<5%) of strong effect. To date, no such variant is known for RLS. Methods: To assess the role of variants with MAF<5% at the known GWAS loci, we screened the coding regions of MEIS1, PTPRD, BTBD9, MAP2K5, SKOR1, TOX3 and BC034767 in 188 RLS cases and 188 controls by high-resolution melting curve analysis and Sanger sequencing. All identified variants with MAF<5% were genotyped in a sample of 3265 cases and 2944 controls. In a second approach, the coding and untranslated (UTR) regions of MEIS1 were screened in 3760 RLS cases and 3572 controls by the same method. Identified variants in MEIS1 were analysed for functional effects by mRNA in vivo complementation assay in zebrafish embryos using optic tectum area as phenotypic read-out. Results: Across all seven genes, we identified 50 non-synonymous (non-syn), synonymous (syn) and nearsplice variants (MAF<5%) present in 78 patients and 46 controls (p=0.004, McNemar test). When genotyped in 3265 cases and 2944 controls, non-syn variants with MAF<1% (p=3.7×10E-12) and MAF<0.1% (p=9.2×10E-5) across all genes and at solely the MEIS1 locus (p=0.002, all chi-squared test) were more common in cases than controls. In MEIS1, a total of 89 variants with MAF<5% were identified. Overall, variants in the 5ÙTR (p=6.2×10E-4) and non-syn variants (p=0.005, both chi-squared test) occurred more frequently in cases than controls. Three single rare MEIS1 variants were significantly associated with RLS prior to correction for multiple testing. In the preliminary assessment of all 17 non-syn MEIS1 variants in zebrafish, null and hypomorphic loss-of-function alleles appear to be more common in the cases than the controls and implicate a loss-of-function as the underlying mechanism. Conclusions: Coding variants with MAF<5% at RLS-GWAS loci are more common in RLS cases than in controls. At the MEIS1 locus, a spectrum of common and rare, non-coding and coding variants appears to contribute to disease development, supporting the concept of allelic series in complex diseases. Several rare variants emerge as candidates for causal RLS variants.

136

Mutations in PNPLA6 Cause a Range of Neurodegenerative Phenotypes. *M.A. Gonzalez[1], M. Synofzik[2], M. Coutelier[3], C. Marques Lourenço[4], T. Haack[5], H. Prokisch[5], R. Schule[2], W. Marques Junior[4], L. Schols[2], G. Stevanin[3], S. Zuchner[1].* 1) Human Genetics and Genomics, Unviersity of Miami Miller School of Medicine, Miami, FL; 2) Hertie Institute for Clinical Brain Research and Center for Neurology, Department of Neurodegenerative Disease, University of Tübingen, Germany; 3) Université Pierre-et-Marie-Curie, Universite Paris VI, Centre de Recherche de l'Institut du Cerveau et de la Moelle epiniere, EPHE, and Centre national de la recherche scientifique, Unité mixte de recherche 7225, Groupe Hospitalier Pitie-Salpêtriere; 4) Department of Neuroscience and Behaviour Sciences, School of Medicine of Ribeirão Preto, University of Sao Polo, Brazil; 5) Institute of Human Genetics, Technische Universität München, Germany.

The large amount of exome sequencing data across related and unrelated diseases has allowed supporting known, and defining new, spectra of phenotypes. Ataxia is a common symptom of a number of neurodegenerative diseases, including spinal ataxias, some forms of complicated hereditary spastic paraplegias (cHSP), and rare syndromes. With ataxia as a defining symptom we have searched across 936 exomes for common recessive genes using the GEnomes Management Application (GEM.app). We have identified PNPLA6 variants as an important and repeated cause in multiple (now deemed related) phenotypes. Boucher-Neuhäuser syndrome (BNS) is a rare autosomal recessive disorder characterized by spinocerebellar ataxia, chorioretinal dystrophy, and a failure of the pituitary to stimulate gonadal development during puberty. Three BNS families had causative PNPLA6 mutations. A Holmes syndrome family carried compound heterozygous mutations in PNPLA6. Mutations in PNPLA6 were previously described in two cHSP families. We identified PNPLA6 mutations in an additional cHSP family and also in a spastic ataxia family. In total, we report 6 families with homozygous or compound heterozygous mutations in PNPLA6. The majority of mutations clustered on the C-terminal catalytic domain, which has been shown to have esterase activity and phospholipase activity for lysophosphatidylcholine (LPC) and phosphatidylcholine (PC). LPC is the result of partial hydryolysis and removal of one of the fatty acids groups of PC. Alteration of phospholipid metabolism has been shown to impact mitochondrial function, which is hypothesized to be critical for the functions of the long axons of the corticospinal tracts. This adds to recently identified genes (DDHD1, DDHD2, CYP2U1 and others), which have highlighted the importance of phospholipase enzyme activity in neurodegeneration. In conclusion, our findings demonstrate that PNPLA6 mutations cause a range of neurodegenerative diseases creating a hitherto not recognized biological connection between several distinct clinical phenotypes.

137

Targeted resequencing of 101 known and candidate epilepsy genes in 600 patients with severe epilepsies identifies recurrently mutated genes. *G.L. Carvill[1], S.B. Heavin[2], J.M. McMahon[2], B.J. O'Roak[3], S.F. Berkovic[2], J. Shendure[3], I.E. Scheffer[2], H.C. Mefford[1]*. 1) Pediatrics, University of Washington, Seattle, WA; 2) Epilepsy Research Center and Department of Medicine, University of Melbourne, Austin Health, Australia; 3) Department of Genome Sciences, University of Washington, Seattle, WA 98195.

Epilepsy is one of the most common neurological disorders, with a lifetime incidence of 3%. The epileptic encephalopathies (EE) are the most severe of all the epilepsies, characterized by multiple refractory seizure types, cognitive arrest or regression, and a poor developmental prognosis. While *de novo* mutations in several genes cause EE, the genetic etiology of the vast majority remain unknown. Using molecular inversion probes for targeted capture followed by multiplexed next generation sequencing, we recently sequenced 19 known EE genes and 46 candidate genes in 500 EE patients (Carvill *et al.*, 2013). Candidate genes were selected from epilepsy-associated CNVs (n=33) and mutations that cause related neurodevelopmental disorders (NDDs) (n=13). Using this approach we made a genetic diagnosis in 10% of our cohort. Importantly, we describe six new EE genes, including *CHD2* and *SYNGAP1* that each account for ~1% of cases. We have since increased our epilepsy cohort to include well over 600 patients, including additional probands with EE (n=66) and patients with intellectual disability with genetic generalized epilepsy (n=52). We have expanded our target genes to include six recently reported EE genes and 30 new candidate genes for gene discovery, including genes identified by whole exome sequencing (n=8), CNV candidates (n=3), NDD-associated genes (n=7) and members of novel gene families important in EE (n=12). Preliminary results from our expanded study demonstrate emerging trends in genetic heterogeneity, with some genes causing discrete phenotypes. For example, mutations in *GRIN2A* cause epilepsy aphasia syndromes, and *GABRA1* mutations cause Dravet syndrome. By contrast, *de novo* mutations in other genes, including *CHD2*, *SYNGAP1*, *SCN8A* and *SCN2A* are implicated in a wider range of EE and NDD phenotypes. Of note, 4/6 new genes we describe are not directly involved in the regulation of neurotransmission at the synapse. *CHD2*, *MBD5* and *MEF2C* genes are involved in regulation of chromatin states that control gene expression, while *HNRNPU* is involved in splicing. These novel biological functions provide new avenues of research for understanding disease mechanisms and development of targeted therapies. In conclusion, we present a cost-effective, efficient method of screening multiple EE genes in large cohorts that will transform molecular diagnosis and highlight novel biological pathways implicated in epileptogenesis that can be targeted in therapeutic approaches.

138

A mouse model of Kabuki syndrome demonstrates defective hippocampal neurogenesis rescued with treatment with AR-42, a histone deacetylase inhibitor. *HT. Bjornsson[1], JS. Benjamin[1], L. Zhang[1], EE. Gerber[1], Y. Chen[1], MC. Potter[2], HC. Dietz[1,3]*. 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 2) Brain Science Institute, Department of Neurology; 3) Howard Hughes Medical Institute, Baltimore, MD.

Kabuki syndrome (KS) is caused by heterozygous loss of function mutations in either of two genes with complementary functions, the mixed lineage leukemia 2 (*MLL2*) and the lysine-specific demethylase 6A (*KDM6A*) genes. *MLL2* is a histone H3K4 methyltransferase that adds trimethylation to H3K4 (an open chromatin mark) while *KDM6A* is a demethylase that removes trimethylation at H3K27 (a closed chromatin mark). A possible hypothesis for the observed gene dosage sensitivity in Kabuki syndrome would be that the pathogenic sequence is dependent upon a relative imbalance between open and closed chromatin states for critical target genes. If this is the case, it may be possible to restore this balance with drugs that promote open chromatin states, such as histone deacetylase inhibitors (HDACi). In an effort to test this hypothesis we have characterized a novel mouse model of KS with a deletion of the SET domain of *Mll2* ($Mll2^{+/\beta Geo}$). Using an antibody against the trimethylated form of H3K4 we have been able to demonstrate a global deficiency of H3K4 trimethylation in the dentate gyrus of $Mll2^{+/\beta Geo}$ animals compared to wild-type ($Mll2^{+/+}$) littermates ($P<0.005$, N=14). Additionally, $Mll2^{+/\beta Geo}$ mice have smaller dentate gyrus volume ($P<0.05$, N=9) in association with a deficiency of neurogenesis, as evidenced by a deficiency of doublecortin positive proliferating neurons ($P<0.05$, N=8). $Mll2^{+/\beta Geo}$ mice perform poorly in Morris water maze testing ($P<0.05$, N=52), suggesting hippocampal memory defects. There is also a deficiency of both H3K4 trimethylation and H4 histone acetylation activity in mouse embryonic fibroblasts from $Mll2^{+/\beta Geo}$ mice when compared to $Mll2^{+/+}$ ($P<0.05$), as assessed using optimized novel fluorescent reporter alleles that allow quantitative, sensitive, specific and dynamic monitoring of these functions using fluorescence-activated cell sorting (FACS) of living cells. These deficiencies in cell culture can be ameliorated upon treatment with HDACi ($P<0.05$). *In vivo*, the deficiency of H3K4 trimethylation in the dentate gyrus is improved upon treatment with the HDACi AR-42 ($P<0.05$) in association with a dose-dependent rescue of neurogenesis (with full normalization at 10 mg/kg/day) for both one month and five month old mice. Our work suggests that a reversible deficiency of adult neurogenesis underlies intellectual disability in Kabuki syndrome and provides both rationale and incentive to test novel therapeutic strategies for KS and related disorders.

139

Human iPSC-based models of neuronal ceroid lipofuscinosis capture progressive pre-storage pathology in multiple cellular compartments. *J.F. Staropoli[1,2,3], X. Lojewski[1,4], S. Biswas[1], L. Haliw[1], A. Simas[1], M.K. Selig[2], K.A. Goss[1], A. Petcherski[1], S.H. Coppel[1], U. Chandrachud[1], S. Sheridan[1], K.B. Sims[5], J.F. Gusella[1], D. Lucente[1], D. Sondhi[6], R.G. Crystal[6], S.J. Haggarty[1], A. Hermann[4], A. Storch[4], S.L. Cotman[1,5].* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Department of Pathology, Massachusetts General Hospital, Boston, MA; 3) Division of Genetics and Genomics, Biogen Idec, Inc., Cambridge, MA; 4) Division for Neurodegenerative Diseases, Department of Neurology, Dresden University of Technology, Germany; 5) Department of Neurology, Massachusetts General Hospital, Boston, MA; 6) Department of Genetic Medicine, Weill Cornell Medical College, New York, NY.

The neuronal ceroid lipofuscinoses (NCLs; also known as Batten disease) are a currently untreatable group of at least 13 lysosomal storage disorders that collectively comprise the most common Mendelian form of progressive neurodegeneration in childhood. The pathologic hallmark of nearly all forms of NCL is the lysosomal accumulation of subunit c of the mitochondrial ATP synthase complex in multiple cell types, including affected neurons. Little is known about the gene products underlying this presumed final common pathway, but it is clear that, as in other neurodegenerative disorders, effective intervention will require a thorough understanding of histopathologic and molecular events that well precede neuronal loss. Here we provide initial characterization of human induced pluripotent stem cell (iPSC) and iPSC-derived neuronal models of the two most common forms of NCL: classic late-infantile NCL, caused by mutations in *TPP1*, and juvenile NCL, caused by mutations in *CLN3*. These NCL genetic cell models displayed overlapping but distinct and progressive pre-storage abnormalities in multiple cellular compartments, including the endosomal-lysosomal system, particularly in multivesicular bodies, as well as in mitochondria, Golgi, and endoplasmic reticulum (ER). Differentiation of neuronal precursor cells to mature neurons recapitulated disease-specific storage material. Moreover, virally overexpressed non-mutated *CLN3* rescued endosomal abnormalities in CLN3 patient cells, while virally overexpressed non-mutated *TPP1* as well as treatment with the nonsense suppressor PTC124 rescued the TPP1 enzyme deficiency in TPP1-deficient neurons carrying the common nonsense mutation c.622C>T/p.Arg208X. These models represent powerful tools to assess markers and mechanisms of early NCL pathology and to screen for compounds that modify disease progression in a genotype-specific manner. JFS and XL contributed equally to this work. JFS appears first only because he is the presenting author.

140

Mutation in EZR inhibits the Ras/MAP pathway and causes autosomal recessive intellectual disability. *R. Abou Jamra[1], L.B. Riecken[2], H. Tawamie[1], K. Geissler[2], A. Schulz[2], R. Buchert[1], S. Uebe[1], M.M. Nöthen[3,4], J. Schumacher[3], A. Ismael[5], A. Ekici[1], H. Sticht[6], A. Reis[1], H. Morrison[2].* 1) Institute of Human Genetics, Friedrich-Alexander-Universitat Erlangen-Nurnberg, Erlangen, Bavaria, Germany; 2) Leibniz Institute for Age Research, Fritz Lipmann Institute, Jena, Thuringia, Germany; 3) Institute of Human Genetics, University of Bonn, Bonn, NRW, Germany; 4) Life and Brain Center, University of Bonn, Bonn, NRW, Germany; 5) Praxis of Pediatrics, Jesser El Sheghour, Idlib, Syria; 6) Institute of Biochemistry, Friedrich-Alexander-Universitat Erlangen-Nurnberg, Erlangen, Bavaria, Germany.

We examined a large consanguineous family with intellectual disability. The family has two sons with severe non-specific intellectual disability and early epilepsy. Brain CT scan showed enlargement and deformation of the ventricles, periventricular leukomalacia, cerebral atrophy, dysplasia of corpus callosum, and reduction in the white matter in both hemispheres. We undertook autozygosity mapping and identified three candidate loci on chromosomes 6, (8.1 Mb), 18 (2.7 Mb), and 22 (7.8 Mb). We then enriched the exome of the index patient with Agilent SureSelect Kit 50 Mb and sequenced it on SOLiD 5500XL. We identified two novel homozygous variants in EZR (p.A129T) and MAP3K4 (p.M577V). The mutated alanine in EZR is highly conserved. In silico analysis using three programs predicted a pathogenic effect of the identified variant. Comprehensive molecular modelling on protein level showed that the mutation in EZR buries the hydrophilic threonine in the hydrophobic core and thus destabilizes the protein structure, probably leading to a strong effect on the protein function. EZR encodes ezrin, a member of the ERM (ezrin, radixin and moesin) protein family, which shares the FERM (four point one ERM homology) domain. Ezrin is necessary for a number of cellular processes, such as cell adhesion, motility, morphogenesis and cell signaling. The identified mutation is located in the FERM domain, which has binding sites for many membrane and signaling molecules. Because we recently showed that ezrin is required for the activity control of the small GTPase Ras we measured the effect of the ezrin mutant specifically on Ras. We transfected NIH3T3 cells with the wild type or mutated EZR. NIH3T3 cells expressing mutant ezrin blocked growth factor induced Ras activity. As a consequence of the inhibition of Ras we observed a decrease in proliferation. These in vitro cellular assays show that this mutation has a drastic effect on the ezrin protein and expression of which leads to an abnormal cellular phenotype. We suppose that this effect exists also in neurons and causes the severe phenotype of the examined family. Further experiments are ongoing. Taken together, we were able to identify EZR as a novel gene causing severe autosomal recessive non-specific intellectual disability. The loss of function of the ezrin mutant with the observed defects in Ras signaling is in line with other phenotypes of neurodevelopmental disorders and defects in Ras/MAP pathway.

141

DE NOVO MUTATIONS IN THE GENOME ORGANIZER CTCF CAUSE INTELLECTUAL DISABILITY. *C. Zweier[1], A. Gregor[1], M. Oti[2], E.N. Kouwenhoven[2], J. Hoyer[1], H. Sticht[3], A.B. Ekici[1], S. Kjaergaard[4], A. Rauch[5], H.G. Stunnenberg[6], S. Uebe[1], G. Vasileiou[1], A. Reis[1], H. Zhou[2,7].* 1) Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany; 2) Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 3) Bioinformatics, Institute of Biochemistry, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany; 4) Department of Clinical Genetics, University Hospital of Copenhagen, Rigshospitalet, Copenhagen, Denmark; 5) Institute of Medical Genetics, University of Zurich, Schwerzenbach, Switzerland; 6) Department of Molecular Biology, Faculty of Science, Radboud University Nijmegen, Nijmegen, The Netherlands; 7) Department of Molecular Developmental Biology, Faculty of Science, Radboud University Nijmegen, Nijmegen, The Netherlands.

An increasing number of genes involved in chromatin structure and epigenetic regulation has been implicated in a variety of developmental disorders, often including intellectual disability. By trio exome sequencing and subsequent mutational screening we now identified two de novo frameshift mutations and one de novo missense mutation in the CTCF gene in individuals with intellectual disability, microcephaly and growth retardation. Furthermore, a patient with a larger deletion including CTCF was identified. CTCF (CCCTC-binding factor) is one of the most important chromatin organizers in vertebrates and is involved in various chromatin regulation processes such as higher order of chromatin organization, enhancer function, and maintenance of three-dimensional chromatin structure. This crucial role in gene regulation prompted us to perform whole transcriptome analyses in blood lymphocytes of three of the patients and eight healthy controls. We found a broad deregulation of genes with a significant overlap between the patients. Down-regulated genes were enriched for genes involved in signal transduction and cell-environment interaction, processes which have been implicated in developmental and cognitive disorders. Together with data from chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) of CTCF in lymphocytes and publicly available ChIA-Pet data of CTCF from the related K562 cell line, we found that CTCF is important for enhancer-driven gene activation and that haploinsufficiency of CTCF affects the genomic interaction of enhancers and their regulated gene promoters that drive developmental processes and cognition.

142

Daylight exposure may modify the effect of variants at *MTNR1B* and *CRY2* on glucose tolerance: the GLACIER Study. *F. Renström[1,2], R. Koivula[1], T. V. Varga[1], G. Hallmans[3], F.B. Hu[2], J.C. Florez[4], H. Mulder[5], P.W. Franks[1,2,6].* 1) Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Skåne University Hospital Malmö, Malmö, Sweden; 2) Department of Nutrition, Harvard School of Public Health, Boston, USA; 3) Public Health & Clinical Medicine, Section for Nutritional Research, Umeå University Hospital, Umeå, Sweden; 4) Massachusetts General Hospital, Harvard Medical School, Boston, USA; 5) Department of Clinical Sciences, Skåne University Hospital Malmö, Malmö, Sweden; 6) Department of Public Health & Clinical Medicine, Section for Medicine, Umeå University Hospital, Umeå, Sweden.

Background: *MTNR1B* and *CRY2* are key players in the regulation of the circadian rhythm, both of which control glucose homeostasis. We recently reported novel associations between genetic variants at these two loci and 2-hrs glucose levels following an oral glucose tolerance test (OGTT) (Renström *et al.*, Diabetes, 2011). This study was performed in a population cohort situated ~400km (250 mi) south of the Arctic Circle in Sweden, where annual daylight hours vary between 4.5-22 hrs. Here we aimed to test the hypothesis that variation in daylight exposure modifies the relationship of rs10830963 (*MTNR1B*) and rs11605924 (*CRY2*) variation with 2-hr glucose levels. **Methods**: The rs10830963 (*MTNR1B*) and rs11605924 (*CRY2*) loci were genotyped in GLACIER, a prospective cohort study of 16,499 adults from the county of Västerbotten in northern Sweden. Daylight exposure was dichotomized according to the vernal and autumnal equinox. Effect modification by daylight exposure was tested by fitting generalized linear models on 2-hr glucose levels including the product term (SNP × season) along with the marginal effects. All analyses were adjusted for age, sex, fasting time and fasting glucose concentrations.**Results**: The *CRY2* locus was positively associated with 2-hr glucose levels during the darker months (β = 0.07 mmol/l per effect allele, P = 0.002, n = 9,605), whereas no association was observed among participants examined during the lighter months (β = -0.02 mmol/l per effect allele, P = 0.38, n = 6,215; $P_{interaction}$ = 0.01). In participants who underwent the OGTT during the dark season, the *MTNR1B* locus was significantly inversely associated with 2-hr glucose levels (β = -0.10 mmol/l per effect allele, P < 0.0001, n = 9,517), whereas no association was observed among participants examined during the bright season (β = -0.03 mmol/l per effect allele, P = 0.35, n = 6,174; $P_{interaction}$ = 0.06). **Conclusions**: Our study provides novel epidemiological evidence of a biologically plausible interaction between the rs10830963 (*MTNR1B*) and rs11605924 (*CRY2*) with daylight exposure on glucose tolerance levels.

143

***NAT1* in an important genetic effect modifier of tobacco smoke exposure in multiple sclerosis susceptibility in 5,453 individuals.** *F.B.S. Briggs[1], B. Acuna[2], L. Shen[2], H. Quach[1], A. Bernstein[3], I. Kockum[4], A.K. Hedström[5], M.W. Gustavsen[6,7], P. Berg-Hansen[6], S.D. Bos[6,7], E. Gulowsen Celius[6], H.F. Harbo[6,7], L. Alfredsson[5], T. Olsson[4], C. Schaefer[2], L.F. Barcellos[1,2].* 1) Genetic Epidemiology and Genomics Laboratory, Division of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, CA; 2) Division of Research, Kaiser Permanente, Oakland, CA; 3) Palm Drive Hospital, Sebastopol, CA; 4) Center for Molecular Medicine, Department of Medicine, Karolinska Institute, Sweden; 5) Institute of Environmental Medicine, Karolinska Institute, Sweden; 6) Department of Neurology, Oslo University Hospital, Oslo, Norway; 7) University of Oslo, Norway.

Multiple sclerosis (MS) is a neuro-inflammatory autoimmune disease with genetic and environmental components. Tobacco smoke (TS) is one of a few environmental exposures known to increase MS risk. However, not all smokers develop MS and only some individuals with MS were smokers. We hypothesized that host genetics contributing to TS metabolism modifies risk conferred by TS. A multi-stage gene-environment (GxE) analysis assessed variation within *NAT1*, *NAT2*, *GSTM1*, *GSTP1* and *GSTT1*, and metabolic phenotypes (*GSTM1* and *GSTT1* null, *NAT1* and *NAT2* slow/fast acetylators) and TS. Data from three large White population-based studies were utilized: Northern California, Norway, and Sweden. The 1st stage analysis was comprised of 2576 subjects with densely imputed genetic data. History of TS conferred MS risk (OR=1.4, P<0.0005), adjusted for age, gender, education, and population ancestry. In the discovery analysis of 1588 subjects, 42 *NAT1* variants showed evidence for interaction with TS ($P_{corrected}$<0.05). 41 *NAT1* SNPs were studied in a replication data set of 988 subjects, and a significant GxE interaction replicated. In the combined 1st stage meta-analysis of 2576 subjects *NAT1* rs7388368 had an OR of interaction (ORI)=1.7 (P<0.0005). In stratified analyses, TS was associated with MS risk among rs7388368A carriers only; homozygote individuals (A/A) had highest risk (C/C: OR=1.1, P=0.3; A/C: OR=1.6, P=0.001; A/A: OR=5.2, P<0.0005). Similar GxE associations were observed when smoking status at age 20 was used (ORI=1.4, P<0.05); and the TS conferred the greatest risk among rs7388368A homozygote subjects (OR=3.4, P=0.006). We have identified and replicated a GxE interaction in MS, identifying *NAT1* as a strong genetic effect modifier of TS exposure on MS risk in 2576 subjects. Rs7388368 variation affects four regulatory motifs where three transcription factors (TFs) bind, and resides adjacent to an insulator. Also, SNPs in linkage disequilibrium with rs7388368 (r^2>0.65), have been correlated with decreased *NAT1* expression in lymphoblastoid cells, suggesting a role in transcription. Therefore, all subjects are being genotyped for *NAT1* SNPs within 8 TF binding sites on a custom TaqMan OpenArray platform. These TF sites are within/near two promoters that result in transcription of various *NAT1* isoforms. A 2nd stage meta-analysis of the independent 2877 subjects demonstrated similar associations for TS and MS risk (OR=1.5, $P=9\times10^{-8}$), and GxE analyses are currently underway.

144

Genome-wide joint meta-analysis for interaction between genetic variants and smoking on waist circumference. *A.E. Justice[1], T.W. Winkler[2], J. Ngwa[3], K.L. Young[1,4], D. Hadley[5], M. Graff[1], J.M. Vink[6], L. Xue[3], T.S. Ahluwalia[7], T. Lehtimäki[8], R.J. Strawbridge[9], M.C. Zillikens[10], M.F. Feitosa[11], N.L. Heard-Costa[12,13], J.H. Zhao[14], J. Luan[14], N. Direk[15], H. Tiemeier[15], H.J. Grabe[16], T.B. Harris[17], R.P.S. Middelberg[18], J.V. van Vliet-Ostaptchouk[19,20], I.M. Nolte[20], J. Kaprio[21], T.O. Kilpeläinen[7], I.B. Borecki[11], R.J.F. Loos[22], K.E. North[1], L.A. Cupples[3,12] on behalf of the GIANT Consortium.* 1) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 2) Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany; 3) Department of Biostatistics, School of Public Health, Boston University, Boston, MA; 4) Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC; 5) Pediatric Epidemiology Center, University of South Florida, Tampa, FL; 6) Department of Biological Psychology, Neuroscience Campus, VU University & VU medical center Amsterdam; 7) The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 8) Department of Clinical Chemistry, Fimlab Laboratories and School of Medicine, University of Tampere, Tampere, Finland; 9) Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet; 10) Department of Internal Medicine, Erasmus MC Rotterdam; 11) Department of Genetics, School of Medicine, Washington University; 12) Framingham Heart Study, National Heart, Lung, and Blood Institute, Framingham, MA; 13) Department of Neurology, Boston University School of Medicine, Boston, MA; 14) MRC Epidemiology Unit, University of Cambridge, Cambridge, UK; 15) Department of Epidemiology, Erasmus Medical Centre; 16) Department of Psychiatry, University Medicine of Greifswald, Germany; 17) Intramural Research Program, National Institute on Aging, Bethesda, MD; 18) Genetic Epidemiology Unit, Queensland Institute of Medical Research; 19) Department of Endocrinology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; 20) Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; 21) Institute for Molecular Medicine, University of Helsinki; 22) The Charles Bronfman Institute for Personalized Medicine, The Mincdich Child Health and Development Institute, The Icahn School of Medicine at Mount Sinai, New York, NY.

Cigarette smokers often display lower body weight than non-smokers, and both men and women gain weight after smoking cessation causing changes in central adiposity, as assessed with waist circumference (WC). Identifying genes that influence WC and whose effects are modified by smoking will help us understand the complex interplay between genetic susceptibility, smoking and central adiposity. While many loci have been associated with WC, little is known about whether current smoking status (SMK) influences these genetic associations. We aim to discover genetic loci that interact with smoking to influence WC adjusted for BMI (WC^a), and to increase the power to detect genetic main effects that may be hidden when the environmental influence of smoking is not considered. To address our aims we evaluated results from 42 studies with GWAS data available in the GIANT (Genetic Investigation of ANthropometric Traits) Consortium providing up to 113,587 individuals. We conducted inverse-variance weighted fixed-effects meta-analyses of the study specific results for four association models separately and combined across sex: 1) SNP main effect on WC^a, stratified by SMK; 2) SNP main effect on WC^a, adjusted for SMK; 3) SNP x SMK interaction effect on WC^a; and 4) we evaluated a joint meta-analysis of the SNP main effect and SNP x SMK interaction effect. A total of 63 loci reached Genome-Wide significance (GWS) ($p<5\times10^{-8}$) in one or more strata, with the greatest number of significant results coming from the joint effects model. Of the 63, seven have been previously associated with waist traits, 11 are near SNPs previously associated with other adiposity traits (e.g. BMI, visceral adiposity, extreme obesity, adiponectin), and one is near a SNP previously associated with smoking behavior. A novel association near PRNP reached GWS for SMK interaction (model 3) in women. PRNP is highly expressed in the nervous system and is important in the functioning of cell signaling, memory, and immune response. Other GWS SNPs lie near strong biological candidates important in early growth and development (e.g. DNMT3A, TBX15, FGFR4) and nervous system functioning (e.g. CABLES1, DOCK3). We have greatly increased the number of loci associated with central adiposity and highlight the influence of the nervous system and developmental processes on adiposity-related traits. These results underscore the importance of accounting for SMK when investigating the genetic basis of central adiposity.

145

Interaction between genome-wide variants and physical activity on body mass index: a meta-analysis of 109,924 individuals. *T.O. Kilpeläinen[1,2], R.A. Scott[2], A. Mahajan[3], L. Xue[4], F. Renström[5], M. Graff[6], D. Hadley[7], T. Workalemahu[8], M. den Hoed[2,9], T.W. Winkler[10], A.Y. Chu[11], N.L. Heard-Costa[12], T. Haller[13], T.S. Ahluwalia[1], J.V. van Vliet-Ostaptchouk[14], P.J. van der Most[14], M.L. Grove[15], L. Quaye[16], S. Snitker[17], E.J.C. de Geus[18], T. Lehtimäki[19], L. Qi[8], P.W. Franks[5,8], I.B. Borecki[20], K. Monda[6], D.I. Chasman[11], K.E. North[6], L.A. Cupples[4,12], R.J.F. Loos[2,21], The GIANT Consortium.* 1) The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, University of Copenhagen, Denmark; 2) MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, UK; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 4) Department of Biostatistics, Boston University School of Public Health, MA; 5) Department of Clinical Science, Genetic and Molecular Epidemiology Unit, Lund University, Malmö, Sweden; 6) Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, NC; 7) Pediatric Epidemiology Center, University of South Florida, Tampa, FL; 8) Department of Nutrition, Harvard School of Public Health, Boston, MA; 9) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Sweden; 10) Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, Germany; 11) Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA; 12) Framingham Heart Study, Framingham, MA; 13) Estonian Genome Center, University of Tartu, Estonia; 14) Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands; 15) School of Public Health, University of Texas Health Science Center at Houston, TX; 16) Department of Twin Research and Genetic Epidemiology, King's College London, UK; 17) Department of Medicine, Division of Endocrinology, Diabetes & Nutrition, University of Maryland School of Medicine, MD; 18) EMGO+ Institute for Health and Care Research, VU Medical Center & VU University, Amsterdam, The Netherlands; 19) Department of Clinical Chemistry, Fimlab Laboratories, Tampere University Hospital and University of Tampere, Finland; 20) Department of Genetics, Washington University School of Medicine, St. Louis, MO; 21) The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York, NY.

The global obesity epidemic underscores the importance of gaining a deeper understanding of the biology of weight regulation to develop better preventive strategies. Obesity has a strong genetic component, but lifestyle factors such as physical activity (PA), may attenuate the impact of genetic susceptibility. Recently, a candidate gene-based meta-analysis confirmed that the body mass index (BMI)-increasing effect of the *FTO* obesity locus is attenuated by PA. To identify novel loci that interact with PA on BMI, we performed a meta-analysis of 38 GWAS, including 109,924 individuals of European descent. We standardized PA across all participating studies by categorizing it into a dichotomous variable (physically inactive vs. active) in each study. Overall, 23% of individuals were categorized as inactive. To screen for interactions, we performed a joint test of the SNP main effect and the SNP x PA interaction term, adjusting for age and age^2 in men and women separately. In addition, we performed a test for the SNP x PA interaction term alone, with the same adjustments. We pooled the results from individual studies using fixed effects inverse variance meta-analysis. The joint P-value reached genome-wide significance ($P<5\times10^{-8}$) for 19 loci in all individuals, or in men or women separately. However, for each of these loci, the association was driven by the SNP's main effect on BMI, rather than by its interaction with PA. When we examined the SNP x PA interaction effects of the 19 loci, separately from main effects, the interaction reached nominal statistical significance for *FTO* ($P=6\times10^{-6}$), *GIPC2* (P=0.02), and *ADCY3* (in women; P=0.03). For these 3 loci, the BMI-increasing effect of the risk allele was smaller in the active group than in the inactive group. In a genome-wide screen for the SNP x PA interaction term alone, no locus reached $P<5\times10^{-8}$. However, the interaction for 4 novel loci (*SV2C*, *SLIT3*, *CGNL1*, *TEKT5*) reached $P<10^{-6}$. In this sample of 109,924 individuals, we confirm that the BMI-increasing effect of *FTO* is attenuated by PA, but find little evidence for other loci interacting with PA. To follow-up our initial findings and to extend our total sample size, we are currently collecting data from approximately 25 additional studies with either genome-wide or Metabochip data. Our study may reveal novel genetic loci whose effects on obesity-risk are modified by PA, and may give important biological insights into the regulation of body weight.

146

Genome-wide analyses highlights gene interaction with processed meat and vegetable intake for colorectal cancer risk. *J. Figueiredo[1], L. Hsu[2,15], E. White[2,15], A. Chan[3,16], B. Zanke[4], J. Potter[2,15], G. Casey[1], C. Hutter[5], H. Brenner[6], B. Caan[7], J. Chang-Claude[8], S. Gallinger[9], R. Hayes[10], T. Hudson[11], L. Le Marchand[12], P. Newcomb[2,15], R. Schoen[13], M. Slattery[14], U. Peters[2,15].* 1) Preventive Medicine, USC, Los Angeles, CA; 2) Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 4) Clinical Epidemiology Program, Ottawa Hospital Research Institute, Toronto, ON; 5) Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Washington DC; 6) Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany; 7) Division of Research, Kaiser Permanente Medical Care Program, Oakland, CA; 8) Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany; 9) Department of Surgery, Mount Sinai Hospital, Toronto, ON; 10) Division of Epidemiology, Department of Environmental Medicine, New York University School of Medicine, New York, NY; 11) Department of Medical Biophysics, University of Toronto, Toronto, ON; 12) Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI; 13) Department of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh PA; 14) Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City UT; 15) Department of Epidemiology, University of Washington School of Public Health, Seattle WA; 16) Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School.

Dietary factors including low consumption of red and processed meat and high intake of fruits, vegetables and fiber are associated with a lower risk of colorectal cancer; however, there is limited information as to whether these dietary factors interact with genetic variants to modify risk of colorectal cancer. To examine this question at a genome-wide level we tested interactions between these dietary factors and about 2.7 million genetic variants on colorectal cancer risk among up to 9,287 cases and 9,117 controls using individual-level data from 10 observational studies. We used logistic regression to detect potential multiplicative gene-diet (GxD) interactions for colorectal cancer as well as our recently developed Cocktail method that involves a screening step based on marginal associations and G-D correlations and a testing step for GxD interactions while correcting for multiple testing using weighted hypothesis testing. Evaluating risks per quartile increment in intake, red and processed meat were associated with statistically significant increased risks of colorectal cancer (OR=1.15, p=1.6E-18 and OR=1.11, p=4.2E-09, respectively) and vegetable, fruit and fiber intake with lower risks (OR=0.93, p=8.2E-05; OR=0.93, p=1.9E-05 and OR=0.91, p=5.6E-05, respectively). From the case-control analysis, we detected a statistically significant interaction between rs4143094 and processed meat consumption (p=8.7E-09). rs4143094 is located 1.42 kb upstream of GATA3 and DNase hypersensitivity marks and histone methylation patterns are consistent with open chromatin and poised/active promoters. Based on our cocktail method, we provided additional support for a statistically significant interaction between variants at the known locus 8q23.3/EIF3H, UTP23 and vegetable intake, which reached genome-wide significance by weighted hypothesis testing (for rs2450114: p-screen=1.4E-08, p-interaction=4.1E-03, p-alpha=5.0E-03). These genetic loci may have interesting biological significance given their location in the genome and further functional analyses are required. Our results identify a novel gene-diet interaction for colorectal cancer, highlighting that diet may modify the effect of genetic variants on diseases risk, which may have important implications for prevention.

147

Genome wide gene-environment interaction study identifies a CYP24A1-related variant as a modifier of colorectal cancer risk associated with menopausal estrogen plus progesterone therapy. *X. Garcia-Albeniz[1], A. Rudolph[2], C.M. Hutter[3], Y. Lin[3], E. White[3], H. Brenner[4], G. Casey[5], S. Gallinger[6], L. Hsu[3], T.J. Hudson[7], L. Le Marchand[8], J. Potter[3], M. Slattery[9], B. Zanke[10], P.A. Newcomb[3], A.T. Chan[11], U. Peters[3], J. Chang-Claude[2].* 1) Epidemiology, Harvard School of Public Health, Boston, MA; 2) Epidemiology, German Cancer Research Center, Heidelberg, Germany; 3) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA; 4) Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany; 5) Preventive Medicine, University of Southern California, Keck School of Medicine, Los Angeles, CA; 6) Surgery, Mount Sinai Hospital, New York, NY; 7) Medical Biophysics, University of Toronto, Toronto, Canada; 8) Epidemiology, University of Hawaii Cancer Center, Honolulu, HI; 9) Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, UT; 10) Clinical Epidemiology. Ottawa Hospital Research Institute, Ottawa, ON; 11) Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA.

Postmenopausal hormone therapy use (PMH) has been consistently associated with a decreased risk of colorectal cancer (CRC). The underlying pathways involved in prevention of CRC by PMH are largely unknown. Our aim was to use a genome-wide gene-environment analysis to identify variants modifying the effect of PMH on the risk of CRC. We included 9 studies from the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) and the Colon Cancer Family Registry. Genome-wide genotype array data from each study was imputed to the HapMap II population (CEU). PMH use was evaluated as use at reference time of “any PMH”, “estrogen only” (E-only), and “estrogen+progesterone” (E+P). To test for multiplicative (GxPMH) interactions, we used a case-control logistic regression analysis as well as a two-stage “Cocktail” method that includes a screening step to prioritize SNPs followed by a testing step for GxPMH interactions using weighted multiple hypothesis testing. In secondary analyses, we applied another two stage method that jointly tests marginal association and GxE interaction as well as an Empirical Bayes method. Final results for each method were obtained by combining study level results with fixed-effects meta-analyses, comprising up to 10,835 postmenopausal women (type of compound known on 9,674). The genome-wide interaction analysis using case-control logistic regression did not yield any genome-wide significant ($p<5.0\times10\text{-}8$) interaction with PMH use. Based on the Cocktail method, we observed a significant interaction of the variant rs964293 with E+P (OR=0.65, $p=2.8\times10\text{-}5$; threshold for significance=$3.1\times10\text{-}4$). This variant also showed significant interaction with E+P use on CRC risk using the Empirical Bayes method (OR=0.62, 95% CI 0.53-0.73, $p=9.1\times10\text{-}9$). We did not detect any genome-wide significant interaction when considering E-only or any PMH use. The variant rs964293 is located in an intergenic region 28 kb upstream of CYP24A1 on chromosome 20q13.2. Among women not using PMH, rs964293 was marginally associated with an increase in risk for CRC (per allele OR=1.08, 95% CI 1.00-1.16, p=0.05) whereas among women taking E+P, rs964293 was associated with a lower risk of CRC (per allele OR=0.72, 95% CI 0.59-0.88, p=0.0011). We identified a CYP24A1-related variant as effect modifier of CRC risk associated with use of E+P, using a genome-wide approach. This finding offers insight in the role of PMH and its downstream pathways in the etiopathogenesis of CRC.

148

Replication of gene-gene interaction models associated with cataracts in the eMERGE Network. *M.A. Hall[1], S.S. Verma[1], E.R. Holzinger[1], R. Berg[2], J. Connolly[3], D.C. Crawford[4], D.R. Crosslin[5], M. de Andrade[6], K.F. Doheny[7], J.L. Haines[4], J.B. Harley[8], G.P. Jarvik[5], T. Kitchner[2], H. Kuivaniemi[9], E.B. Larson[5,10], G. Tromp[9], S.A. Pendergrass[1], C.A. McCarty[11], M.D. Ritchie[1].* 1) Center for Systems Genomics, The Pennsylvania State University, University Park, PA; 2) Marshfield Clinic, Marshfield, WI; 3) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 4) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 5) Department of Genome Sciences, University of Washington, Seattle, WA; 6) Mayo Clinic, Rochester, MN; 7) Center for Inherited Disease Research, IGM, Johns Hopkins University SOM, Baltimore, MD; 8) Cincinnati Children's Hospital, University of Cincinnati, Department of Pediatrics, Cincinnati, OH; 9) Geisinger Health System, Danville, PA; 10) Group Health Research Institute, Seattle, WA; 11) Essentia Rural Health, Duluth, MN.

Bioinformatics approaches to examine epistasis provide the means to discover the interactions between multiple genes and pathways that are likely the basis of complex disease. Despite its importance, extensive computational demands and adjusting for multiple testing make uncovering these interactions a challenge when explored with an exhaustive combinatorial search. Here, we address this issue using Biofilter 2.0 to identify putative SNP-SNP models for cataract susceptibility, reducing the number of models for analysis. With Biofilter 2.0, we created biologically relevant SNP-SNP models from genes with published associations, including genes belonging to the same pathway or having known biological interactions. Using PLATO software, we evaluated these models using logistic regression, adjusting for sex and principal components in 3,907 samples (1,354 controls, 2,553 cases) of European (3872), African (1), Asian (14), and other (13) descent from the Marshfield Clinic Personalized Medicine Research Project, part of the Electronic Medical Records & Genomics (eMERGE) Network. All highly significant models from the Marshfield Clinic (likelihood ratio test (LRT) $p < 0.0001$) were then tested in a replication dataset of 3,483 individuals (537 controls, 2,946 cases) of European (3251), African (113), Asian (66), and other (53) descent, using independent samples from additional sites in the eMERGE Network: Mayo Clinic, Group Health Cooperative, and Vanderbilt University Medical Center. Over 100 SNP-SNP models were found in the replicating sample at LRT $p < 0.01$, and 8 models replicated with high significance (LRT $p < 10\text{-}4$). The most significant replicating SNP-SNP models and their nearest genes included rs7749147 (FYN) - rs11017910 (DOCK1), rs9790292 (TGFBR2) - rs8110090 (TGFB1), rs10176426 (UGT1A10) - rs17863787 (UGT1A6), and rs11723463 (UGT2B4) - rs1112310 (UGT1A10). Notably, the genes UGT1A10 and UGT1A6, members of the UDP glucuronosyltransferase 1 family, and UGT2B4, of the UDP glucuronosyltransferase 2 family are involved in the porphyrin and chlorophyll metabolism pathway. This pathway has demonstrated association with cataracts, and therefore, bears further inquiry. These findings indicate the role of epistasis in susceptibility to cataracts and demonstrate the utility of Biofilter 2.0 as a biology-driven method, which can be applied to any GWAS dataset for investigation of the complex genetic architecture of common diseases.

149

Evidence from multiple genome-wide association studies of a hub of gene-gene interactions affecting human HDL cholesterol levels. *L. Ma[1,2], C. Ballantyne[3], A. Brautbar[4], A. Keinan[1].* 1) Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 2) Department of Animal and Avian Sciences, University of Maryland, College Park, MD; 3) Section of Cardiovascular Research, Department of Medicine, Baylor College of Medicine, Houston, TX; 4) Department of Medical Genetics, Marshfield Clinic, Marshfield, WI.

Epistasis, an important genetic component underlying the basis of complex traits in many species, has been suggested to underlie some of missing heritability in genome-wide association studies. In this study, we first sought to identify gene-gene interactions affecting HDL cholesterol (HDL-C) levels between a focused set of SNPs in a sample of 2,091 European American (EA) individuals in a candidate gene study. Two additional EA samples from the Atherosclerosis Risk in Communities study (ARIC; n = 9,713) and from the Multi-Ethnic Study of Atherosclerosis (MESA; n = 2,685) were considered for replication. We identified a gene-gene interaction between rs1532085 and rs12980554 ($P = 7.1\times10^{-7}$) in their effect on HDL-C levels, which is significant after Bonferroni correction ($P_c = 0.017$) for the number of SNP pairs tested. It successfully replicated in the ARIC study ($P = 7.0\times10^{-4}$; $P_c = 0.02$). Rs1532085, an eQTL of *LIPC*, has been involved in another, well-replicated interaction affecting HDL-C that we recently published. To further investigate the role of this SNP in gene-gene interactions, we additionally tested for gene-gene interaction between this SNP and any other SNP in the ARIC study and found it to be involved in a few additional suggestive interactions (FDR = 0.25), one of which significantly replicated in MESA ($P = 0.03$). *LIPC* is known to play a key role in the lipid and lipoprotein metabolism pathway. *LIPC*, and rs1532085 in particular, has also been previously associated by itself with the level of HDL-C, as well as with other lipid traits and metabolic syndrome. Collectively, we discovered several novel gene-gene interactions on HDL-C levels, all involving an eQTL of *LIPC*, thus suggesting a hub role of this SNP and *LIPC* in general in the interaction network regulating HDL-C.

150

Epistasis Analysis for Quantitative Trait with Next-generation Sequencing Data. *F. Zhang, E. Boerwinkle, M. Xiong.* University of Texas School of Public Health.

The critical barrier in interaction analysis for rare variants is that most traditional statistical methods for testing interaction were originally designed for testing the interaction between common variants and are difficult to be applied to rare variants for their prohibitive computational time and low power. The great challenges for successful detection of interactions with next-generation sequencing (NGS) data are (1) lack of methods for interaction analysis with rare variants, (2) suffering from severe multiple testing, and (3) heavy computations. To meet these challenges, we shift the paradigm of interaction analysis between two loci to interaction analysis between two sets of loci or genomic regions and collectively test interaction between all possible pairs of SNPs within two genome regions. In other words, we take a genome region as a basic unit of interaction analysis and use high dimensional data reduction and functional data analysis techniques to develop a novel functional regression model to collectively test interaction between all possible pairs of SNPs within two genome regions. By intensive simulations, we demonstrate that the functional regression models for interaction analysis of the quantitative trait has the correct type 1 error rates and much higher power to detect interaction than the current pair-wise interaction analysis. The proposed method was applied to exome sequence data from the NHLBI's Exome Sequencing Project (ESP) and CHARGE-S study. We discovered 27 pairs of genes showing significant interactions after applying the Bonferroni correction (P-values <4.58E-10) in ESP and 11 of them were replicated in CHARGE-S study.

151

Targeted capture and sequencing identifies causative alleles in simplex and multiplex consanguineous Palestinian families with orofacial clefts. *H. Shahin[1], U. Sharaha[1], M.K. Lee[2], A. Watts[2], M-C. King[2], J. van Aalst[3], T. Walsh[2]*. 1) Dept Life Sci, Bethlehem Univ, Bethlehem, Palestinian Territory; 2) Division of Medical Genetics, University of Washington, Seattle, WA; 3) Division of Plastic and Reconstructive Surgery, University of North Carolina, Chapel Hill, NC.

Orofacial clefts are among the most common of all human congenital birth defects. They are divided into two categories on the basis of embryological findings: cleft lip with or without cleft palate (CL/P) and cleft palate only (CP). CL/P affects between 1/2500 and 1/500 live births with considerable geographic variability in rates. Overall, approximately 70% of CL/P and 50% of CP are non-syndromic. Genes responsible for orofacial clefts were first identified by linkage and positional cloning studies of families and more recently by exome sequencing. In order to more efficiently identify causative alleles in genes known to cause orofacial clefts, we developed a targeted capture panel to sequence the coding exons and flanking splice sites of 137 genes. The capture size of 494kb enabled deep multiplexed sequencing while maintaining high coverage and detection of single base substitutions, indels and exonic copy number variants. Using this capture panel, we evaluated 44 Palestinian patients with orofacial clefts undergoing corrective surgeries, but with no prior genetic testing. Of the 44 patients, 20 were simplex cases with no family history of clefting, and 24 were probands of multiplex families. In four of the 20 simplex cases, we identified heterozygous mutations, either truncations or functionally validated damaging missense mutations, in *GLI2*, *MSX1*, *SATB2*, and *TP63*. In four of the 24 probands from multiplex families, we identified a homozygous missense mutation in *COL2A1* (Gly855Ser). Genotyping all 41 members of these four extended kindreds revealed perfect co-segregation of COL2A1 p.G855S with clefting. Among homozygotes for COL2A1 p.G855S, there was considerable variability of clinical presentation in addition to clefting, including partial hearing loss, ocular problems, joints abnormalities, and dwarfism. Our approach enables comprehensive testing of all known clefting genes in a single assay and can be readily applied to large cohorts of patients with or without additional syndromic features. Sequencing in this way is likely to increase our knowledge of the type of clefting and additional clinical features associated with mutations in each gene.

152

Dominant mutations in *GRHL3* cause Van der Woude syndrome and disrupt oral periderm development. *M. Peyrard[1], E.J. Leslie[2], Y.A. Kousa[3], T.L. Smith[4], M. Dunnwald[2], M. Magnusson[5], B.A. Lentz[2], P. Unneberg[6], I. Fransson[1], H.K. Koillinen[7], J. Rautio[8], M. Pegelow[9], A. Karsten[9], L. Basel-Vanagaite[10,11], W. Gordon[12], B. Andersen[12], T. Svensson[5], J.C. Murray[2], R.A. Cornell[4], J. Kere[1,5,13], BC. Schutte[14]*. 1) Department of Biosciences and Nutrition, Karolinska Institutet, and Center for Biotechnology, 14183 Huddinge, Sweden; 2) Department of Pediatrics and Interdisciplinary Program in Genetics, University of Iowa, Iowa City 52242, Iowa, USA; 3) Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing 48824, Michigan, USA; 4) Department of Anatomy and Cell Biology, University of Iowa, Iowa City 52242, Iowa, USA; 5) Department of Biosciences and Nutrition, Science for Life Laboratory, Karolinska Institutet, 17121 Solna, Sweden; 6) Department of Biochemistry and Biophysics Science for Life Laboratory, Stockholm University, 17121 Solna, Sweden; 7) Department of Clinical Genetics, Helsinki University Hospital, 00029 Helsinki, Finland; 8) Cleft Palate and Craniofacial Center, Department of Plastic Surgery, Helsinki University Hospital, 00029 Helsinki, Finland; 9) Department of Orthodontics, Stockholm Craniofacial Team, Institute of Odontology, Karolinska Institutet, 17177 Stockholm, Sweden; 10) Pediatric Genetics, Schneider Children's Medical Center of Israel and Raphael Recanati Genetic Institute, Rabin Medical Center, Petah Tikva 49100, Israel; 11) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel; 12) Department of Biological Chemistry, University of California Irvine, Irvine 92697, California, USA; 13) Research Programs Unit, University of Helsinki, and Folkhälsan Institute of Genetics, 00014 Helsinki, Finland; 14) Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing 48824, Michigan, USA.

Van der Woude syndrome (VWS) is an autosomal dominant form of cleft lip and palate (CL/P). About 70% of VWS cases are caused by mutations in the gene encoding for the transcription factor Interferon Regulatory Factor 6 (*IRF6*). Because IRF6 is essential for development of the oral periderm in mouse and zebrafish, we hypothesized that a subset of the remaining 30% of VWS cases are caused by mutations in other genes regulating periderm development. Supporting locus heterogeneity in VWS, one large pedigree where the proband has lip pits, the hallmark of VWS, showed linkage to a locus on 1p33-p36 (*VWS2*) rather than to *IRF6* at 1q32-q41. From this pedigree, we selected 8 affected individuals, including the proband, and 3 healthy individuals, for whole-exome sequencing of the ~700 genes in this linkage interval. We identified a likely causative mutation, F324Lfs, in the Grainy head-like 3 (*GRHL3*) gene. GRHL3 is a transcription factor with a deeply conserved role in formation of epithelial barriers. We also found likely disease-causing mutations in *GRHL3* in seven additional families with VWS, of variable ethnicity and where no causative *IRF6* mutations had been detected. While we observed no consistently unique phenotypes in these families, individuals with a *GRHL3* mutation are more likely to have cleft palate and less likely to have cleft lip or lip pits than individuals with an *IRF6* mutation. In a zebrafish-based assay, we found that all five human *GRHL3* variants tested, F324Lfs included, disrupted periderm development during gastrulation. These results suggest that VWS is caused by dominant-negative alleles of *GRHL3*, in contrast to haploinsufficiency of *IRF6* seen in most cases. Using murine models, we observed that *Grhl3*$^{-/-}$ embryos, similar to *Irf6*$^{-/-}$ embryos, have oral epithelial adhesions, a phenotype seen occasionally in human VWS, and exhibit abnormal development of the oral periderm. Finally, the oral phenotype of double heterozygous (*Irf6*$^{+/-}$; *Grhl3*$^{+/-}$) murine embryos is consistent with an additive rather than a synergistic interaction between the two genes, suggesting that Irf6 and Grhl3 function in separate but converging pathways in periderm development. Together our data demonstrate that mutations in two genes, *IRF6* and *GRHL3*, lead to nearly identical phenotypes of CL/P, and support the hypothesis that failure of oral periderm development contributes to VWS. This work will increase the specificity and sensitivity of clinical diagnostics for VWS.

153

A homeotic maxillary to mandibular transformation in humans resulting from loss of selective ligand affinity of the endothelin receptor type A. *C. Gordon[1], M. Oufadem[1], Y. Kurihara[2], A. Picard[3], S. Breton[4], S. Pierrot[3], M.-A. Delrue[5], M. Biosse-Duplan[1], M. Guion-Almeida[6], P. Moura[6], D. Garib[7], D. Weaver[8], A. Munnich[1,9], P. Corvol[10], H. Kurihara[2], R. Zechi-Ceide[6], S. Lyonnet[1,9], J. Amiel[1,9].* 1) Université Paris Descartes, INSERM U-781, Paris, France; 2) Department of Physiological Chemistry and Metabolism, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 3) Service d'ORL Pediatrique, Hôpital Necker-Enfants Malades, Paris, France; 4) Service d'Imagerie Pédiatrique, Hôpital Necker-Enfants Malades, Paris, France; 5) Unité de génétique médicale, CHU Bordeaux, Bordeaux, France; 6) Department of Clinical Genetics, Hospital for Rehabilitation of Craniofacial Anomalies, University of São Paulo (HRCA-USP), Bauru, SP, Brazil; 7) Department of Orthodontics, Bauru Dental School, University of São Paulo, Bauru, São Paulo, Brazil; 8) Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana; 9) Département de Génétique, Hôpital Necker-Enfants Malades, Paris, France; 10) INSERM U833, Collège de France, Paris, France.

A homeotic transformation is a fascinating genetic error of development first described in Drosophila. In humans, Auriculocondylar syndrome (OMIM 602483 and 614669) is thought to result from the homeotic transformation of the mandible into a maxilla. We present two unrelated and sporadic cases with a severe and syndromic form of mandibulofacial dysostosis (MFD) featuring malformation of the eyelids and alopecia. Craniofacial 3D CT scans of both patients showed a striking symmetry between the abnormal mandible and maxilla, suggestive of a mirror image version of mandibular structures in the upper jaw. The endothelin core system (i.e. ligands plus receptors) is a vertebrate-specific innovation implicated in the development of several neural crest cell (NCC)-derived tissues, and in mammals is comprised of three ligands and two receptors belonging to the G protein-coupled seven transmembrane domain receptor family. Endothelin receptor type A (Ednra) displays selective affinity for endothelin 1 (Edn1) relative to Edn2 and Edn3, and Edn1/Ednra signalling plays a critical role in specification of mandibular identity in post-migratory NCCs. Edn1 expression is restricted to the mandibular prominence and Ednra is expressed throughout the first pharyngeal arch. The phenotype of the two patients we report was considered similar to a mutant mouse in which the Edn1 cDNA is knocked-in to the Ednra locus, leading to their coexpression and resulting in a maxillary to mandibular transformation. EDNRA was thus considered as a candidate disease-causing gene for the syndromic MFD of the two patients. Sanger sequencing identified the same de novo, heterozygous missense mutation in EDNRA in both patients. In in vitro assays, the mutation increases affinity for EDN3 approximately 100 fold, while not affecting the affinity for EDN1. Expression analysis of Edn3 during mouse craniofacial development revealed strong expression in the epithelium of the caudal surface of the maxillary prominence, but not in the mandibular prominence, at E11.5. Whether the increase in affinity for EDN3 leads to increased levels of EDNRA signalling is unclear at present, but is consistent with a model in which the normal expression of EDN3 in the maxillary prominence would lead to a maxillary-to-mandibular transformation via overactivation of mutant EDNRA. These findings confirm the importance of selective EDN1-EDNRA affinity in human jaw development and gnathostome evolution.

154

Putative gain of function mutations in FAM111A result in Kenny-Caffey syndrome and Osteocraniostenosis: identification of an upstream regulator of the PTH axis. *S. Unger[1,2], M.W. Gorna[3], A. Le Bechec[4], S. Do Vale-Pereira[2], F. Bedeschi[5], S. Geiberger[6], G. Grigelioniene[7], E. Horemuzova[8], F. Lalatta[5], E. Lausch[9], C. Magnani[10], S. Nampoothiri[11], G. Nishimura[12], D. Petrella[13], F. Rojas-Ringeling[14], A. Utsunomiya[15], B. Zabel[9], S. Pradervand[16], K. Harshman[16], B. Campos-Xavier[2], L. Bonafé[2], G. Superti-Furga[3], B. Stevenson[4], A. Superti-Furga[2].* 1) Service de Génétique Médicale, CHUV / University of Lausanne, Lausanne, Switzerland; 2) Dept of Pediatrics, Lausanne University Hospital, University of Lausanne, Lausanne, Switzerland; 3) CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria; 4) Vital-IT - High Performance Computing Center, Swiss Institute of Bioinformatics, Univ. of Lausanne, Lausanne, Switzerland; 5) Clinical Genetic Unit, Fondazione Ospedale Maggiore Policlinico, Mangiagalli e Regina Elena, Milano, Italy; 6) Department of Pediatric Radiology, Astrid Lindgrens Barnsjukhus, Karolinska Sjukhuset, Stockholm, Sweden; 7) Dept of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; 8) Dept of Women's and Children's Health, Karolinska Institutet, Pediatric Endocrinology Unit, Karolinska Hospital, Stockholm, Sweden; 9) 7 Paediatric Genetics Division, Centre for Paediatrics and Adolescent Medicine, Freiburg University Hospital, Freiburg, Germany; 10) Dept of Neonatal Intensive Care Unit, University of Parma, Parma, Italy; 11) Amrita Institute of Medical Sciences & Research Center, Dept of Pediatric Genetics, AIMS Ponekkara, Cochin, Kerala, India; 12) Dept of Radiology and Medical Imaging, Tokyo Metropolitan Kiyose Children's Hospital, Kiyose, Japan; 13) Dept of Pathology, Niguarda Ca' Granda Hospital, Milano, Italy; 14) Genética Clínica, Hospital Clínico Universidad de Chile, Santiago, Chile; 15) Dept of Pediatrics, Graduate School of Biomedical Science, Hiroshima University, Hiroshima, Japan; 16) Genomic Technologies Facility, Faculty of Biology and Medicine, Univ. of Lausanne, Lausanne, Switzerland.

The Kenny-Caffey syndrome (KCS; MIM 127000) and the more severe Osteocraniostenosis (OCS; MIM 602361) are conditions characterized by impaired skeletal development with small and dense bones, short stature, and primary hypoparathyroidism with hypocalcemia. While dominant transmission of KCS has been reported, including the mother and son originally described by Kenny, Linarelli and Caffey, OCS is usually lethal and has only been observed in sporadic cases. We speculated that KCS and OCS might share a common genetic etiology, and we studied a cohort of individuals with KCS or OCS with whole exome and direct gene sequencing. All ten individuals studied had heterozygous mutations in the gene FAM111A; one mutation was identified in four unrelated patients with KCS, and another one in two unrelated patients with OCS, and all occurred de novo in affected individuals (Unger et al, AJHG, in press, June 2013). Thus, OCS and KCS are allelic disorders, and FAM111A appears to be crucial to a pathway that governs PTH production, calcium homeostasis, and skeletal growth. What is the link between FAM111A and the parathyroid hormone (PTH) axis? FAM111A codes for a 611 amino acid protein with homology to trypsin-like peptidases that is localized to the cytoplasm and the nucleus. While its native function is unknown, it has been found to bind to the large T-antigen of SV40 and to restrict viral replication. Putative loss of function mutations have been observed in public databases, apparently without clinical correlates. Molecular modeling of FAM111A shows that residues affected by OCS/KS mutations do not map close to the active site, but are clustered at, or close to, a restricted region on its surface. This suggests that rather than impairing the catalytic activity of FAM111A, the KCS/OCS mutations may affect the interaction with other proteins. Based on the published data on interaction between FAM111A and SV40-LT and on the protein modeling results, we propose a model in which FAM111A activity is physiologically regulated by binding to an inhibitor; in the presence of KCS/OCS mutations, binding is impaired and activity is unrestricted. How FAM111A gain of function would result in the disruption of PTH activation remains to be determined. However, pinpointing the molecular defect in KCS and OCS upstream of the PTH system itself opens the possibility of therapeutic approaches using PTH analogues to treat chronic hypocalcemia and impaired bone growth.

155

Mutations in *PIEZO2* cause Gordon syndrome, Marden-Walker syndrome and distal arthrogryposis type 5. *M.J. McMillin[1], A.E. Beck[1,2], J.X. Chong[1,2], K.M. Shively[1], K.J. Buckingham[1], H.I. Gildersleeve[1], M. Splitt[4], A.S. Aylsworth[5], I.P.C. Krapels[6], C.J. Curry[7], M.I. Aracena[8,9], J.T. Hecht[10], J.A. Hurst[11], R.H. Scott[11], K. Devriendt[12], J.M. Graham, Jr.[13], J.D. Smith[3], H.K. Tabor[14], J. Shendure[3], D.A. Nickerson[3], M.J. Bamshad[1,2,3], University of Washington Center for Mendelian Genomics.* 1) Department of Pediatrics, University of Washington, Seattle, WA; 2) Seattle Children's Hospital, Seattle, WA; 3) Department of Genome Sciences, University of Washington, Seattle, WA; 4) Institute of Genetic Medicine, Center for Life, Newcastle upon Tyne Hospitals, Newcastle, UK; 5) Departments of Pediatrics and Genetics, University of North Carolina, Chapel Hill, NC; 6) Department of Clinical Genetics, Maastricht University Medical Centre, Maastricht, Netherlands; 7) Genetic Medicine Central California, Fresno/UCSF; 8) Unidad de Genética, Hospital Dr. Luis Calvo Mackenna, Santiago-Chile; 9) División de Pediatría, Facultad de Medicina, Pontificia Universidad Católica de Chile; 10) Department of Pediatrics, University of Texas Medical School, Houston, TX; 11) NE Thames Genetic Service, Great Ormond Street Hospital, London; 12) Center for Human Genetics, University Hospitals Leuven, Leuven, Belgium; 13) Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angles, CA; 14) Treuman Katz Center for Pediatric Bioethics, Seattle Children's Research Institute, Seattle, WA.

Gordon syndrome (GS), or distal arthrogryposis type 3, is a rare, autosomal dominant disorder characterized by congenital contractures of the hands and feet, cleft palate, dysmorphic facial features, and short stature. Analysis of exome sequencing data from five families with GS identified mutations in the gene piezo-type mechanosensitive ion channel component 2 (*PIEZO2*) in each family. Specifically, a novel substitution predicted to result in a missense mutation (p.R2686H) was transmitted from an affected mother to her three affected offspring in one GS family, and was confirmed to be de novo in three sporadic GS cases. A three base pair deletion predicted to result in the deletion of a single amino acid (p.E2727del) was confirmed to be de novo in the fifth GS family. The p.E2727del mutation is one of two mutations in *PIEZO2* recently reported to cause distal arthrogryposis type 5 (DA5). DA5 is distinguished from other DAs by the presence of ocular abnormalities, such as ptosis, ophthalmoplegia and keratoconus. However, there is considerable overlap between the phenotypes of GS and DA5. Accordingly, we used molecular inversion probe (MIP) sequencing to screen *PIEZO2* in 16 familial and 12 sporadic cases of DA5, three families with GS and seven families with distal arthrogryposis and cleft palate. *PIEZO2* mutations were found in 23/28 (82%) DA5 families and in the three additional GS families. In the DA5 cohort, the most common mutations were the p.E2727del mutation (n=10) and p.R2686H (n=4), the same two mutations originally found in the GS cases. Finally, the phenotypes of both GS and DA5 overlap with Marden-Walker syndrome (MWS) and indeed, sequencing of *PIEZO2* in a sporadic case of MWS revealed a novel mutation at a recurrent site resulting in a predicted p.R2686C substitution. Collectively, while DA5, GS and MWS have traditionally been considered distinct disorders, our findings indicate they are etiologically related and perhaps represent different manifestations of a single syndrome.

156

Mutation in the SH2 domain of PIK3R1 cause SHORT syndrome with partial lipodystrophy. *S. Johansson[1,2], K.K. Chudasama[1,2], J. Winnay[3], T. Claudi[4], R. König[5], I. Haldorsen[6], B. Johansson[1], J.R. Woo[3], D. Aarskog[1], J.V. Sagen[1], C.R. Kahn[3], A. Molven[6], P.R. Njolstad[1,7].* 1) KG Jebsen Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, Norway; 2) Center for Med Genet and Mol Medic, Haukeland Univ Hosp, Bergen, Norway; 3) Joslin Diabetes Center, Harvard Medical School, Boston, MA 02215, USA; 4) Department of Medicine, Nordlandssykehuset, N-8092, Bodø, Norway; 5) Department of Human Genetics, University of Frankfurt, G-60325, Frankfurt, Germany; 6) Department of Clinical Medicine, University of Bergen, Bergen, Norway; 7) Department of Pediatrics, Haukeland University Hospital, N-5021, Bergen, Norway.

SHORT syndrome (OMIM #269880) is characterized by Short stature, Hyperextensibility of joints, Ocular depression, Rieger anomaly and Teething delay as well as partial lipodystrophy and facial dysmorphic features. Until now, the genetic cause of SHORT syndrome has been unknown. We studied two unrelated families with short stature, ocular depression, Rieger anomaly, partial lipodystrophy and progeroid face. Although they had partial lipodystrophy, serum cholesterol and triglyceride levels were within the normal range. Moreover, serum alanine transaminase and γ-glutamyltransferase were normal, and abdominal computer tomography scans showed no sign of hepatic steatosis. Using whole-exome sequencing, we identified the same heterozygous PIK3R1 mutation (c.1945C>T corresponding to p.Arg649Trp) as the cause of the disorder in both families. PIK3R1 encodes p85α, a regulatory subunit of the phosphatidylinositol (PI) 3-kinase signaling pathway, which is critical for fundamental cellular processes such as metabolism, proliferation and survival. The mutation led to impaired interaction of p85alpha with IRS-1 and reduced AKT-mediated insulin signaling in patient fibroblasts and in reconstituted Pik3r1 knockout pre-adipocytes. Results from transgenic mice expressing the mutant protein will be presented at the meeting. Normal PI 3-kinase activity is critical for adipose differentiation and insulin signaling; the mutation in PIK3R1 therefore provides a unique link between lipodystrophy, growth and insulin signaling, adding new possibilities for diagnosis and targeted treatment in these diseases.

157

MAP4 defect underlines centrosomal organization as a central mechanism in growth regulation. *C.T. Thiel[1], D. Zahnleiter[1], N. Hauer[1], K. Kessler[1], Y. Sugano[2], S. Neuhaus[2], A.B. Ekici[1], H. Blessing[3], H. Sticht[4], H.-G. Doerr[3], A. Reis[1].* 1) Institute of Human Genetics, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany; 2) Institute of Molecular Life Sciences, University of Zurich, Zurich, Switzerland; 3) Department of Pediatrics and Adolescent Medicine, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany; 4) Institute of Biochemistry, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany.

Shortness of stature is a common medical concern in childhood, since 3 % of the general population present with a body height below -2 SD scores (SDS). After excluding already known defects the underlying cause remains unknown in approximately 80 % of patients. Recently, mutations in centrosomal proteins have been associated with severe syndromal forms of short stature with microcephaly. In a consanguineous family with two affected children with severe growth retardation of -7 SDS and relative macrocephaly (-2 SDS) we used genome wide homozygosity mapping and whole exome sequencing to identify a homozygous missense mutation in the coding region of the *MAP4* gene. This mutation was excluded in 372 control individuals, the 1000genomes project or ESP5400, and segregated in the family. We performed a quantitative real-time RT-PCR and detected a 70 % reduced *MAP4* expression level in the patient compared to healthy controls indicating a reduced stability of the mutant transcript. MAP4 is a major protein for microtubuli assembly during mitosis and ubiquitously expressed. The identified mutation in our patients led to a new phosphorylation site which is most likely phosphorylated by kinases in cell cycle regulation like cdk1 and predicted to alter microtubuli organization. Defects in the cell cycle were confirmed in the patient fibroblast cell line by detecting centrosome amplifications in 40 % of the mitotic cells. These numeric centrosomal aberrations were also present during interphase where centrosomes form the basal body initiating ciliogenesis. The identification of cilia duplications and disturbed cilia associated pathways in the patient cells confirmed a combined effect on mitosis and ciliogenesis. High expression levels in the somite boundaries of the zebrafish proposed a role in growth and body segment patterning by MAP4. We noted significant overlap of the patient's phenotype with Seckel syndrome and Microcephalic Osteodysplastic Dwarfism type Majewski also caused by defects in centrosomal proteins. Surprisingly, our patients did present with relative macrocephaly. This observation is explained by the only brain expressed MAP2 gene, likely to compensate the MAP4 defect in our patients. Therefore, our results confirmed dysregulation of centrosomal proteins to be associated with severe short stature, whereas disturbed neuron development, leading to microcephaly, underlies complex regulation mechanisms.

158

Best understanding of structural and functional impact of FGFR3 mutations at the same position (K650N, K650M, K650E) leading to both mild and lethal dwarfism. *D. Komla Ebri[1], C. Benoist-Lasselin[1], N. Kaci[1], P. Busca[2], G. Prestat[2], A. Munnich[1], F. Barbault[3], L. Legeai-Mallet[1].* 1) INSERM U781, Université Paris Descartes - Sorbonne Paris Cité, Institut Imagine, Paris, France; 2) UMR CNRS 8601, Université Paris Descartes, France; 3) 3ITODYS - Université Paris Diderot, France.

Three different missense mutations in the tyrosine kinase domain of fibroblast growth factor receptor 3 (FGFR3) affecting a lysine residue at position 650 lead to dwarfism with a spectrum of severity, hypochondroplasia (p. Lys650Asn), severe achondroplasia with developmental delay and acanthosis nigricans (p. Lys650Met), and thanatophoric dysplasia (p. Lys650Glu). Fgfr3 mutations induce a constitutive activation of the receptor characterized by a sustained phosphorylation responsible of growth plate defects. To understand the link between the severity of the clinical phenotype and a single amino acid change localized in the tyrosine kinase domain (TKD), we developed in silico and in vitro studies. Computational studies were conducted to get an atomic description of the p. Lys650Met, p. Lys650Glu and p. Lys650Asn built using a validated structural model of the FGFR3 kinase domain. Structural analyses indicate that a salt bridge between R655 and E686 is the cornerstone of the Tyr647 solvent exposition. This salt bridge is, significantly, disturbed with Lys650Glu mutation and is destroyed with Lys650Met and Lys650Asn mutants ($p < 0.01$). The activation of FGFR3 is initiated with the transphosphorylation of specific tyrosines in TKD, and induces conformation changes in FGFR3 activation loop. These FGFR3 activation loop mutation-related changes have been quantified by measuring the distance between the backbone gravity center of 650-656 aa (activation loop) and the backbone gravity center of 562-565 aa and 620-623 aa. We demonstrated a correlation within the major modifications and the severity of the chondrodysplasias (Lys650Asn 12.2 Å; Lys650Glu 16.4 Å; Lys650Met 16.8 Å). To confirm these structural changes, we studied in vitro the impact of the FGFR3 mutants on signaling pathways. Transient transfections of three DNA mutants at position 650 in chondrocyte show a gradient of phosphorylation levels correlated with the severity of the disease. Using HEK cells transfections, we noted an increased proliferation associated with the severity of the phenotype. Interestingly, we observed a higher activation of the Mapk, STAT and β-catenin signalling pathways related to the disease severity. In conclusion, the mutation of lysine 650 alters differently the conformation of the kinase domain thus leading to activate unusual signalling pathways. Various biological mechanisms seem to be responsible for mild and lethal dwarfism.

159

X-linked osteoporosis and fractures: an unexpected genetic cause. *F.S. van Dijk[1], M.C. Zillikens[2], D. Micha[1], M. Riessland[3], C.L.M. Marcelis[4], C.E. de Die-Smulders[5], J. Milbradt[3], A.A.M. Franken[6], G.J. Harsevoort[7], K.D. Lichtenbelt[8], J.E. Pruijs[9], M.E. Rubio-Gozalbo[10,11], R. Zwertbroek[12], M. Hammerschmidt[13], R. Bijman[1], C.M. Semeins[14], A.D. Bakker[14], V. Everts[14], J. Klein-Nulend[14], N. Campos-Obando[2], A. Hofman[15], A.J.M.H. Verkerk[2], A.G. Uitterlinden[2,15], A. Maugeri[1], E.A. Sistermans[1], Q. Waisfisz[1], H. Meijers-Heijboer[1], B. Wirth[3], M.E.H. Simon[16], G. Pals[1].* 1) Clinical Gen, VU Med Ctr, Amsterdam, Netherlands; 2) Department of Internal Medicine, Erasmus MC, Rotterdam, the Netherlands; 3) Institute of Human Genetics, Institute of Genetics and Center for Molecular Medicine Cologne, University of Cologne, Cologne, Germany; 4) Department of Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands; 5) Department of Clinical Genetics, Maastricht University Medical Centre, Maastricht, the Netherlands; 6) Department of Internal Medicine, Isala Clinics Zwolle, Zwolle, the Netherlands; 7) Department of Orthopedics, Isala Clinics Zwolle, Zwolle, the Netherlands; 8) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, the Netherlands; 9) Department of Pediatric Orthopaedics, University Medical Center Utrecht, Wilhelmina Childrens Hospital, Utrecht, the Netherlands; 10) Department of Pediatrics, Maastricht University Medical Center, Maastricht, the Netherlands; 11) Department of Pediatrics and Laboratory Genetic Metabolic Diseases, Maastricht University Medical Center, Maastricht, the Netherlands; 12) Department of Internal Medicine, Westfries Gasthuis, Hoorn, the Netherlands; 13) Department of Developmental Biology, University of Cologne, Cologne, Germany; 14) Department of Oral Cell Biology, ACTA, University of Amsterdam and VU University Amsterdam, MOVE Research Institute Amsterdam, Amsterdam, the Netherlands; 15) Department of Epidemiology, Erasmus MC Rotterdam, the Netherlands; 16) Department of Clinical Genetics, Erasmus MC, Rotterdam, the Netherlands.

Objectives: Osteoporosis with fractures as an X-linked trait, was first reported as a rare type of OI by Sillence in 1980. OI is genetically heterogeneous with an estimated 90% due to dominant mutations in the COL1A1 or COL1A2 gene and approximately 10% due to recessive mutations and other unknown causes. Our goal is to discover new genetic causes of OI and related disorders. Methods: We investigated all referred cases that had been clearly labelled as OI (all types) but where no genetic cause had been identified by sequence analysis of all known OI genes. Depending on the pedigree, specific genetic methods including exome sequencing were used to identify the causative gene. Results : In one large family with a referral diagnosis of OI type I, normal collagen type I electrophoresis and absence of COL1A1/2 mutations, X-linked inheritance appeared likely based on the pedigree. Exome sequencing of the X chromosome identified a putative pathogenic frameshift mutation in a candidate gene in affected family members. Four additional families with pathogenic mutations were discovered. In vitro and in vivo studies have been performed which support pathogenicity of the identified mutations. Furthermore, a rare variant in our candidate gene appeared upon testing in a large group of elderly subjects, associated with a twofold increased fracture risk in elderly female carriers in the normal population. Conclusions: For the first time we report a genetic cause of X-linked osteoporosis and fractures supported by in vitro and in vivo studies. Furthermore, the discovery of a rare variant, associated with a twofold increased fracture risk in elderly females, indicates genetic variation in the identified candidate gene as a novel etiological factor involved in osteoporosis.

160

Identification of Rare Genetic Variants in High-Risk ASD Families and Their Role in a Large ASD Case/Control Population. *C. Hensel[1], N. Matsunami[2], D. Hadley[3], G.B. Christensen[4], C. Kim[3], E. Frackelton[3], K. Thomas[3], R. Pellegrino da Silva[3], J. Stevens[2], L. Baird[2], B. Otterud[2], K. Ho[1], T. Varvil[2], T. Leppert[2], C. Lambert[4], M. Leppert[2], H. Hakonarson[3,5].* 1) Lineagen, Inc., Salt Lake City, UT; 2) Department of Human Genetics, University of Utah, Salt Lake City, UT; 3) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 4) Golden Helix, Inc. Bozeman, MT; 5) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA.

Background: Genetic variation plays a significant etiological role in autism spectrum disorders (ASDs), and numerous studies documenting the relevance of copy number variants (CNVs) and single nucleotide variants (SNVs) in ASD have been published. **Objectives:** This study was designed with three goals in mind. The first goal was to identify CNVs present in high-risk ASD families and to determine which of those CNVs contribute etiologically to ASD in the general population. The second goal was to confirm the findings of several published ASD CNV studies using a larger case/control population, to determine the potential clinical utility of those CNVs in the genetic analysis of children with ASD. The third goal was to determine if any SNVs identified as potential risk variants in high-risk ASD families supported their potential role as risk alleles in the same case/control population. **Methods:** CNVs in high-risk ASD families were identified using the Affymetrix SNP6.0 microarray. SNVs were identified by sequence capture in regions of genetic linkage and in published ASD candidate genes. CNVs and SNVs subsequently were evaluated in a set of 3000 ASD cases and 6000 controls using a custom Illumina iSelect array followed by molecular confirmation of significant variants. **Results:** We identified 153 putative ASD-specific CNVs in 55 affected individuals from 9 multiplex ASD families. These CNVs were not observed in control samples from Utah CEPH families. Case/control analysis revealed that 14 CNVs from high-risk ASD families were observed in unrelated ASD cases and had odds ratios greater than 2. We also identified CNVs not detected in our high-risk families using SNVs probes on the array, suggesting that some genetic regions can be impacted at both the structural and sequence levels. Findings for published CNVs indicated that many appeared to increase the ASD risk only slightly, since these CNVs had odds ratios less than 2. Multiple rare SNV were observed in unrelated ASD cases and in no controls, suggesting that variants in this gene may be risk factors for ASD. **Conclusions:** Genetic variants identified in high-risk ASD families also play a role in ASD etiology in unrelated ASD cases. The absence of 10 of these variants from public ASD databases suggests that they represent previously unidentified risk variants. These variants lay the groundwork for the development of a more sensitive test to use in the genetic evaluation of children with ASD.

161

Discovery of cryptic chromosomal abnormalities in clinically-referred youth with neuropsychiatric disorders. *V. Pillalamarri[1,8], A. Doyle[1,3,5,6,8], H. Brand[1,5], M.R. Stone[1], I. Blumenthal[1], C. O'dushlaine[3], E. Braaten[5], J. Rosenfeld[7], S. McCarroll[3,4,6], J. Smoller[1,2,5], M.E. Talkowski[1,3,5].* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard; 4) Department of Genetics, Harvard Medical School; 5) Departments of Psychiatry, Neurology, Harvard Medical School; 6) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard; 7) Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA; 8) Co-First Author.

Unequivocal evidence suggests that rare structural variation (SV) is among the strongest genetic risk factors for neuropsychiatric disorders (NPD), including autism and schizophrenia. Recently, we showed that balanced chromosomal abnormalities (BCAs) detected at cytogenetic resolution represent a unique class of highly penetrant SV resulting in heterozygous inactivation at a single point in the genome. To date, the impact of submicroscopic BCAs have not been routinely investigated as they remain cryptic to all conventional technologies other than deep whole-genome sequencing (WGS). Here, we define the landscape of cryptic BCAs in a clinically-referred youth cohort with NPDs, including autism and early onset psychiatric phenotypes as part of a longitudinal data collection. We sequenced 33 subjects using customized large-insert or "jumping" library WGS to a median insert coverage of ~60x with an insert size of ~2.6kb. We find 98.1% sensitivity to detect cryptic BCAs. We used a convergent genomic approach to support the broader significance of associated loci by comparing copy number variant (CNV) data from over 34,000 clinical diagnostic cases and 14,000 controls, as well as existing GWAS and exome data. Our approach uncovered a spectrum of BCAs including *intra-chromosomal* inverted insertions, *inter-chromosomal* excision/insertion and exchange events, and *semi-balanced* BCAs with small CNVs at the breakpoint. As expected, ~55-65% of cryptic BCAs represent differences from the reference assembly and are observed in all individuals with adequate coverage. Across 33 subjects, we detected about 50 balanced exchanges per individual, with most events being polymorphic. However, we observed 10 genes and one microRNA (mir1256) that were disrupted by a private cryptic BCA and not observed in other individuals. All events were validated by PCR and the parental origin was evaluated. Convergent genomics and network analyses implicated four of these loci, each with a direct, first-order interaction to a previously implicated ASD locus (*AKAP13*, *IQGAP1*, *ETV4*, *CTNNA3*), as well as a putative novel locus (*UBE2F*; CNV burden $p = 8.3\times10^{-4}$). For *CTNNA3*, we discovered an upstream disruption by a reciprocal translocation in a second independent case with ASD. We note as many as five putative contributors to the phenotypes of 33 patients. These data suggest that cryptic BCAs are an important and uncharacterized component of the genetic architecture of NPDs.

162

Mechanism, prevalence, and more severe neuropathy phenotype of the Charcot-Marie-Tooth, Type 1A triplication. *V. Gelowani[1], P. Liu[1], F. Zhang[2], V. Drory[3,6], S.B. Shachar[4], E. Roney[1], W.B. Burnette[5], J. Li[5], A. Orr-Urtreger[4,6], J.R. Lupski[1,7,8].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China; 3) Department of Neurology, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel;; 4) Genetic Institute, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel; 5) Department of Neurology, Center for Human Genetics Research, Vanderbilt University; 6) Sacker Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel;; 7) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 8) Texas Children's Hospital, Houston, TX, USA.

Early studies in the 1930s on the *Bar* locus in drosophila raised questions about molecular origin, mechanism, and frequency of triplications. To address these questions experimentally in humans we chose as our genetic model Charcot-Marie-Tooth, type 1 (CMT1A, which is one of the most common dominant disease traits and due to duplication. We hypothesized that a) triplications convey a more severe neuropathy phenotype compared to duplications, b) triplications arise from duplications through nonallelic homologous recombination (NAHR), and c) that the frequency of triplication arising from a pre-existing duplication is higher compared to that of duplication arising from a wild type allele.

We identified families with one member having the disease at the very extreme end of severity. Array CGH and interphase FISH identified individuals with either neutral copy number (unaffected), duplication (trait manifestation), or triplication (severe disease) within the CMT1A families. These data support the gene dosage hypothesis as the patients carrying triplications were more severe by both clinical examination and objective electrophysiological measurements compared to patients with duplications. To investigate the mechanism of crossover, we designed microsatellite genotyping assays to phase haplotypes. In both families, the *de novo* triplication arose from a pre-existing duplication during maternal meiosis. The triplication arose through one interchromosomal NAHR event in one family, and involved two steps in the other family. Theoretically, there are more ways for a chromosome to have copy number gain through NAHR when there is a duplication at the locus than when there is not. Data extracted from de-identified CMT samples analyzed by MLPA (courtesy A. Medeiros, R. Moore, Dr. J. Higgins, Quest/Athena Diagnostics) showed that the frequency of triplication among CNV gains is ~1/500. Thus, the frequency of triplication arising from a pre-existing duplication is higher, by about 200-fold, compared to the frequency of *de novo* duplication (1.73×10^{-5}) arising from a wildtype allele. We conclude that CMT1A triplication causes a severe clinical phenotype and its prevalence may have been substantially underestimated. Patients with severe neuropathy and with a CMT1A duplication family history should be evaluated for triplication using adequate molecular assays that can distinguish triplication from duplication.

163

Whole Genome Sequencing of two individuals with excessive numbers of de novo CNVs. *P. Liu[1], K. Walter[2], A. Wuster[2], T. Gambin[1], V. Gelowani[1], K. Writzl[3], S. Lindsay[2], C. Carvalho[1], M. Withers[1], C. Gonzaga[1], J. Wiszniewska[1], A. Patel[1], B. Rautenstrauss[4], R. Gibbs[1], M. Hurles[2], J. Lupski[1].* 1) Molec & Human Gen, Baylor College Med, Houston, TX; 2) Wellcome Trust Sanger Institute, Hinxton, UK; 3) Institute of Medical Genetics, UMC, Ljubljana, Slovenia; 4) Medical Genetics Center, Bayerstrasse 3-5, Munich, Germany.

Large de novo copy number variation (dnCNV) was estimated to occur constitutionally about once in every 100 individuals. It is extremely rare to observe individual genomes with two or more independently generated dnCNVs. Here we describe two subjects with excessive numbers of dnCNVs. They were referred to clinic because of developmental problems and multiple congenital anomalies. Genome-wide and region-specific array CGH analyses identified eight (50kb to 6.4Mb) and ten (210kb to 4.7Mb) dnCNVs in the two subjects, respectively. In order to characterize the full spectrum of dnCNVs and to glean insights into mechanism, whole genomes of the two trios were sequenced by Illumina Technology to an average depth of sequence coverage of 25X. Two CNV callers, BreakDancer and Pindel, were used to identify new dnCNVs or complexities not detected by aCGH, and to help obtain breakpoint sequences. All the dnCNVs identified from WGS analysis only were in proximity to the large dnCNVs originally revealed by arrays, suggesting that they likely belong to individual complex genomic rearrangements, and that the spectrum of dnCNVs in these two subjects does not seems to comprise smaller sizes. In addition, duplications in both subjects appear in the form of tandem duplication as well as insertional translocation. In a few loci, breakpoint complexities and triplications were identified, suggesting that the rearrangements were likely produced by replication mechanisms. Using variants specific to each parent, haplotypes were phased in patient genomes. In both patients, the dnCNVs were a mixture of paternal and maternal events, consistent with postzygotic timing of the CNV occurrence. Interestingly, in one of the subjects, the vast majority of dnCNVs was derived from the maternal chromosomes. In the same patient, de novo point mutations preferentially occurred in the paternal chromosome, may be reflecting advanced paternal age. The genomic distribution of de novo point mutations are being examined and validated in a comprehensive manner to investigate potential association with sites of dnCNV. Our results document a genome-wide spectrum of dnCNVs in rare cases of subjects with excessive number of mutations, and we suggest that errors in the cellular DNA replication machinery could lead to a propensity for long distance template switching events and multiple independent de novo rearrangements. Our findings have important implications for genomic disorders, cancer and evolution.

164

Utilization of Next Generation Sequencing to detect and assign pathogenicity to balanced rearrangements identified by conventional cytogenetics. *U. Aypar[1], K. Pearce[1], E. Thorland[1], J. Evans[2], V. Sarangi[2], C. Wang[2], Y. Asmann[3], N. Hoppman[1].* 1) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 2) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 3) Department of Health Sciences Research, Mayo Clinic, Jacksonville, FL.

Balanced rearrangements are found in approximately 1 in 175 phenotypically normal individuals. However, balanced rearrangements can result in an abnormal phenotype due to cryptic deletions or duplications at the breakpoints, disruption of a gene, position effects resulting from the rearrangement, or uniparental disomy. When detected by karyotyping in a patient with abnormal phenotypic features, parental studies are typically performed in order to determine pathogenicity; however even when de novo, such rearrangements are thought to be pathogenic only 6.7% of the time. Chromosomal microarray (CMA) is useful in detecting cryptic imbalances in some cases; however, the majority have normal array results, and in these cases, no additional clinical testing is currently available to further assess pathogenicity. Therefore, we performed next generation sequencing in four patients, three with abnormal phenotypic features and one with a normal phenotype, in whom a de novo balanced rearrangement was detected by conventional chromosome analysis and normal CMA results were obtained. Our goal is to develop a reflex test for patients with de novo balanced rearrangements in order to provide a non-equivocal test result. Samples with a (1;9)(q44;p13)dn, t(2;3)(q23;q27)dn, t(2;8)(q35;q22.1)dn, and inv(6)(p12;q27)dn were sequenced by paired-end whole genome sequencing on the Illumina HiSeq2000. A variety of bioinformatics algorithms designed to detect structural variants were used, including Delly and CREST. In two of the three cases with abnormal phenotype, disruption of a gene with known pathogenic mutations was identified. The third sample had no genes disrupted by the rearrangement. The patient with a normal phenotype had disruption of two genes with no known clinical significance. Based on these findings, we conclude that the clinical utility of next generation sequencing in the context of cytogenetically balanced rearrangements is significant, and incorporation of this technology to routine cytogenetic testing is recommended.

Rearrangement	Breakpoint 1	Breakpoint 2
46,XX,t(1;9)(q44;p13)dn	1q44: AKT3	9p13: no gene
46,XX,t(2;3)(q23;q27)dn	2q23: no gene	3q27: TP63
46,XX,t(2;8)(q35;q22.1)dn	2q35: MREG	8q22.1: LINC00534
46,XX,inv(6)(p12;q27)dn	6p12: no gene	6q27: no gene

.

165

Sequencing of unbalanced translocation junctions reveals mutational mechanisms and gene fusions. *B. Weckselblatt, M.K. Rudd.* Human Genetics, Emory University, Atlanta, GA.

Unbalanced chromosome translocations are a major contributor to intellectual disability and other neurodevelopmental disorders. Most translocation breakpoints are not recurrent and the mechanisms of translocation formation are largely unknown. Sequencing translocation breakpoints can point to specific types of double-strand break repair and reveal breakage-susceptible regions of the genome. In addition, sequence analysis of breakpoint junctions can reveal more complex rearrangement structures than expected from copy number studies alone. In a cohort of 55 human subjects with unique unbalanced translocations, we are characterizing chromosome breakpoints using a combination of array comparative genome hybridization, targeted sequence capture, and next-generation sequencing. Of the 33 sequenced translocations, 22 have between zero and four basepairs (bp) of microhomology at the breakpoint junction, consistent with nonhomologous end joining (NHEJ) or microhomology-mediated break-induced replication (MMBIR). Four junctions had 450-3023 bp of sequence homology typical of nonallelic homologous recombination (NAHR) between interspersed repeats. Notably, seven translocations had short inserted sequences at the junctions, suggesting a DNA replication-based mechanism of repair. Mapped insertions originate from sequence nearby the junction or up to 56 kilobases away from the translocation breakpoint. A recent study of 12 unbalanced translocation breakpoints suggested that the primary mechanism of rearrangement was NAHR. However, our sequence analysis of unbalanced translocations junctions revealed NHEJ/MMBIR as the dominant form of repair, which is consistent with other large-scale studies of copy number variation (CNV) breakpoints. Ongoing studies include sequence analysis of translocations between three, four, or five different chromosomes that may arise via chromothripsis. Junction sequencing has also uncovered two potential novel fusion genes, *FSTL5-PRDM16* and *SMOC2-PROX1*. In both translocations, disrupted genes are in the same direction, breakpoints lie within introns, and the resulting fusion is predicted to be in-frame. As in cancer genomes, fusion genes can lead to gain-of-function and contribute to disease. Breakpoint sequencing of our large collection of chromosome rearrangements provides a comprehensive analysis of molecular mechanisms of translocations.

166

Identification of a Deletion in the LRP1b Gene associated with Megalencephaly in the Sudden Infant Death Syndrome. *D.S. Paterson[1], H.C. Kinney[1], K. Schmitz Abe[2], F. Rahimov[2], F.L. Trachtenberg[3], E.A. Haas[4], H.F. Krous[4], K. Markianos[2].* 1) Pathology, Boston Childrens Hospital, Boston, MA; 2) Program in Genomics , Department of Pediatrics Boston Childrens Hospital, Boston, MA; 3) New England Research Institutes, Watertown, MA; 4) Rady Children's Hospital San Diego and University of California, San Diego School of Medicine, La Jolla, CA.

The sudden infant death syndrome (SIDS) is defined as the sudden death of an infant less than 12 months of age that is associated with a sleep period and remains unexplained after a complete autopsy and death scene investigation. It is the leading cause of postneonatal death in the United States with an overall incidence of 0.57/1,000 live births. The cause of SIDS remains unknown, but it is thought to have complex etiology involving multiple environmental and genetic risk factors. To date however, genetic studies in SIDS have failed to identify any variant that is necessary or sufficient for a SIDS death to occur. In other sporadic diseases with complex etiology, including autism, schizophrenia and Crohn's disease, high penetrance cytogenetic abnormalities such as copy number variations (CNVs) play a role in disease pathogenesis. In this study, we performed genome wide analysis of SIDS cases in the San Diego SIDS Dataset in order to identify CNVs that potentially cause or contribute to the pathogenesis of SIDS. Using a comparative genomic hybridization array with DNA from 50 SIDS cases we identified a total of 384 CNVs, including a novel 100kb deletion in chromosome 2q22.1affecting the low density lipoprotein receptor-related protein 1b (LRP1b) gene. Using qPCR we confirmed the deletion in 2/50 original SIDS cases and in an additional 6/100 SIDS cases from the San Diego SIDS Dataset. The deletion was absent in 144 ethnicity and sex matched controls from Coriell Institute Human Variation Panels and in 1,200 healthy controls and ~8000 cases collected for other phenotypes. Moreover, the affected cases exhibited "somatic overgrowth", i.e., increased body metrics for age including megalencephaly, which has previously been reported, but was of unknown etiology, in SIDS. Thus, we report the identification of a novel deletion that affects the protein coding region of the LRP1b gene and that is present in ~ 5% (8/150) of SIDS cases but that has not been identified in over 9,000 non-SIDS samples. These observations suggest that gene variants altering LRP1b function play an important role in the pathogenesis of at least a subset of SIDS cases and are partially responsible for megalencephaly associated with the disease. .

167

A comprehensive microarray prenatal study: efficacy for both copy number and copy neutral changes. *S. Schwartz[1], R. Pasion[1], H. Cabral[1], R. Burnside[1], I. Gadi[1], E. Keitges[2], L. Kline[1], V. Jaswaney[1], K. Phillips[1], H. Risheg[2], B. Rush[1], J. Shafer[1], H. Taylor[1], J. Tepperberg[1], P. Papenhausen[1]*. 1) Laboratory Corporation of America, Research Triangle Park, NC; 2) Dynacare/Laboratory Corporation of America, Seattle, WA.

Utilizing an Affymetrix platform we have now performed over 8500 prenatal ancillary analyses, the largest to date and only major study employing a platform that includes SNPs. The prenatal reporting thresholds were 1 Mb for deletions and 2 Mb for duplications. However, alterations as low as 50 kb were reported when significant pathogenic genes were involved. Results from this work revealed that ~4.8% of over 4,000 pregnancies ascertained with ultrasound (US) had pathogenic copy number changes. This frequency varied based on the specific US anomaly, ranging from ~1.9% for "soft" markers to ~9.5% for multiple abnormalities. The indication of an US heart defect was associated with copy number changes in ~9.7% of the analyses (with fewer than 30% due to 22q deletions). The frequency of UPD and consanguinity in patients with any US abnormality was ~3.6%; but, prenatal analyses with multiple US anomalies correlated with consanguinity in ~6.0% of the studies. The frequency of pathogenic copy number changes in chromosomally normal AMA/anxiety patients was ~1.1%. These studies have not only proven the importance and efficacy of utilizing a platform with genotyping capabilities, but have yielded some illuminating results including: (a) The importance of establishing clear reporting cut-offs, which were based on the previous analysis of almost 50,000 pediatric arrays; (b) The frequency of prenatal detection of variants of unknown significance was only 1.0-1.3%; (c) The recognition of changing referral patterns for prenatal array analysis with the frequency of AMA/anxiety referral studies which has almost doubled since the NICHD study (18.5% to 32.7%); (d) The usefulness of finding pathogenic copy number abnormalities not only in pregnancies ascertained with US anomalies, but in all patients; (e) The elevated frequency of consanguinity in referral due to US abnormalities suggests the causal importance of autosomal recessive diseases; (f) The technology was also very beneficial for resolving structurally abnormal chromosomes; (g) The analysis of cells in culture longer than 2 weeks revealed an increased frequency of apparent culture artifacts which must be reported appropriately; (h) Lastly, these studies showed that a firm background in interpreting pediatric array variation is necessary since both new pathogenic syndromes and consanguinity have been detected prenatally at a higher frequency than expected and need to be reported in the appropriate clinical context.

168

Non-random, locus-specific differences in DNA accessibility are present in homologous metaphase chromosomes. *W.A. Khan[1,3], P.K. Rogan[2,3,4], J.H.M. Knoll[1,3,4]*. 1) Department of Pathology; 2) Departments of Biochemistry and Computer Science; 3) University of Western Ontario, London, Ontario, Canada; 4) Cytognomix, London, Ontario, Canada.

Condensation differences between heterochromatin and euchromatin along the lengths of homologous, mitotic metaphase chromosomes are well known. This study describes differences in metaphase compaction between homologous euchromatic loci. We report molecular cytogenetic data showing local differences in condensation between homologs that are related to differences in accessibility (DA) of associated DNA probe targets. Reproducible DA was observed at ~10% of 450 distinct genomic regions mapped by single copy fluorescence in situ hybridization (scFISH). Fourteen short (1.5-5kb) sc and low copy (lc) FISH probes (from chromosomes 1, 5, 9, 11, 15, 16, 17, 22) targeting genic and non-genic regions with and without DA were developed and hybridized to cells from 10 individuals with cytogenetically-distinguishable homologs. Differences in hybridization were non-random for 6 genomic regions (*RGS7*, *CACNAB1*, *HERC2, PMP22*:IVS3, *ADORA2-B*:IVS1, *ACR*) and were significantly-biased towards the same homolog ($p < 0.01$; n = 355 cells). The imprinted paternal chromosome 15 in a three-generation pedigree also showed non-random bias in DA. DNA probes within *CCNB1*, *C9orf66*, *ADORA2B*:Ex 1-IVS1, *PMP22*:IVS4-Ex 5, and a nongenic region within 1p36.3 did not show DA, while *OPCML* showed unbiased DA. A subset of probes was mapped onto chromosome topography by FISH-correlated atomic force microscopy (AFM). To quantify DA and pinpoint probe locations, we performed 3D-structured illumination super-resolution microscopy (3D-SIM). 3D anaglyph videos showed genomic regions with DA having nearly 5-fold larger differences in volumetric integrated probe intensities between homologs. Additional non-DA probes (*NOMO1*, *NOMO3*) hybridized to grooves in chromosome topography and exhibited a narrow range of probe depths (average: 0.08 μm) along axial and lateral axes of the 2 homologs. In contrast, probe for targets with DA (*HERC2*, *PMP22*:IVS3, *ACR*) significantly differed in probe depth (average: 0.77 μm) and volume ($p < 0.05$) between each homolog. Interestingly, genomic regions without DA are enriched in epigenetic marks (DHS, H3K27Ac, H3K4me1) of accessible interphase chromatin to a greater extent than regions with DA, suggesting these differences may be correlated with epigenetic marks established during the previous interphase. In summary, we present several lines of evidence that regional differences in condensation between homologs are programmed during metaphase chromosome compaction.

169

Rare variants contributing to age-related macular degeneration - Results from the International AMD Genomics Consortium. *W.M. Igl for the International AMD Genomics Consortium.* Department of Genetic Epidemiology, University of Regensburg, 93053 Regensburg, Bavaria, Germany.

Purpose: Age-related macular degeneration (AMD) is a common cause of blindness in older people with strong genetic contribution from common variants (CVs). In contrast, the role of rare variants (minor allele frequency<1%) to AMD has not yet been studied genome-wide. To delineate the role of rare variants (RVs) in AMD, we genotyped about 50,000 samples at a single genotyping center with a custom-modified Illumina exome chip, specifically based on sequencing experiments.

Methods: We analyzed currently available genotype data of 23,362 unrelated individuals of European descent, totaling 12,030 AMD cases vs. 11,332 controls of comparable age. We tested 103,947 autosomal, bi-allelic RVs for association with AMD using Fisher's Exact Test. All results were genomic controlled and tested for significance at α=0.05/100,000=5E-7. Sensitivity analyses using Exact and Firth's bias-corrected logistic regression adjusted for sex, age, ancestry principal components, and/or known CVs (Fritsche et al., NatGenet 2013) were performed for significant variants.

Results: Among the significant RVs were the known functional variant rs121913059 (R1210C, *CFH*, Effect Allele Frequency (T)=0.00013, odds ratio=24.88, p=2.06E-19) and the newly described rs147859257 (K155Q, *C3*, EAF(G)=0.0040, OR=3.12, p=1.21E-24). Additional novel RVs were discovered at loci known for common AMD variants: *CFH/CFHlike* (15 variants, 6.51E-31≤p≤4.83E-8), *CFB/C2* (2 variants, 1.27E-7≤p≤3.27E-7), and *ARMS2/HTRA1* (2 variants, 3.71E-11≤p≤2.91E-10). One locus as yet unpublished for AMD was detected with a sole significant RV in the *C9* gene (p=5.86E-16), which is part of the complement system. The effects were not substantially modified by adjusting for sex, age, or principal components. Several of the signals were substantially weakened by conditioning on known CVs; the known variant in *CFH* (rs121913059, p=2.41E-12), the *C3* variant (rs147859257, p=9.28E-19), and six other novel variants in *CFH/CFHlike* (2.10E-16≤p≤4.70E-7) and *C9* (p=1.64E-11) remained significant.

Conclusions: Using world-wide sample collections, this rare variant genome-wide association study (rGWAS) has identified strong associations with AMD in known and novel loci. While some RV effects are independent from CVs highlighting novel AMD mechanisms, others coinciding with CV signals potentially depict underlying functional entities worth further follow-up.

170

Identification of new rare coding variants associated with hemoglobin levels and platelet counts. *P. Livermore Auer[1,4], N. Chami[2], U. Schick[1,3], S. de Denus[2], C. Carlson[1], M. Dube[2], R.D. Jackson[5], J.D. Rioux[2], C.L. Kooperberg[1], U. Peters[1], J.C. Tardif[2], L. Hsu[1], A.P. Reiner[1,3], G. Lettre[2]*. 1) Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Universite de Montreal and Montreal Heart Institute, Montreal, Quebec, Canada; 3) University of Washington, Seattle, WA; 4) University of Wisconsin-Milwaukee, Milwaukee, WI; 5) The Ohio State University, Columbus, OH.

Introduction: Hematological traits are clinically important diagnostic and prognostic parameters. Inter-individual variation in these blood-related traits is heritable and genome-wide association studies (GWAS) have already implicated dozens of loci. The development of new genotyping arrays that build on variant catalogues from deep re-sequencing efforts now allows extension of the search for associations with blood cell phenotypes to low-frequency and rare coding variants. **Materials and Methods:** We genotyped 18,388 participants from the Women Health Initiative study and 6,796 patients from the Montreal Heart Institute biobank on the Illumina ExomeChip array. We analyzed hemoglobin levels, hematocrit, and white blood cell and platelet counts as quantitative phenotypes for association with single nucleotide variants (SNV) using linear regression and gene-based tests (T1, SKAT-O). To minimize the impact of phenotypic outliers on rare variant association results, extreme values of blood phenotypes were winsorized (bottom 0.5% and top 99.5%). **Results:** We replicated several associations with common SNVs previously identified by GWAS: *TMPRSS6* and *HFE* variants with hemoglobin levels ($P<5\times10^{-17}$); variants within the *17q12* locus, the *HLA* region and the *ABO* gene with white blood cell count ($P<1\times10^{-7}$); and SNVs near *ARHGEF3*, *SH2B3* and *JMJD1C* with platelet counts ($P<1\times10^{-16}$). We also found potential causal variants at several loci previously identified by GWAS, including a splice site mutation in *C1orf150* (minor allele frequency (MAF)=7%, $P=2\times10^{-11}$) and a *SH2B3* Glu400Lys missense variant (MAF=0.1%, $P=2\times10^{-8}$) associated with platelet count. Finally, we identified novel associations with rare coding alleles: a rare missense variant in the erythropoietin (*EPO*) gene (MAF=0.4%, effect=-0.3 g/dL per rare allele, $P=7\times10^{-7}$) and a splice site beta-thalassemia variant (*HBB*) (MAF=0.02%, effect=-2.3 g/dL per rare allele, $P=9\times10^{-9}$) associated with hemoglobin levels, and the somatic *JAK2* V617F mutation associated with platelet count (MAF=0.05%, effect=124×10^{9} platelets/L, $P=3\times10^{-22}$). **Conclusions:** Our results confirm the utility of exome-wide genotyping in large populations to validate and fine-map previously identified genetic association signals. It is also a powerful tool to discover new genes and biological pathways involved in complex diseases and traits.

171

Large-scale Whole Genome Sequencing Study for Bone Mineral Density: the UK10K Consortium. *H. Zheng[1], V. Forgetta[1], S. Wilson[2,3], C. Greenwood[1], N. Timpson[4], N. Soranzo[5], T. Spector[3], B. Richards[1,3].* 1) McGill University, Montreal, Quebec, Canada; 2) University of Western Australia, Western Australia, Australia; 3) King's College London, London, UK; 4) Bristol University, Bristol, UK; 5) Sanger Institute, UK.

Aim: To identify genetic variants, including rare variants, associated with bone mineral density (BMD), we performed an association study using whole genome sequencing data within UK10K Consortium (http://www.uk10k.org/). Methods: 3781 samples from TwinsUK (discovery study) and ALSPAC were whole genome sequenced in Sanger Institute. The haplotypes of sequenced individuals were then imputed into an additional 3896 TwinsUK GWAS samples (replication study). We tested the association of TwinsUK samples at lumbar spine (L1-L4) (LS), femoral neck (FN) and forearm (Total radius and ulna) BMD separately in these two datasets. The phenotypes were standardized adjusting for age, age2 and weight. We undertook single variant tests, as well as a collapsing test of coding variants, with GEMMA and skatMeta, respectively. We then meta-analyzed the association results from the sequenced and imputed TwinsUK datasets. We set significant and suggestive thresholds for single variant test at 5×10-8, and 5×10-6, and set significance threshold for collapsing tests at 2.5×10-6 (0.05 over ~20,000 genes across the genome). Results: In the single variant test meta-analysis of data from the discovery and replication studies for forearm BMD, 143 variants from WNT16 locus were significant, and another 125 variants were suggestive. For LS and FN BMD, we identified 156 (including 11 LRP5 variants) and 125 suggestive variants. In the metaSKAT analysis of coding variants, we identified WNT16 and HSPB1 to be strongly associated with forearm BMD, however, when we excluded common variants (MAF>5%) in gene region, WNT16 was no longer significant, suggesting common missense variants in WNT16 drove the association signal. Nine rare coding variants in HSPB1 were collapsed and had an association P-value of 9.8×10-7, and this association was largely driven by rs28708645 (MAF= 0.005, P= 9.4×10-8). We also identified 3 genes (including LRP5) and 11 genes for LS and FN BMD, respectively. Conclusion: In this whole genome sequencing data analysis for BMD, we identified a novel, rare missense variant in a non-canonical HSPB1 transcript (ENST00000447574) as an important genetic determinant of forearm BMD, and confirmed the association of variants in WNT16 and LRP5 with forearm and LS BMD, respectively, and highlighted other novel loci and genes. These variants will soon undergo additional replication in upcoming de-novo genotyping.

172

Identification of 6 novel loci associated with amino acid levels using single-variant and gene-based tests. *T.M. Teslovich[1], J.R. Perry[2], J.R. Huyghe[1], A.U. Jackson[1], A. Stančáková[3], H.M. Stringham[1], P.S. Chines[4], J.M. Romm[5], H. Ling[5], I. McMullen[5], R. Ingersoll[5], E.W. Pugh[5], K.F. Doheny[5], J. Kuusisto[3], F.S. Collins[4], K.L. Mohlke[6], M. Laakso[3], M. Boehnke[1].* 1) Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI; 2) Genetics of Complex Traits, Exeter Medical School, University of Exeter, Exeter, UK; Wellcome Trust Centre for Human Genetics, UK; and Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 3) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 4) Genome Technology Branch, National Human Genome Research Institute, Bethesda, Maryland, USA; 5) Center for Inherited Disease Research, Johns Hopkins University, Baltimore, Maryland, USA; 6) Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA.

Amino acid levels (AA) are correlated with several disease states, including Alzheimer disease and type 2 diabetes. While candidate gene and GWA studies have implicated common variants underlying variability in AA, the impact of low-frequency (LF; MAF .5% - 5%) and rare (MAF < .5%) variants on AA is largely unknown. To determine the effect of LF (nonsynonymous) variants on AA, we genotyped ~242,000 SNPs on the Illumina HumanExome Beadchip in 8,850 Finnish men from the population-based METSIM study. 60,517 variants were non-monomorphic and passed QC filters. The levels of nine amino acids (Ala, Gln, Gly, His, Ile, Leu, Phe, Tyr, Val) were measured by nuclear magnetic resonance in all subjects; after log-normalization, we generated inverse normalized residuals adjusting for age, age^2, and BMI, and tested for association using a linear mixed model assuming an additive genetic effect. We identified 17 associations at genome-wide significance ($p<5\times10^{-8}$) at 12 unique loci. The 3 novel loci are all associated with Gly level. Intergenic variant rs4841132 (MAF=17%, $p_{Gly}=2\times10^{-23}$), proximal to *PPP1R3B*, was included on the array due to its associations with fasting insulin, fasting glucose, and free cholesterol in medium HDL cholesterol particles. Synonymous variant rs16954698 (MAF=12%, $p_{Gly}=8\times10^{-13}$) creates a new initiation codon for *PKD1L2*; however, an adjacent gene, glycine cleavage system H (*GCSH*), is a strong biological candidate, and rs16954698 may be tagging an untyped causal variant in or near *GCSH*. Finally, 2 independent LF, nonsynonymous variants in glycine dehydrogenase (*GLDC*) are strongly associated with Gly (Q996H, MAF=.9%, $p=4\times10^{-58}$; V735L, MAF=4%, $p=3\times10^{-17}$). We performed gene-level tests of association using SKAT and the variable threshold (VT) test to aggregate nonsynonymous and essential splice site variants (SKAT MAF < 1%). SKAT and VT analyses each identified 3 genes associated with AA at genome-wide significance ($p<2\times10^{-6}$). While three of the associations are driven by single variants with large effect, we identified 3 gene-level associations due to aggregate evidence across multiple variants: *ALDH1L1* associated with Gly, SKAT $p=9\times10^{-8}$, 9 variants, minimum single-marker p (MSMP)=1×10^{-5}; *BCAT2* associated with Val, VT $p=2\times10^{-6}$, 3 variants, MSMP=5×10^{-5}; and *HAL* associated with His, VT $p=1\times10^{-7}$, 7 variants, MSMP=8×10^{-5}. Taken together, our results suggest that low-frequency and rare variants may have a substantial impact on amino acid levels.

173

Exome analysis in 65,653 European samples identifies novel low-frequency and common variants for type 2 diabetes. *C. Fuchsberger[1], A. Mahajan[2], J. Flannick[3], D. Pasko[4], N. Grarup[5], N. Robertson[2], X. Sim[1], N. Burtt[3], A. Morris[2] on behalf of the GoT2D Consortium.* 1) Ctr Statistical Genetics, Univ Michigan, Ann Arbor, MI, USA; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 3) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 4) University of Exeter, Exeter, UK; 5) University of Copenhagen, Copenhagen, Denmark.

Genome-wide association studies have led to impressive advances in the identification of common genetic variants associated with type 2 diabetes (T2D). The next challenge is to explore the role of low-frequency (LF; 1-5% MAF) and rare (<1% MAF) coding variants in T2D. Custom genotyping arrays, such as the Illumina HumanExome Beadchip, provide a cost-effective alternative to sequencing and based on our comparison with ~5,000 European individuals with deep (>70x) sequenced exomes capture 83.7% of the genetic variation with MAF >0.5%. To discover novel T2D loci and examine whether LF and rare coding alleles could explain established common variant GWAS association signals, we studied 23,483 T2D cases and 42,080 controls of European ancestry from 7 studies. We tested single variants for association with T2D using a linear mixed model (EMMAX) that accounts also for sample structure. In a fixed-effects meta-analysis of up to 175,403 high-quality autosomal variants, we identified one LF non-synonymous variant approaching study-wide significance (P<2.8×10-7) in a gene not mapping to an established T2D locus: FAM63A (missense Y285N; odds ratio (OR) 1.37 95% CI [1.22-1.53]; P=3.4×10-7, 1.6% MAF). The T2D risk allele of this variant was associated with increased fasting plasma glucose in 33,553 non-diabetic individuals (P=0.009). Three common variants in previously unreported T2D loci achieved study-wide significance: PAM (missense D563G; OR 1.17 [1.10-1.25]; P=1.1×10-7; 5.6% MAF), GPSM1 (missense S391L; OR 1.09 [1.06-1.12]; P=2.6×10-8; 25% MAF), and BAG6 (intronic; OR 1.09 [1.05-1.13]; P=2.9×10-8; 36% MAF).). The T2D risk alleles of two of these SNPs were also associated with reduced insulinogenic index. Gene-level tests (SKAT-o) of rare non-synonymous variants did not reveal any study-wide significant genes supported by multiple studies. These results highlight that LF variants with moderate effects are also contribute to T2D risk, but that LF coding variants of large effect (OR>1.5) are unlikely to be frequently associated with T2D.

174

Loss of function mutations in *SLC30A8* protect against type 2 diabetes. *J. Flannick[1], G. Thorleifsson[2], N.L. Beer[3], S.B.R. Jacobs[1], N. Grarup[4], NNF-CBMR Genetics, Go-T2D/T2D-GENES consortia, deCODE genetics, Pfizer-Broad-MGH-Lund T2D project.* 1) Molecular Biology, Massachusetts General Hospital, Boston, MA; 2) deCODE Genetics, Reykjavik, Iceland; 3) Oxford Centre for Diabetes, Endocrinology, and Metabolism, University of Oxford, Oxford, U.K; 4) The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

Type 2 diabetes (T2D) affects ≈300,000,000 people worldwide and is rising in prevalence, but current approaches to prevention only partially delay progression to disease. Loss of function coding mutations that protect against disease risk in human populations (e.g. Nav1.7 and pain, *PCSK9* and myocardial infarction) provide in vivo validation of targets for therapeutic development, but none have been previously described for T2D.

Through sequencing and genotyping in ~150,000 individuals from 5 ethnicities, we identified 12 rare protein truncating variants in *SLC30A8*, a gene that encodes a beta-cell zinc transporter (ZnT8) and that carries a common variant (p.W325R) previously associated with T2D risk, glucose and insulin levels. In aggregate, carriers of these rare variants had a 65% reduction in risk of T2D (p=1.7 × 10-6), and non-diabetic carriers of a frameshift mutation (p.K34SfsX50) in Iceland showed reduced glucose levels (-0.17 s.d., p= 4.6×10-4). The two most common variants, each individually associated with protection against T2D, encode unstable forms of ZnT8 and are degraded when expressed in vitro.

Although previous experiments in animal and cellular model systems had led to the hypothesis that reduction in *SLC30A8* activity increases T2D risk, these findings from human genetics indicate that haploinsufficiency of *SLC30A8* protects against T2D in human populations. These results thus unequivocally establish *SLC30A8* as the causal gene responsible for the common variant association signal, phase the directional relationship between gene activity and T2D risk, and illustrate the value of human population data for the generation of therapeutic hypotheses.

175

Whole genome sequencing of 2,850 Central-Northern European type 2 diabetes cases and controls reveals insights into functional mechanisms underlying disease pathogenesis. *K. Gaulton[1], J. Flannick[2], C. Fuchsberger[3], H.M. Kang[3], N. Burtt[2], J. Ferrer[4], M. Stitzel[5], M. Kellis[6], M. McCarthy[1], D. Altshuler[2], M. Boehnke[3], the GoT2D consortium.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford UK; 2) Medical and Population Genetics, The Broad Institute of Harvard and MIT, Cambridge MA; 3) Center for Statistical Genetics, University of Michigan, Ann Arbor MI; 4) Imperial College London, London UK; 5) National Human Genome Research Institute, Bethesda MD; 6) Department of Computer Science, MIT, Cambridge MA.

Type 2 diabetes (T2D) is a complex disorder with incompletely known etiology that affects millions of individuals worldwide. To further our understanding of the genetic factors and biological processes underlying T2D pathogenesis, we carried out whole genome sequencing of 2,850 T2D cases and controls of Central and Northern European origin as part of the GoT2D project. Low-pass (~4x) sequencing identified ~25M single-nucleotide variants, including >99% with minor allele frequency >0.1% in the sequenced individuals, allowing near complete evaluation of the contribution of variants in this allele frequency range to T2D risk. Subsequent imputation of variant genotypes into ~35K GWAS samples revealed novel loci harboring common variants approaching genome-wide significant association to T2D (P<1×10-7: HORMAD2, HSD17B12, CENPW) and putative lower frequency (MAF<.05) secondary signals at four known loci TCF7L2, CCND2, KCNQ1, and CDKAL1 (all P<1×10-6). Using this unbiased survey of variation, we then assessed to what extent broad classes of functional elements contribute to T2D using regulatory state and transcription factor binding maps from pancreatic islets, adipocytes, and nine ENCODE cell types. Variants overlapping sets of functional elements were tested for enriched association to T2D compared to sets of control variants (matched on genomic properties or in shuffled sites). Common associated variants were collectively enriched at enhancer elements (P = .005), and low-frequency associated variants at promoter elements (P = .004). We found heterogeneity across cell types whereby common variants are most prominently enriched at enhancers active in hepatocytes, adipocytes, and pancreatic islets and bound by specific factors active in these cell types such as NKX2.2, MAFB, and TCF7L2 (all P<.05). Patterns were unchanged when removing variants within 500kb of a known GWAs signal, demonstrating that enriched element types can prioritize novel susceptibility loci not strictly genome wide significant. These results suggest information about the non-coding genome can provide significant insight into the genetic and biological basis of T2D, and support the central importance of global regulatory mechanisms in specific cell types to disease pathogenesis. More broadly this study confirms whole genome sequencing as a valuable tool to dissect genetic factors and functional mechanisms contributing to complex disease.

176

Exome chip scan of 74,000 subjects of European descent and 18,000 subjects of African descent identify novel genes with functional mutations influencing adiposity traits. *I.B. Borecki, The CHARGE Adiposity Working Group.* Div Statistical Genomics, Washington Univ Sch Med, St Louis, MO.

The rapid increase in prevalence of obesity in the industrialized world along with associated metabolic derangements is a major public health concern costing billions of dollars each year. An understanding of the biological pathways that regulate adiposity is crucial to the development of intervention strategies. Published genome-wide association studies have revealed 40 and 14 loci across ethnicities influencing body mass index (BMI) and waist-to-hip ratio adjusted for BMI (WHRaBMI), respectively, many of which are annotated as having potential regulatory function. Here, we test the hypothesis that putatively functional exonic variants (missense and splice) also influence variation in these adiposity traits,. We organized a meta-analysis effort within the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium including 73,768 European descent (EU) and 17,713 African descent (AF) subjects for BMI, and 50,601 EU and 13,224 AF subjects for WHRaBMI, across 18 studies. The subjects were genotyped using the Illumina HumanExome v1.0 BeadChip, which covers ~18,500 genes. Variant-wise association tests were combined using fixed effects meta-analysis as implemented in METAL, and two gene-wise burden tests were implemented using the sequence kernel association test in the SKAT package. Using a criterion of -logP >6, variants in 9 genes were associated with BMI in EU: *BDNG, GIPR, MPR17L2, IFI30, ANAPC4, ADCY3, SLC39A8, KCNH2* and *MAST3*; and 1 gene in AF, *PLS1*. Some of these are novel genes that have not been previously recognized in association with adiposity (*ANAPC4, KCNH2* and *MAST3*), and they function in the cell cycle, in mitochondrial membranes, modulating gene expression in intestinal mucosa with inflammatory response, and in assembly of intestinal epithelium. Five variants in 4 genes were associated with WHRaBMI in EU: *RSPO3* and *COBLL1* and in novel genes *CCDC92, RAPGEF3*. These genes have been associated with Wnt signaling, adiponectin levels, and anxiety and depression using GWAS markers. Here we demonstrate for the first time, association of exonic mutations with a measure of fat distribution in population-based samples. These results support the value of exon screening as a means of identifying novel genes and variants influencing adiposity traits. A deeper understanding of adiposity-related pathways and processes they influence can potentially lead to novel therapeutic strategies.

177

Large duplications are associated with increased risk of obesity. *J.S. El-Sayed Moustafa[1], M. Falchi[1], R.G. Walters[2], I. Prokopenko[1], A.I.F. Blakemore[3], L.J.M. Coin[1,4], P. Froguel[1,5].* 1) Department of Genomics of Common Disease, Imperial College London, London, United Kingdom; 2) Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, Oxford, UK; 3) Section of Investigative Medicine, Imperial College London, London, UK; 4) Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia; 5) Centre National de la Recherche Scientifique (CNRS) - Unité Mixte de Recherche de l'Université Lille 2 (UMR) 8199, Institut Pasteur, Lille, France.

Obesity is rapidly becoming an increasingly serious public health concern, placing affected individuals at increased risk of various other health conditions. Despite an estimated heritability of 40-70%, much of the genetic variance of adiposity remains unexplained. We have investigated the contribution of large, rare genomic structural variants above 500kb to obesity susceptibility. We employed our copy number variant (CNV) prediction algorithm cnvHap to conduct genome-wide CNV prediction in a child obesity sample of 683 cases and 655 controls, and an adult obesity sample of 695 cases and 197 controls, all genotyped on the Illumina Human CNV370 Duo platform. Analyses were limited to CNVs of a minimum length of 500kb, and of a frequency below 1% in control subjects. All identified CNVs were confirmed visually, as well as using an alternative CNV prediction algorithm (pennCNV), with a subset of CNVs also experimentally validated using non-array-based methods including multiplex ligation-dependent probe amplification (MLPA) and quantitative PCR (qPCR). A total of 188 rare CNVs above 500kb were identified in our study sample. Permutation-based global large CNV burden analysis was conducted using PLINK v1.07. We found rare duplications above 500kb to be significantly enriched in child and adult obese cases compared to normal-weight controls (P=8.89×10^{-3} and P= 5.05×10^{-3}, respectively, pooled analysis P=1.07×10^{-3}). Functional annotation analysis also revealed enrichment for genes associated with schizophrenia within GSVs identified in obese subjects (P=2.2×10^{-5}). These results provide novel evidence that large duplications above 500kb may contribute to the "missing heritability" of obesity. The overlap between genes located within large CNVs above 500kb in obese subjects and genes associated with psychiatric disorders such as schizophrenia may also help explain the observed clinical and genetic overlap between these disorders.

178

A role for host-bacteria interactions in shaping patterns of genetic variation across human populations. *R. Blekhman[1,2], J.K. Goodrich[1,3], K. Huang[4], Q. Sun[5], R. Bukowski[5], J.T. Bell[6], T.D. Spector[6], A. Keinan[7], R.E. Ley[1,3], D. Gevers[4], A.G. Clark[1].* 1) Molecular Biology and Genetics, Cornell University, Ithaca, NY; 2) Neurology and Neuroscience, Weill Cornell Medical College, New York, NY; 3) Microbiology, Cornell University, Ithaca, NY; 4) 4.Genome Sequencing and Analysis Program, The Broad Institute of MIT and Harvard, Cambridge, MA; 5) Computational Biology Service Unit, Cornell University, Ithaca, NY; 6) Twin Research & Genetic Epidemiology, King's College London, U.K; 7) Biological Statistics and Computational Biology, Cornell University, Ithaca, NY.

The composition of bacteria in and on the human body varies widely across human individuals, and has been associated with multiple health conditions. While host genetic factors are expected to control the human microbiome through the immune system and metabolic pathways, elucidation of the genetic influence on microbiome composition has proven to be a challenge. Here, we present a genome-wide association study aimed at identifying human genetic variation associated with microbial diversity in multiple body sites. By mining the shotgun metagenomic data from the Human Microbiome Project for incidental host DNA reads, we gathered information on host genetic variation for 93 individuals for whom bacterial abundance data are also available. Using a two-stage discovery and validation approach, we identified 26 candidate human genes that are associated with microbiome composition in 15 host body sites. These genes are significantly enriched in immunity functional categories, and form an interaction network highly enriched with immunity-related functions. To investigate the evolutionary history of bacteria-associated host genes, we used available sequencing data from the 1000 Genomes Project, and find that these genes show a significant excess of highly differentiated allele frequencies among human populations. Moreover, these genes are over-represented with genes that have been identified in recent genome scans for positive selection and balancing selection. Combined, these results highlight the role of host immunity in determining bacteria levels across the body, and underline a possible role for the microbiome in driving the evolution of bacteria-associated host genes.

179

Using Ancient Genomes to Detect Positive Selection on the Human Lineage. *K. Prüfer, M. Lachmann, C. Theunert, M. Ongyerth, G. Renaud, M. Dannemann, T. Neandertal Genome Consortium, S. Pääbo.* Max Planck Institute for evolutionary Anthropology, Leipzig, Germany.

At least two distinct groups of archaic hominins inhabited Eurasia before the arrival of modern humans: Neandertals and Denisovans. The analysis of the genomes of these archaic humans revealed that they are more closely related to one another than they are to modern humans. However, since modern and archaic humans are so closely related, only about 10% of the archaic DNA sequences fall outside the present-day human variation whereas for 90% of the genome, Neandertal or Denisova DNA sequences are more closely related to some humans than to others. The fact that the archaic sequence often falls within the diversity of modern humans can be used to detect selective sweeps that affected all modern humans after their split from archaic humans since such sweeps will result in genomic regions where both the Neandertal and Denisova genomes fall outside the modern human variation. The genetic lengths of such external regions are proportional to the strength of selection, since stronger selection will lead to faster sweeps allowing less time for recombination to decrease their size. We have implemented a test for such external regions as a hidden Markov model. At each polymorphic position the model emits *ancestral* or *derived* based on whether the tested archaic genome carries the ancestral or derived variant of SNPs observed in present-day humans. The model was applied to 185 African genomes from the 1000 genomes phase 1 data. We identified thousands of external regions using the Neandertal and Denisova genomes, separately. Approximately one third of the regions are overlapping between the two genomes. These regions are significantly longer than regions only identified in only one of the archaic genomes. Based on this excess of overlap for long regions, we devise a measure to identify a set of regions that are candidates for selective sweeps on the human lineage since the split from Neandertal and Denisova.

180

Reference sample guided pooled sequencing identifies loss-of-function patterns across human populations. *A. Eran[1], M. Carneiro[1], G. del Angel[1], E. Banks[1], R. Poplin[1], M. Lek[1,2], G. van der Auwera[1], S. Fisher[1], S. Gabriel[1], D. Altshuler[1], D. MacArthur[1,2], M. Depristo[1], The 1000 Genomes Project Consortium.* 1) Broad Institute of Harvard and MIT, Cambridge, MA; 2) Massachusetts General Hospital, Boston, MA.

Originally deemed disease causing, about 100 loss-of-function (LoF) mutations may be detected in every typical human genome sequenced today. Understanding the typical LoF variation distribution is therefore essential for accurate genomic variation inference, especially in a clinical setting. Here we present an accurate cost-effective targeted sequencing framework, based on reference sample guided pooled variant calling, and apply it to confidently survey LoF variation in ~1000 individuals from 13 distinct populations. About 13,000 putative LoF variants originally detected in the 1000 Genomes Project were targeted, pooled along a reference sample, deeply sequenced, and jointly genotyped using an empirically derived site-specific error model. When compared to microarrays and The 1000 Genomes Project genotypes, deep targeted reference sample guided pooled sequencing showed concordance of over 99% with SNPs and 94% with indels. Using this approach, we examined population-specific LoF burden at the single gene and pathway levels. We find that protein-coding genes enriched with LoF variation in any single population were frequently involved in environmental response (p<1e-8) and agreed with known demographic histories. These tend to act extracellularly or at the plasma membrane, and include disease-implicated loci such as *ITGA7, TMEM67, AIRE*, and *HLA-DQB1*. Molecular mechanisms subject to differential gene inactivation between populations include natural killer cell mediated cytotoxicity in Africans (p<1e-8), transition metal ion binding in Europeans (p<1e-8), and adaptive immunity in East Asians (p<1e-8). These results improve our understanding of LoF variation across human populations, and illustrate the value of reference sample guided pooled sequencing for large-scale population studies.

181

Patterns of IBD sharing inferred from whole genome sequences of 962 European Americans. *C.V. Van Hout[1], F. Yu[2], X. Liu[3], E. Boerwinkle[2,3], A.G. Clark[1].* 1) Molecular Biology and Genetics, Cornell University; 2) Human Genome Sequencing Center, Baylor College of Medicine; 3) Human Genetics Center, University of Texas Health Science Center at Houston.

Past demography of a population is expected to leave a signature in the distribution of numbers and lengths of segments of the genome that are shared identity by descent (IBD) between individuals. Recently, methodological advances and the availability of whole genome sequencing data for large numbers of individuals within a single population have enabled high resolution inference of IBD block sharing, allowing investigation of recent human genealogic history. We identified pairwise-shared IBD blocks with the Beagle software (Browning & Browning, Genetics, 2013; Gusev et al., Genome Res., 2009) in whole genome sequence in a sample of 962 European Americans from the ARIC study. Assuming unilineal relationships, we estimated empirical pairwise kinship coefficients as a function of the genetic lengths of shared blocks. The distribution of kinship coefficients for 462,241 pairs of putatively unrelated study participants had an inter-quartile range of 0.00006-0.0003 and mean 0.0002 (between 5th and 6th cousins). This implies that in this sample, on average 92 pairs of participants are expected to share any given locus IBD. Recent origins of common ancestry would be expected to yield long blocks of IBD sharing, and as the time back to a common ancester grows, we expect the length of the IBD block to shrink at the same time that sequence differences will accumulate by mutation. We summarized the distribution of shared block lengths and the number of nucleotide differences in each block. Consistent with established models of human population history, we observed a block length distribution heavily skewed toward shorter segments. In addition, we observed an increase in the mean and variance of the number of nucleotide differences per kilobase in shorter shared IBD blocks, consistent with their increased age. The mutation accumulation within IBD blocks and block length reduction by recombination provide types of molecular clocks whose consistency under specified demographic models was tested. We conclude that IBD block inference from whole-genome sequence data provides and exceptionally rich platform for analysis of recent genealogic kinship and demography.

182

Title: Reconstructing the Genetic Demography of the United States. *R. Sebro[1], N. Laird[2], N. Risch[3,4,5].* 1) Radiology; Genetics, University of California, San Francisco, San Francisco, CA; 2) Department of Biostatistics, Harvard School for Public Health; 3) Institute for Human Genetics, University of California, San Francisco;; 4) Department of Epidemiology and Biostatistics, University of California, San Francisco; 5) Division of Research, Kaiser Permanente, Oakland, CA.

The United States (U.S) is a complex, multiethnic society shaped by immigration and admixture, but the extent to which these forces influence the overall population genetic structure of the U.S is unknown. We utilized self-reported ancestry data collected from the decennial U.S Census 2010 and allele frequency data from over 2000 SNPs for over 40 of the most common ancestries in the U.S. that were available from the Pan Asian Single Nucleotide Polymorphism (PASNP), Population Reference Sample (POPRES), 1000 Genomes, and Human Genome Diversity Panel (HGDP) databases. We utilized the relative proportions of individuals of each ancestry within each county, state, region and nation and calculate the weighted average allele frequency in these areas. We reconstructed the genetic demography of the U.S by examining the geographic distribution of Wright's Fst. Shannon's diversity index, H was calculated to assess the apportionment of genetic diversity at the county, state, regional and national level. This analysis was repeated stratifying by race/ethnicity. We analyzed households with spouses, using the phi-coefficient as a measure of assortative mating for ancestry. This analysis was repeated stratifying by age of the spouses (older or younger than 50). Most of the genetic diversity is between ancestries within county, but this varies by race/ethnicity, and ranges from 95% for Whites to 43% for Hispanics illustrating that the White ancestries are relatively homogeneously scattered throughout the U.S whereas the Hispanic ancestries show significant clustering by geography. Analysis of the mating patterns show strong within ethnicity assortative mating for American Indian/Alaska Natives, Asians, Blacks, Hispanic, Native Hawaiians/Pacific Islanders, and Whites, with ϕ = 0.30, 0.864, 0.92, 0.863, 0.478 and 0.832 respectively (P<1×10-324 for each) and significantly less correlation in the younger cohort. These results show demographic patterns of social homogamy which are slowly decreasing over time. One major implication is that data collected from different locations around the U.S are susceptible to both within- and between-location population genetic substructure, leading to potential biases in population-based association studies.

183

Assessing functional potential along the human genome by integrating comparative, population, and functional genomic data. *I. Gronau[1], B. Gulko[2], L. Arbiza[1], M.J. Hubisz[1], A. Siepel[1,2].* 1) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 2) Graduate Field of Computer Science, Cornell University, Ithaca, NY.

More than a decade after publication of the human genome sequence, little is still known about the effects of mutations in non-coding DNA. While genome-wide functional assays such as ChIP-seq, DNase-seq and RNA-seq are informative about possible biochemical functions of non-coding DNA, they provide little indication as to the importance of these functions to fitness at the organismal level. As a result, there is still much debate about the fraction of the human genome that is functional in the sense that mutations to these sequences influence fitness. Here, we aim to address this question using a novel approach that combines functional genomic data from a large collection of assays together with comparative genomic data and genome-wide human variation data. Using these different data sources, we assign each nucleotide in the genome a functional potential score (**FPS**), which represents the probability that this nucleotide is directly influenced by natural selection.

FPSs are computed in a two steps: first, functional genomic data are used to group together sites with similar functional indicators (e.g., RNA transcription, open chromatin and histone modifications); then, using patterns of sequence variation and divergence from closely related primate species, the fraction of sites under selection is estimated separately for each group and used as the FPS of all sites in that group. This estimation is carried out using a recently developed method called INSIGHT (Gronau et al., *MBE*, 2013), which is specifically designed to detect signatures of recent natural selection across dispersed collections of sites. We use this approach to compute genome-wide FPSs and display them in a genome browser track.

We find that focusing on signatures of recent natural selection provides our FPSs with a clear advantage over existing evolutionary conservation scores when used to classify putative functional non-coding elements such as enhancers, eQTLs and GWAS SNPs. We also examine the contribution of the different functional indicators to the FPS and find complex non-linear interactions between them. Finally, using a weighted average of FPS along the genome, we estimate that roughly 8% of the human genome is under selection. In conclusion, we anticipate that functional potential scores will provide a powerful tool in the ongoing efforts to characterize function in non coding regions of the human genome.

184

Genome-wide analysis of cold adaption in indigenous Siberian populations. *A. Cardona[1], T. Antao[2], L. Pagani[1], D.J. Lawson[3], C.A. Eichstaedt[1], B. Yngvadottir[1], C. Tyler-Smith[4], M.T.T. Shwe[5], J. Wee[5], I.G. Romero[6], S. Raj[7], R. Villems[8], M. Metspalu[8], R. Nielsen[9], E. Willerslev[10], B.A. Malyarchuk[11], M.V. Derenko[11], T. Kivisild[1].* 1) Division of Biological Anthropology, University of Cambridge, Cambridge, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 3) Department of Mathematics, University of Bristol, Bristol, United Kingdom; 4) Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 5) National Cancer Centre Singapore, Singapore; 6) Department of Human Genetics, University of Chicago, Chicago, US; 7) Department of Molecular Biology and Genetics, Cornell University, New York, US; 8) Department of Evolutionary Biology, University of Tartu, Estonia; 9) Department of Integrative Biology and Department of Statistics, University of California Berkeley, Berkeley, US; 10) Centre for GeoGenetics, Natural History Museum of Denmark and Department of Biology, University of Copenhagen, Copenhagen, Denmark; 11) Genetics Laboratory, Institute of Biological Problems of the North, Magadan, Russia.

After the dispersal out of Africa, where hominins evolved in low latitude and warm environments for millions of years, our species has colonised different climate zones of the world, including high latitudes and cold environments. The extent to which human habitation in (sub-) Arctic regions has been enabled by cultural buffering, short-term acclimatization and genetic adaptations is not clearly understood. Present day indigenous populations of Siberia show a number of phenotypic features, such as increased basal metabolic rate, low serum lipid levels, increased blood pressure, short stature and broad skulls that have been attributed to adaptation to the extreme cold climate. We have genotyped 200 individuals from ten indigenous Siberian populations for 730,525 SNPs across the genome to identify genes and non-coding regions that have undergone unusually rapid allele frequency and haplotype homozygosity change in the recent past. At least three distinct population clusters could be identified among the Siberians, each of which showed a number of unique signals of selection. We present a candidate list of cold adaption loci containing novel regions and genes that showed significant signals of positive selection. Amongst other functions, our candidate regions contain genes associated with energy regulation, non-shivering thermogenesis, vascular smooth muscle contraction, regulation of blood pressure and oxidative stress.

185

Genetic variation is a major source of transcriptional variation in human induced pluripotent stem cells. *N. Kumasaka, F. Rouhani, A. Bradley, L. Vallier, D. Gaffney.* Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

Induced pluripotent stem cells (iPSCs) are a potentially powerful model system for studying human disease and development. However, iPSC lines have been reported to exhibit substantial transcriptional and epigenetic variability, perhaps driven by stochastic effects during cellular reprogramming or epigenetic memory of the somatic tissue from which they were derived. Here we report the first systematic comparison of the transcriptome of ESCs, iPSCs and their somatic progenitors using high depth sequencing. We performed RNA-seq in a panel of 44 human iPSC lines and somatic tissues obtained from the same four unrelated, healthy subjects. We show that genetic background variation between different individuals is likely to be a major source of transcriptional variation between iPSC lines. In contrast, the transcriptome of iPSCs is largely stable across different tissues of origin and transcriptional memory appears to be uncommon. Our data illustrate that variation between biological replicates of iPSCs derived the same reprogramming event is not dissimilar to that between different passages of adult cells or ESCs. We detect extensive signatures of allele-specific expression at non-imprinted genes in iPSCs and also replicate known expression quantitative trait loci (eQTLs) detected in somatic tissues. Our results suggest that epigenetic and transcriptional variation between iPSCs and ESCs may often reflect underlying genetic differences rather than the effects of cellular reprogramming or residual memory of the tissue of origin. In addition, we clearly show that genetically driven changes in key cellular phenotypes such as transcription are readily detectable using iPSCs as a model system.

186

Using enhancer activity regulatory motifs to explore evolutionary trajectories and disease mechanisms. *L.D. Ward[1,2], W. Meuleman[1,2], P. Kheradpour[1,2], A. Kundaje[1,2], M. Kellis[1,2], Roadmap Epigenomics Mapping Consortium.* 1) CSAIL, Massachusetts Institute of Technology, Cambridge, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA.

The broad range of cell types being studied by the Roadmap Epigenome Mapping Consortium allow us to apply techniques we have previously developed to (a) explore the relative strength of ancient and recent selection on the regulatory programs in different cell types; (b) discover regulatory motifs associated with the tissue specificity of enhancers; and (c) to generate hypotheses about the regulatory mechanisms underlying disease. We used k-means clustering to classify distal enhancers by their cross-tissue activity. These clusters were more enriched for measures of inter- and intra-species sequence constraint than single-tissue annotations; the most mammalian-conserved being a cluster of enhancers strongest in fetal brain, lung, and kidney, and adult brain (14.2% conserved) and a cluster of constitutively poised enhancers active in pancreas, spleen, and gastric tissue (13.9% conserved.) We then looked at derived allele frequencies of SNPs in the 1000 Genomes Phase 1 data to infer the strength of negative selection on the human lineage, improving on our previous methodology to account for ascertainment biases due to read depth. Human constraint was strongest at a cluster of ubiquitous enhancers located near housekeeping genes and a cluster of enhancers active in mesenchymal-derived cells with skin-related gene enrichment. Interestingly, a high derived allele frequency was associated with enhancers with a poised signature in fibroblasts and skeletal muscle and active in fetal and adult brain, near genes annotated as regulating interneuron differentiation and neural tube patterning, suggesting a developmental program targeted by recent positive selection. We then used enrichment analysis to discover putative regulatory motifs that distinguish enhancers belonging to activity clusters. These regulatory motifs allow us to (a) pinpoint signals of selection within enhancers and (b) better dissect disease-associated haplotypes to develop hypotheses about causal motif-altering variants. We have incorporated these maps of "driver" motif instances into the latest version of our online genome annotation tool HaploReg, which allows sets of haplotype blocks from genetic studies to be visualized directly with ENCODE and Roadmap regulatory elements, results from other GWAS and eQTL/meQTL studies, and disruption and creation of regulatory motifs. HaploReg also performs systems-level enrichment analyses on GWAS against these regulatory features.

187

Adjusting Family Relatedness in Data-driven Burden Test of Rare Variants. *Q. Zhang, L. Wang, I.B. Borecki, M.A. Province.* Division od Statistical Genomics, Washington University School of Medicine, St Louis, MO.

Family data represents a rich resource for detecting association between rare variants (RVs) and human complex traits. However, most RV association analysis methods developed in recent years are data-driven burden tests which can adaptively learn weights for individual variants from observed data, but require permutation to estimate significance, thus are not readily applicable to family data because random permutation will destroy family structure. Direct application of these methods to family data will usually result in a significant inflation of false positives. To overcome this issue, we have developed a generalized, weighted sum mixed model (WSMM) and corresponding computational techniques that incorporate family information into data-driven burden tests, and allow appropriate and efficient mixed-model-based permutation tests in family data. Using simulated and real datasets (including the GAW17 simulated data, the FamHS exome chip data, and the LLFS exome sequence data), we demonstrate that the proposed WSMM method can be used to appropriately adjust for dependence among family members and has a good control of type I errors. We also compare WSMM with famSKAT (the family based version of a widely-used, non-data-driven score test method), showing that WSMM has significantly higher power in some cases (without losing power in most other cases). WSMM provides a flexible analysis framework that accommodates arbitrary family structures of any complexity, and it can be easily extended for binary and time-to-onset traits, and combined with different data-driven burden test methods.

188

Fast and accurate pedigree-based imputation from sequenced data in a founder population. *O.E. Livne[1], L. Han[1], G. Alkorta-Aranburu[1], W. Wentworth-Sheilds[1], L.L. Pesce[2,3], C. Ober[1], M. Abney[1], D. Nicolae[1,4,5].* 1) Human Genetics, The University Chicago, Chicago, IL; 2) Department of Pediatrics, The University of Chicago, Chicago, IL; 3) Computation Institute, The University of Chicago and Argonne National Laboratories, Chicago and Argonne, IL; 4) Department of Medicine, Section of Genetic Medicine, The University of Chicago, Chicago, IL; 5) Department of Statistics, The University of Chicago, Chicago, IL.

Despite decreasing DNA sequencing costs, the effects of rare genetic variants on disease risk remain hard to evaluate due to the very large required sample size, which is often prohibitively expensive or impractical to obtain. Founder populations have therefore attracted attention because many rare variants in the general populations rise to higher frequencies due to drift following the bottleneck, providing more power for association studies. Algorithms for phasing and imputation of related individuals exist, yet often fall short of maintaining high accuracy for rare variants. We present a new fast and accurate imputation algorithm that utilizes genome-wide SNP genotypes for 1414 members of the South Dakota Hutterite population, Whole Genome Sequencing (WGS) data from Complete Genomics, Inc. for 98 of those individuals, and pedigree data connecting each of them to all others and to 64 founders. First, phased haplotypes are constructed based on nuclear families and on a hidden Markov model of identity-by-descent (IBD) among the samples. The phased haplotypes are then used to build a complete IBD segment dictionary, indexed by a novel network-based method that allows fast lookup and ensures the consistency of the global IBD structure. We phased >99% of the SNP genotypes, and imputed ~11.6 million bi-allelic variants (SNPs, insertions, deletions) discovered in the WGS data to on average ~77% of the chromosomes of the 1414 individuals. Once IBD segments were indexed, the imputation required only <0.1 second per variant, and a total of 6 node hours on the University of Chicago Beagle supercomputer. Median concordance between imputed and directly genotyped data was >0.995, and was independent of minor allele frequency. We also determined high-confidence IBD-2 segments between pairs of individuals, used to perform a generalized Mendelian error check to assess the WGS data's quality. In those regions, variant calling error rates were lowest for SNPs (0.3%), intermediate for deletions (1.5%), and highest for insertions (52%). Pedigree imputation has other advantages over LD-based imputation, such as inference of the parental origin of haplotypes and the ability to impute ancestors with no available DNA. This work was supported by NIH grants HL085197 and HL21244, and in part by NIH through resources provided by the Computation Institute and the Biological Sciences Division of the University of Chicago and Argonne National Laboratory, under grant S10 RR029030-01.

189

Multiple Genetic Variant Association Testing by Collapsing and Kernel Methods with Pedigree or Population Structured Data. *DJ. Schaid[1], SK. McDonnell[1], JP. Sinnwell[1], SN. Thibodeau[2].* 1) Dept Hlth Sci Res, Mayo Clinic, Rochester, MN; 2) Dept. lab Med Path, Mayo Clinic, Rochester, MN.

Searching for rare genetic variants associated with complex diseases can be facilitated by enriching for diseased carriers of rare variants by sampling cases from pedigrees enriched for disease, possibly with related or unrelated controls. This strategy, however, complicates analyses because of shared genetic ancestry, as well as linkage disequilibrium among genetic markers. To overcome these problems, we developed broad classes of "burden" statistics and kernel statistics, extending commonly used methods for unrelated case-control data to allow for known pedigree relationships, for autosomes and the X chromosome. Furthermore, by replacing pedigree-based genetic correlation matrices with estimates of genetic relationships based on large-scale genomic data, our methods can be used to account for population structured data. By simulations, we show that the Type-I error rates of our developed methods are near the asymptotic nominal levels, allowing rapid computation of p-values. Our simulations also show that a linear weighted kernel statistic is generally more powerful than a weighted "burden" statistic. Because the proposed statistics are rapid to compute, they can be readily used for large-scale screening of the association of genomic sequence data with disease status.

190

Evidence for causality of rare variants based on exact sharing probabilities in affected relatives. *I. Ruczinski[1], A. Bureau[2], MM. Parker[1], MA. Taub[1], ML. Marazita[3], JC. Murray[4], JE. Bailey-Wilson[5], CD. Cropp[5], E. Mangold[6], M. Noethen[6], JB. Hetmanski[1], P. Balakrishnan[1], H. Wang[1], H. Ling[1], AF. Scott[1], TH. Beaty[1].* 1) Johns Hopkins University, Baltimore, MD; 2) Universite Laval, Quebec, Canada; 3) University of Pittsburgh, Pittsburgh, PA; 4) University of Iowa, Iowa City, IA; 5) National Human Genome Research Institute, Baltimore, MD; 6) Universitaet Bonn, Bonn, Germany.

Family based study designs are regaining popularity because large-scale sequencing can help to interrogate the relationship between disease and variants too rare in the population to be detected through any test of association in a conventional case-control study, but may nonetheless co-segregate with disease within families. When only a few affected subjects per family are sequenced, evidence that a rare variant may be causal can be quantified from the probability of sharing alleles by all affected relatives given it was seen in any one family member under the null hypothesis of complete absence of linkage and association. We present a general framework for calculating such sharing probabilities when two or more affected subjects per family are sequenced, and show how information from multiple families can be combined by calculating a p-value as the sum of the probabilities of sharing events as (or more) extreme. We also examine the impact of unknown relationships and propose methods to approximate sharing probabilities based on empirical estimates of kinship between family members obtained from genome-wide marker data. We apply this method to a study of 55 multiplex families with apparent non-syndromic forms of oral clefts from four distinct populations. Whole exome sequencing was performed by the Center for Inherited Disease Research (CIDR) on two or three affected members from each family. The rare single nucleotide variant rs149253049 in the gene ADAMTS9 was shared by affected relatives in three Indian families ($p=2\times10^{-6}$), illustrating the power of this sharing approach.

191

A generalized sparse regression model with adjustment of pedigree structure for variant detection from next generation sequencing data. *S. Cao[1,2], H. Qin[2,3], H. Deng[2,3], Y. Wang[1,2,3].* 1) Department of Biomedical Engineering, Tulane University, New Orleans, LA, USA; 2) Center for Bioinformatics and Genomics, Tulane University, New Orleans, LA, USA; 3) Department of Biostatistics and Bioinformatics, Tulane University, New Orleans, LA, USA.

Complex diseases and traits are likely to be explained by both genotypes (e.g., common and rare genetic variants or SNPs) and environmental factors. Many association methods have been developed for detecting rare or common variants, and usually consider family design and unrelated individual design separately. To overcome the limitations of these methods, we develop a sparse regression model with the adjustment of pedigree structure and the incorporation of prior information. According to the pedigree's impact on continuous phenotypes, we propose a modified Kinship matrix to adjust the correlation between pedigrees. To incorporate prior knowledge, we regularize the model with weighted penalty terms. To get the sparse solution, we evaluate and implement a fast threshold algorithm for solving the regression model with $L_{1/2}$ norm regularization. We also use the smooth gradient algorithm to solve the sparse model penalized with Lp ($0<p<1$) regularization term. After getting the solution path, we use the AIC (Akaike Information Criteria) and stability selection hybrid methods to determine the sparsity level. To evaluate our methods, we compare our method with the single marker test (ψ^2 test), Elastic-net, OMP (i.e., Orthogonal Matching Pursuit) and FOCUSS (i.e., FOcal Underdetermined System Solver). To validate the results, we use the Encyclopedia of DNA Elements (ENCODE) data to simulate the different pedigree structures and test our methods on the Genetic Analysis Workshop 17 and 18 data. The results on both simulation and real data analysis show that our proposed sparse regression models can discover more true causal variants while maintain a lower false discovery rate. In addition, the models tend to detect common and rare variants evenly; the detection of true causal rare variants is not overwhelmed by unrelated common variants. In conclusion, our proposed approach has the following advantages: (i) The model can adjust pedigree structures; (ii) The Lp ($0<p<1$) norm regularization model can yield higher true positive rate while lower false discovery rate than other methods; (iii) The weighted regularization term provides a flexible way to incorporate prior knowledge; (iv) Our model can be easily extended to accommodate environmental covariates.

192

Haplotype phasing across the full spectrum of relatedness. *J. O'Connell[1,2], O. Delaneau[2], N. Pirastu[3], S. Ulivi[4], M. Cocca[5], M. Traglia[5], J. Huang[6], J.E. Huffman[7], I. Rudan[8], R. McQuillan[8], R.M. Fraser[8], H. Campbell[8], O. Polasek[9], C. Hayward[7], A.F. Wright[7], V. Vitart[7], P. Navarro[7], J.F. Zagury[10], J.F. Wilson[8], D. Toniolo[5], P. Gasparani[3], N. Soranzo[6], J. Marchini[1,2].* 1) The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Department of Statistics, University of Oxford, Oxford, United Kingdom; 3) Institute for Maternal and Child Health - IRCCS Burlo Garofolo, University of Trieste, Trieste, Italy; 4) Institute for Maternal and Child Health - IRCCS Burlo Garofolo, Trieste, Italy; 5) Division of Genetics and Cell Biology, San Raffaele Scientific Institute, Milano, Italy; 6) Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 7) MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, Scotland; 8) Centre for Population Health Sciences, University of Edinburgh, Edinburgh, Scotland; 9) Faculty of Medicine, University of Split, Split, Croatia; 10) Laboratoire Génomique, Bioinformatique, et Applications (EA4627), Conservatoire National des Arts et Métiers, Paris, France.

Many existing cohorts contain a range of relatedness between genotyped individuals, either by design or by chance. Haplotype estimation (phasing) in such cohorts is a central step for many downstream analyses. Cohorts sampled from population isolates offer the opportunity for 'long range' phasing of individuals, which involves leveraging recent ancestry between individuals for extremely accurate haplotype inference. This idea was first popularised in a well known paper by Kong et al. (2008) but tractable software for this approach is not available. Extended pedigrees may also be present amongst a wider cohort of unrelated individuals, a Lander-Green algorithm (eg. Merlin) is the traditional method of choice for pedigree phasing but this approach has several limitations. Using genotypes from six cohorts from isolated populations and one cohort from a non-isolated population we have investigated the performance of different phasing methods designed for 'unrelated' individuals. We find that SHAPEIT2 produces much lower switch error rates in all cohorts compared to other methods, especially in identity-by-descent (IBD) sharing regions (less than 0.1% error rate). This occurs because the SHAPEIT2 algorithm implicitly looks for stretches of shared haplotypes between individuals, and can be thought of as a generalization of the long range phasing approach. We show that SHAPEIT2's performance in IBD regions also translates to very accurate phasing for pedigrees. We introduce a novel HMM that can further improve accuracy by integrating family information with the SHAPEIT2 haplotypes, giving us an effective method for dealing with extended pedigrees. The model allows us to accurately detect recombination events in a manner that is robust to genotyping error. We show that our method detects numbers of recombination events that align very well with expectations based on genetic maps whereas Merlin produces inflated recombination rates due to its sensitivity to genotyping error. Our technique even has some ability to detect recombination events in parent-child duos that are not part of a wider pedigree, something that is impossible with a pedigree-only approach. In summary, this work demonstrates methodology for haplotype inference in cohorts with any degree of relatedness that produces haplotypes with unparalleled accuracy.

193

The Theory of Genetic Interactions and its Application to the Problem of Missing Heritability. *A. Young[1], R. Durbin[2].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire.

Twin studies have estimated high heritabilities for complex traits in humans, yet genome wide association studies have failed to identify variants that collectively explain much of the estimated heritability. Insufficient power to detect genetic interactions leaves their contribution largely unknown. It is known, however, that genetic interactions upwardly bias heritability estimators. Quantifying the influence of genetic interactions will therefore lead to a better understanding of the problem of missing heritability. This quantification will become more precise with increasing sample sizes and play an important role in modelling the relation between genotype and phenotype. A theoretical framework is built to provide a basis for quantifying the influence of genetic interactions on phenotypic correlations and heritability estimators. The genetic correlations between individuals in a population isolate are derived in terms of identity-by-descent sharing. The covariance structure of a general trait in a diploid population with an arbitrary ancestral size is derived. This shows the relation between patterns of phenotypic similarity and patterns of identity-by-descent sharing dependent on genetic architecture. Existing heritability estimators are derived as consequences of the general theory. The general theory and the estimators are summarised in an elegant graphical representation. This shows how estimator bias changes dependent on genetic architecture and suggests novel estimators. The theory is illustrated by an application to data from the Northern Finland Birth Cohort. Using identity-by-descent sharing patterns, the covariance structure is fitted to estimate the narrow sense heritability of height, high density lipoprotein levels, and mean cell haemoglobin count in the Northern Finland Birth Cohort.".

194

Tracing individual ancestry in a principal components space. *C. Wang[1], L. Liang[1], G. Abecasis[2], X. Lin[1].* 1) Biostatistics, Harvard School of Public Health, Boston, MA; 2) Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI.

Joint analysis of combined data from multiple sources is an effective approach to increase sample size and statistical power in genetic association studies. Such analysis requires accurate estimation of individual ancestry to adjust for potential population stratification among the combined data. We have previously developed a method that can use small amounts of sequence data to accurately place the ancestry of individuals in a principal components ancestry map generated using a reference set of individuals. Here, we modify the method to also analyze directly genotyped samples. This allows us to place targeted sequenced samples and array-genotyped samples into the same reference ancestry map, facilitating analysis of combined data from targeted sequencing and array-genotyping experiments. We apply these methods to estimate worldwide continental ancestry and fine-scale ancestry within Europe using the Human Genome Diversity Project (HGDP) and Population Reference Panel (POPRES) as the reference panels. Our results show that ~1,000 random SNPs can lead to accurate estimation of continental ancestry while ~20,000 random SNPs are required for accurate estimation of ancestry within Europe. Further, we examine two custom arrays, the ExomeChip and MetaboChip. We show that for samples genotyped on these two arrays, we can accurately estimate their continental ancestry using the HGDP data as reference. However, estimating the fine-scale ancestry within Europe is difficult, partly due to the small number of overlapping markers between these two arrays and the POPRES data. To address this problem, future studies should consider developing a densely genotyped reference panel of diverse European populations.

195

Multiple HLA loci and energy metabolism genes are targeted by recent positive selection in an Ethiopian population. *F. Tekola-Ayele[1], A. Adeyemo[1], E. Hailu[2], A. Aseffa[2], G. Davey[3], M.J. Newport[3], CN. Rotimi[1].* 1) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Armauer Hansen Research Institute, Addis Ababa, Ethiopia; 3) Brighton and Sussex Medical School, Brighton, UK.

The genetic, population demographic, ecological, and socio-cultural diversity of African populations is a useful resource to capture genomic regions targeted by specific local selection pressures, and to refine and interpret the putative selective forces. However, only a handful of African population groups have been represented in studies involving genome-wide scans of selection. Here, we performed a genome-wide scan to identify regions that show evidence of recent positive selection in a sample of 120 individuals from the Wolaita ethnic group from southern Ethiopia. The Wolaita people have inhabited the mid- and high-lands of southern Ethiopia for millennia and inhabit one of the most densely populated regions in the country. To identify selection signal in this ethnic group, we included 988 unrelated individuals from 11 global populations in the HapMap 3 database. We performed three types of analyses: the integrated haplotype score (iHS), cross-population extended haplotype homozygosity (XP-EHH), and fixation index (FST). We found enrichment of signals of recent positive selection in several genes in the HLA locus ($p<0.00001$). Our selection signals overlapped with loci that showed genome-wide significant association with several inflammatory diseases including podoconiosis, a devastating lymphedema of the lower legs resulting from massive inflammatory response to chronic exposure of feet to irritant soils in poor barefoot individuals. Podoconiosis is endemic in the Wolaita area. Second, we found selection signals in energy metabolism genes including PPAR-alpha that plays a key adaptive role in energy metabolism during prolonged food deprivation. This finding is consistent with the historical consumption of "enset" (Ensete ventricosum) as a main food source in the Wolaita area. "Enset", also known as "The Tree against Hunger", has high carbohydrate and starch but low fat and protein content, and is used as a staple and co-staple food source. Therefore, the putatively advantageous selected allele may inhibit PPAP-alpha expression, resulting in reduced oxidation of lipids which is balanced by enhanced carbohydrate oxidation. We also found a novel selection signal in CDKAL1, a gene known to be associated with type 2 diabetes. In all, we found evidence of strong selection for HLA and metabolism genes in an Ethiopian population probably due to adaptive response to burden of pathogens and nutritional composition of the staple diet.

196

Statistical model for the joint estimation of mRNA isoforms and individual-specific expression from RNA-seq data. *F. Mordelet[1], B.E. Engelhardt[1,2,3].* 1) Institute for Genome Science and Policy, Duke University, DURHAM, NC; 2) Department of Biostatistics & Bioinformatics, Duke University, Durham, NC; 3) Department of Statistical Science, Duke University, Durham, NC.

There are around 21,000 transcribed genes in the human genome that encode somewhere between 250,000 and one million proteins in human cells. A gene in its pre-mRNA form consists of a chain of introns and exons. Ultimately, only exons are transcribed into mRNA, and introns are removed or`spliced''; often one or several exons get spliced as well in a variety of different ways. Through this alternative splicing mechanism, the same gene may produce many different protein sequences, with often with distinct biological functions. Those different products of the same exon sequences are called isoforms. Along with the biological complexity added by alternative splicing comes a role in many human complex traits and disease, including cancer and HIV. In order to detect differentially spliced genes and genetic variants that regulate isoform transcription, it is crucial to correctly estimate individual-specific expression levels of each isoform. To achieve this, we propose a Bayesian nonparametric statistical framework to model RNA-sequencing data. This model, called a Hierarchical Dirichlet Process, views sequencing reads from an individual as random observations from a multinomial distribution associated with an unobserved isoform. Our model has the novel properties that i) it does not require the specification of the number of isoforms a priori but estimates this from the data; ii) it allows sharing of information across individuals, which is useful since isoforms may be shared across individuals; iii) it does not assume a uniform rate of transcription across isoform sites, where current models use a Poisson-based distribution with uniform rate assumptions. RNA-seq studies have shown the existence of sequence-specific and position-specific read count biases, causing the baseline distribution of reads to be considerably non-uniform along the isoform sequence. We use fast Markov chain Monte Carlo (MCMC) methods to jointly estimate both the unobserved read assignments to isoforms and individual-specific isoform levels. This allows us to efficiently map sequenced reads to isoforms, learning at the same time the set of isoforms that are expressed for a given gene and how much each isoform is expressed in each individual. These fitted models are then used to identify the set of isoforms across individuals for a particular cell type, test for differential expression of isoforms, and identify genetic variants associated with individual-specific isoform levels.

197

Choosing an RNA-seq Aligner for QTL and ASE Analysis in the Genotype-Tissue Expression Project. *D.S. DeLuca[1], T. Lappalainen[2], P. Kheradpour[1], M. Sammeth[3], J. Monlong[3], P. Ribeca[4], E. Palumbo[4], A. Battle[5], E. Gelfand[1], R. Guigo[3], K. Ardlie[1], G. Getz[1], The GTEx Consortium.* 1) The GTEx Project, The Broad Institute, Cambridge, MA; 2) University of Geneva, Department of Genetic Medicine and Development, Genève, Switzerland; 3) Universitat Pompeu Fabra, Center for Genomic Regulation, Barcelona, Spain; 4) Centro Nacional de Análisis Genómico, CNAG, Barcelona, Spain; 5) Stanford University, Daphne Koller Group, Stanford, CA.

For many applications, the first step in analyzing RNA sequence data is to align reads to a reference genome. The alignment step is fundamentally important because it directly impacts all downstream characterizations such as expression, splicing events and allele specific expression. RNA-seq has reached a critical stage where widespread adoption has resulted in the availability of a selection of alignment tools, but has not yet matured to the point of convergence on an established gold standard. As a result, investigators have a range of alignment options but little indication as to how the choice will affect their analyses. For the Genotyping-Tissue Expression project (GTEx), we have developed a series of criteria with which to compare alternative aligners that will reflect the performance in the project's targeted goals. A primary goal of GTEx is to create a public atlas for human gene expression and its regulation, to enable discovery of expression quantitative trait loci (eQTL) and establish associations with disease. In the project's pilot phase, GTEx typed 190 postmortem human donors from which 1814 total tissues (from 47 separate tissue sites) were profiled by RNA-seq to a median depth of 80 million aligned reads. We wanted to use the aligner that would allow for maximum discovery of eQTLs, splice-QTLs and allele specific expression. We chose to compare the relatively established Tophat aligner with the more recently developed GEM package. Initial metrics indicated that GEM produces a greater number of alignments, begging the question as to whether these additional alignments are providing biological signal or noise. We demonstrate that these additional reads do correlate strongly with expected expression levels, exhibit distributions consistent with biological expectations, and provide additional power for allele specific expression. Given the full panel of criteria, we conclude that while both aligners perform well overall, GEM does exhibit some advantages. Because the strategy of alignment comparison established here is broadly applicable, we expect that this analysis will provide a path forward to improved decision making when choosing an aligner for eQTL and related analysis.

198

mRNA and small RNA sequencing of 465 HapMap cell lines: the feasibility of multicenter RNA-seq studies. *P.A.C. Hoen[1], M.R. Friedlander[2], J. Almlof[3], M. Sammeth[2,4], I. Pulyakhina[1], S.Y. Anvar[1,5], J.F.J. Laros[1,5], O. Karlberg[3], J.T. den Dunnen[1,5], G.J.B. van Ommen[1], I.G. Gut[4], R. Guigo[2], X. Estivill[2], A.C. Syvanen[3], E.T. Dermitzakis[6,7,8], T. Lappalainen[6,7,8], GEUVADIS consortium.* 1) Department of Human Genetics, Leiden University Medical Center, Leiden, Netherlands; 2) Centre for Genomic Regulation (CRG), Barcelona, Spain; 3) Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 4) Centro Nacional de Analisis Genomico (CNAG), Barcelona, Spain; 5) Leiden Genome Technology Center, Leiden University Medical Center, Leiden, the Netherlands; 6) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 7) Institute for Genetics and Genomics in Geneva (iG3), University of Geneva, Geneva, Switzerland; 8) Swiss Institute of Bioinformatics, Geneva, Switzerland.

RNA-sequencing is an increasingly popular technology for genome-wide analysis of transcript structure and abundance. However, the sources of technical and inter-laboratory variation have not been assessed in a systematic manner. To address this, seven centers of the GEUVADIS consortium sequenced mRNAs and small RNAs of 465 HapMap lymphoblastoid cell lines (LCLs) for which the full genome sequence was available from the 1000Genomes consortium. Five samples were sequenced in every center and 168 samples were sequenced in two centers. When comparing individual LCLs, the biological variation is limited. Nevertheless, the five samples that were sequenced in each laboratory clustered by sample and not by laboratory. The clustering by sample was much stronger for exon quantifications than for transcript quantifications. When investigated further, laboratory differences mainly manifested in the average GC-percentage, the width of the distribution of GC-percentages and the insert sizes. A similar analysis was performed for small RNA sequencing. Again, the replicates sequenced in all laboratories grouped by samples rather than laboratories. Clustering divided the samples into those dominated by miRNA and those dominated by rRNA. The proportions of miRNA and rRNA reads were more similar within samples than within laboratories. The miRNA contents clearly varied between RNA extraction batches. Therefore, differences in relative miRNA/rRNA contents are likely introduced during RNA isolation, before the samples were distributed across the laboratories. The heterogeneity in small RNA contents did not bias the relative quantification of individual miRNAs. In conclusion, distributed RNA-sequencing appears to be feasible. It is particularly attractive for large population-based and cross-biobank studies, where sequencings costs and sample logistics may require combination of data from individual studies and laboratories. The combined sequencing data from this project significantly extended our understanding of the genetic basis of transcriptome variation and generated an unprecedented resource of genomic variants affecting expression (eQTLs), splicing, and transcription start site and polyadenylation site usage.

199

Complete resequencing of extended genomic regions using fosmid targeting and PacBio's Single Molecule Real-Time (SMRT®) long-read sequencing technology. *D.E. Geraghty[1], C.W. Pyo[1], K. Wang[1], R. Wang[1], Y.S. Pyon[1], K. Eng[2], B. Bowman[2], S. Ranade[2].* 1) Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA., United States of America; 2) Pacific Biosciences of California Inc., Menlo Park, CA, United States of America.

A limitation of genome-wide methods for identifying genetic variation is the inability to acquire phased, extended, and complete genomic sequence from targeted regions. In that regard, we have substantially improved methods for both sample preparation and single-molecule, long-sequence-read generation that allow for complete, haplotype-resolved resequencing across extended genomic subregions. As a specific application, we have targeted subregions of HLA in case and control chromosomes with a major focus on Type 1 Diabetes (T1D). Despite decades of research, the causative genetic factors in the MHC that contribute to T1D susceptibility have not been completely and unambiguously identified and it is likely that some of the relevant genetic variants are yet to be discovered. Evidently, the most certain way to identify them is to resequence the conserved portions from cases and controls, with the goal of testing the relatively simple hypothesis that susceptibility loci lie within the MHC and specifically within the conserved extended haplotypes (CEHs) of the MHC that are associated with disease. Towards that end, we are completing a pilot project resequencing 800 kb segments that include the DR4 CEH using next-generation sequencing methods for both targeted DNA acquisition and for sequencing. Three of the four DR4 haplotypes resequenced are from T1D patients, and one from a control individual and all data are to be complete and phased over each 800 kb segment. The approach developed and data acquired demonstrate cost-effective linear scale-up, supporting feasibility of extending the analysis to several hundred cases and controls generating phased chromosomal genomic sequences of ~ 800 kb that encompass the extent of relevant CEHs.

200

Platinum Genomes: A systematic assessment of variant accuracy using a large family pedigree. *M.A. Eberle[1], M. Kallberg[2], H.-Y. Chuang[2], P. Tedder[1], S. Humphray[1], D. Bentley[1], E.H. Margulies[1].* 1) Scientific Research, Illumina Cambridge, Ltd, Saffron Walden, Essex, United Kingdom; 2) Dept Bioinformatics, Illumina, Inc, San Diego, CA., USA.

As next-generation sequencing technologies become widely adopted for clinical applications, it is extremely important that we have the ability to systematically assess the accuracy of variant calls generated from these data. However, at present, no such "truth" dataset of variant calls exists for a diploid or cancer genome. Instead, we have relied on measures of concordance with calls from alternative technologies (such as Sanger sequencing or microarrays), or by testing for inheritance errors using parent-parent-child trios as a proxy for sensitivity/accuracy measures. Both approaches have limitations that preclude us from measuring the true accuracy of sequencing technologies and variant calling algorithms. We have initiated a project to systematically identify all variation in a large three-generation family (the CEPH/Utah pedigree 1463). Both the raw sequence data and variant calls are being made publicly available. There are several key features of our initial approach to generating an extremely high confidence set of variant calls: First, all sequence data have been generated with the latest PCR-free techniques and sequenced to a higher-than-usual depth (~50x) to maximize sensitivity in low-coverage regions. Second, we have determined the haplotype inheritance structure and used this information to boost sensitivity to detect errors. Third, several variant calling algorithms have been used to leverage joint calling approaches and maximize the detection of a broad set of SNVs and indels.

To illustrate the increased sensitivity of error/accuracy calculations when multiple siblings are analyzed in parallel we have analyzed the single-nucleotide polymorphisms (SNPs) and indels within the parents and eleven offspring of this family. Based on this we have identified over 4.7M SNPs and 640K indels that we predict are correctly genotyped across the parents and 11 siblings corresponding to an additional ~360K SNPs and ~95K indels per sample compared against a normal quality-filtered call set. For the variants that show Mendelian conflicts in the pedigree we have identified that the majority are related to cell line mutations including ~2000 cell line de novo SNPs per sample and large cell line deletions. We will present this method and assess the role that de novo cell line mutations and alignment errors play in deviations from Mendelian inheritance.

201

Sensitive and Quantitative Measurement of Nuclease Mediated Genome Editing at Human Endogenous Loci using SMRT Sequencing. *A. Hendel[1], E. Kildebeck[1], E. Fine[2], J. Clark[1], G. Bao[2], M. Porteus[1].* 1) Department of Pediatrics, Stanford University, Stanford, CA; 2) Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta.

Targeted genome editing with engineered nucleases allows researchers to introduce precise sequence modifications at almost any site within the genome. These genome modification strategies are initiated by a double-strand break created by an engineered nuclease (e.g. zinc finger nuclease, TAL effector nuclease, or an RNA-guided endonuclease). The precise genome modification, however depends on whether the break is repaired in a mutagenic fashion by non-homologous end-joining (NHEJ) or using a provided donor DNA molecule by homologous recombination (HR). If the break is repaired by NHEJ, small insertions or deletions are created at the site of the break. If the break is repaired by HR, however, defined sequence changes, can be introduced at the site of the break creating nucleotide specific modifications to the genome. An important aspect is to develop sensitive and quantitative methods to measure these two different outcomes at human endogenous loci. While there are different reporter based assays to measure these two events, our work is the first that allows simultaneous measurement of both events at endogenous loci without having to use reporter genes. Because of the large size of donor DNA repair templates used in genome editing, there has not been a sequencing platform capable of providing a comprehensive picture of how frequently NHEJ is occurring alongside targeted HR. We now show that the long read lengths afforded by Single Molecule, Real-Time (SMRT) sequencing provide the platform needed to simultaneously measure genome editing outcomes by either mutagenic NHEJ or by HR at endogenous human loci. Using this method, we were able to analyze the frequency of multiple genome editing outcomes simultaneously at an endogenous locus in human cell lines with a detection sensitivity of ~ 0.1%. We used this method to investigate the variables critical to increasing targeted genome editing, determine conditions that maximize HR while minimizing mutagenic NHEJ, and measure genome editing in difficult to target hematopoietic stem and progenitor cells. In sum, our new method to quantify genome editing outcomes has significant advantage over prior methods, because it does not require making a stable reporter cell line, it measures outcomes at endogenous loci, it is sensitive to an unprecedented level, and it can be utilized in cell types for which creating reporter cell lines is simply not possible including human hematopoietic stem and progenitor cells.

202

Mining genomic feature sets and identifying significant biological relationships with BedTools2. *A. Quinlan, N. Kindlon.* Center for Public Health Genomics, University of Virginia, Charlottesville, VA.

Modern DNA sequencing technologies are enabling unprecedented explorations of the spectrum of functional elements in diverse cell types. The fundamental result of such large projects is a complex, multi-dimensional collection of "signals" such as ChIP-seq peaks, DNA methylation sites, and RNA-seq measurements, that are scattered throughout the genomes of hundreds of different cell types. While these datasets are crucial to gaining insight insight into genome regulation in the context of human disease, even basic analyses pose substantial computational and statistical challenges. The datasets are large, complex, and employ myriad file formats. Moreover, revealing new biological relationships such as co-associated regulatory elements depends upon choosing a relevant statistical metric. The inherent analytical complexity, computational burden, and debate about the choice of appropriate statistics motivated us to develop a standardized analysis toolkit for the genomics community.

Building upon our widely used bedtools genomic analysis software, we have developed bedtools2, a scalable toolkit for mining genomic feature sets and identifying significant biological relationships among them. We have completely re-engineered the core algorithms in bedtools2 to scale to analyses involving hundreds of datasets described in any common genomics file format (e.g., BAM, BED, VCF).

In the context of biological discovery, the most exciting new functionality in bedtools2 is a comprehensive set of statistical measures for revealing associations between sets of genomic features (e.g., do these transcription factors co-associate more than expected by chance?). Here we present the new statistical tests in bedtools2, compare our tests to existing approaches such as the ENCODE Genome Structure Correction (GSC) metric, and provide needed insights into which metrics are most appropriate to common biological questions. We demonstrate typical misuses of these metrics and illustrate how our tests and associated visualization tools can reveal new biological insights. Given the speed and analytical flexibility of bedtools2, we anticipate that our new toolkit will be an invaluable resource for geneticists studying the impact of genetic variation and regulatory elements on human disease phenotypes.

203

Creating a Single Haplotype Human Genome Assembly. *T. Graves[1], W. Warren[1], B. Fulton[1], K. Meltz Steinberg[1], R. Agarwala[2], V. Schneider[2], D. Church[2], E. Eichler[3], R. Wilson[1].* 1) Washington University School of Medicine, Saint Louis, MO; 2) National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Building 38A, 8600 Rockville Pike, Bethesda, MD; 3) University of Washington, Genome Sciences, Seattle, WA.

The human genome reference sequence has provided a foundation for studies of genome structure, human variation, evolutionary biology and human disease. Many of these studies have also revealed, however, that there are regions of the human reference genome that are not represented optimally. At the time the reference genome was completed it was clear that there were some loci recalcitrant to closure with the technology and resources available at that time. It was not clear, however, the degree to which structural variation and diversity affected our ability to produce a truly representative genome sequence at these loci. Many of these regions in the genome, particularly the structural variant loci, are often associated with repetitive sequences. In order to discriminate between repeat copy and allelic copies, the sequence from a single haplotype across these regions is necessary. To this end, we have utilized a hydatidiform mole source, CHM1 to finish highly complex, repetitive regions to high quality. Our aim is to develop a single allelic representation of the entire human genome, the 'platinum' reference. In order to achieve this, we have generated ~100X whole genome shotgun sequence as Illumina paired end data, as well as over 450 BAC sequences from the CHM1 libraries. The whole genome data has been assembled using a reference-guided assembly and the finished BAC sequences have been incorporated into this assembly. We have compared the CHM1 assembly to the current reference, GRCh37 to identify single nucleotide variants, structural variants, and missing sequence from the reference. In addition, we have aligned the CHM1 Illumina sequence to the CHM1 assembly to evaluate the efficacy of our assembly strategy.

204

A generalized human reference as a graph of genomic variation. *E. Garrison[1], D. Kural[1,2], A. Ward[1], W.P. Lee[1], G. Marth[1].* 1) Biology, Boston College, Chestnut Hill, MA; 2) Seven Bridges Genomics, Cambridge, MA.

The linear reference genome provides a straightforward basis for analysis, but this convenience also limits the ability of researchers to understand complex forms of genomic variation. The short reads used in resequencing studies must be mapped to the single haplotype of the reference genome, generating ascertainment bias towards small variants that are unlikely to disrupt read placement, such as SNPs and short indels. Consequently, the detection of more complex divergences from the reference genome, such as longer indels, structural variants, and clustered variants requires large expenditures in sequencing and analysis costs.

Much of genomic variation in humans is shared, and thus the haplotypes detected in many individuals can be pooled into a combined reference containing the vast majority of variation likely to be found in a newly-sequenced sample. To generate this combined representation of genomic variation, we propose a graph genome reference (GGR). Nodes in the GGR are sequences observed in the population and edges represent possible linkages between adjacent sequences. The haplotypes of individuals in a population are thus a subset of possible paths through this graph, a property which allows the use of the graph as a tool to reduce bias when detecting known sequence variants via resequencing.

Here, we describe the application of this structure to variant detection. We have developed a method to align short reads to sequence graphs, which we use to collect evidence for putative or known variants in a variety of analytical contexts. By realigning short sequence reads to a GGR of putative variants, we produce an ensemble variant detector capable of coherently integrating signals from a wide array of resequencing and assembly-based variant detection approaches. Our method enables the accurate characterization of contexts in which variants overlap or are embedded in other variants. The application of haplotype-based variant detection to reads aligned to a GGR allows the determination of physical linkage of variants using primary sequencing observations. Using a GGR of known, high-confidence variants as a basis for mapping provides the benefits of multi-sample variant detection without requiring the centralized analysis of raw sequencing data.

205

Clinical experience implementing chromosomal microarray analysis (CMA) in a clinical psychiatric practice for adults with autism spectrum disorders (ASD) and related neurodevelopmental disorders. *K.B. Teed[1], A. Vahabzadeh[2], J.C. Cubells[1,2].* 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 2) Emory Autism Center, Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA.

The number of children diagnosed with autism spectrum disorders (ASD) began increasing sharply in the early 1990s, and the first wave of this ballooning demographic is now entering adulthood. While the number of adults with ASD is growing rapidly, clinical guidelines for application of CMA and other genomic technologies in evaluation and management of ASD remain focused largely on pediatric populations. CMA is now routinely used as a first-tier test for work-up of ASD in pediatric settings, but is not always considered for adults in psychiatric or other behavioral-health settings. We propose that CMA should be a routinely considered test for adults with ASD. To support this contention, we present data and clinical experience from more than 6 years of CMA testing at an academic tertiary care center for adults with ASD and related neurodevelopmental disorders. A total of 45 patients were tested from 2007 to 2013, chosen based on clinical judgment and availability of insurance to cover testing. Fifteen (33%) were found to have either suspected pathogenic or established pathogenic copy number variants (CNVs). The abnormal findings included previously reported ASD-associated CNVs, such as one case of 15q13.1-13.3 duplication, and one case of 9q34 deletion (Kleefstra syndrome), as well as a variety of previously unreported chromosomal differences. The molecular findings altered specific management recommendations in one case, in which discovery of a deletion disrupting *MAOA* and *MAOB* on the X chromosome of a male patient led to specific dietary and pharmacologic restrictions aimed at minimizing exposure to tyramine and other sympathomimetic compounds. The observations in this clinical series, which will be presented in detail, suggest that adult patients with ASD and related neurodevelopmental disorders can benefit from CMA. We hypothesize that such benefits will increase as our knowledge about genes and recurrent CNV continues to expand.

206

New insights into the spectrum of pathogenic variation in epilepsy gained from molecular diagnostic testing of 1600 individuals. *S. Aradhya, E. Butler, D. McKnight, A. Shanmugham, C. Downtain, A. Entezam, G. Richard.* GeneDx, 207 Perry Pkwy, Gaithersburg, MD.

Determining the etiology of epilepsy, which affects ~1% of the population, is challenging but can have important implications for treatment, prognosis and recurrence risk counseling. We performed molecular testing for 1,600 individuals using next-generation sequencing and copy number analysis for 53 genes most often associated with major forms of primary and syndromic epilepsy. A known or predicted pathogenic variant was observed in 261 individuals (16%), underscoring the role of genetic factors in disorders of neuronal excitability. The highest molecular diagnostic rate (21%) was obtained for a subset of genes associated with infantile-onset epilepsy, including infantile epileptic encephalopathy or spasms. 6% of positive findings were gene deletions or duplications, emphasizing that copy number analysis increases the diagnostic yield. 478 individuals (30%) had a variant of unknown significance (VUS), including 169 who only had one heterozygous VUS in an autosomal recessive gene, most likely unrelated to the phenotype or representing carrier status. The remainder had likely benign or no reportable variants. Our data showed a pronounced incidence of autism and/or intellectual disability not only in individuals with mutations in MECP2, STXBP1, SCN1A, and CDKL5, but also in those with variants in KCNQ2, SLC2A1, SCN2A, SPTAN1, and GABRG2. Although mutations in ion channel genes are a common cause of epilepsy, half of all pathogenic findings in our cohort were in non-ion channel genes. For example, 23% of individuals had mutations associated with disorders in the Rett/Angelman syndrome spectrum. We also found pathogenic mutations associated with rare forms of epilepsy, such as pyridoxine-dependent epilepsy and others, for which a molecular diagnosis can guide treatment. The vast majority of individuals with positive findings had definitive pathogenic variants in only one gene, suggesting that genetic forms of epilepsy are mostly monogenic. However, 35% of these individuals also had one or more VUS, which may contribute to the phenotype, and modifier effects due to variants in other genes not included in the assay may also exist. Our data provide new insight into the mutation spectrum and frequency of pathogenic variants for specific epilepsy subtypes, broaden the genotype-phenotype correlations for many genes studied in this cohort, and illustrate the utility of a multi-gene testing strategy.

207

Pathogenic rare copy number variants in community-based schizophrenia suggest a potential role for clinical microarrays. *A.S. Bassett[1,9], G.A. Costain[1], A. Lionel[2,3], D. Merico[2], P.J. Forsythe[4], K. Russell[1], C. Lowther[1], T. Yuen[1], J. Husted[5], D.J. Stavropoulos[6,7], M. Speevak[8], E.W.C. Chow[1,9], C.R. Marshall[2,3], S.W. Scherer[2,3].* 1) Clinical Genetics Research Program, Centre for Addiction and Mental Health, Toronto, Ontario M5S 2S1, Canada; 2) The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario M5G 1L7, Canada; 3) Department of Molecular Genetics and the McLaughlin Centre, University of Toronto, Toronto, Ontario M5G 1L7, Canada; 4) Horizon Health and Dalhousie University, Saint John, New Brunswick E2L 4L4, Canada; 5) Health Studies, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada; 6) Cytogenetics Laboratory, Department of Pediatric Laboratory Medicine, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada; 7) Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario M5S 1A1, Canada; 8) Department of Genetics, Credit Valley Hospital, Mississauga, Ontario L5M 2N1, Canada; 9) Department of Psychiatry, University of Toronto, Toronto, Ontario M5T 1R8, Canada.

Individually rare, large copy number variants (CNVs) contribute to genetic vulnerability for schizophrenia. Unresolved questions remain, however, regarding the anticipated yield of clinical microarray testing in schizophrenia. Using high resolution genome-wide microarrays and rigorous methods, we investigated rare CNVs in a prospectively recruited community-based cohort of 459 unrelated adults with schizophrenia and estimated the minimum prevalence of clinically significant CNVs that would be detectable on a clinical microarray. Blinded review by two independent clinical cytogenetic laboratory directors of all large (>500 kb) rare CNVs in cases and well-matched controls showed that those deemed clinically significant were highly enriched in schizophrenia (16.4-fold increase, $p<0.0001$). In a single community catchment area, the prevalence of individuals with these CNVs was 8.1%. Rare 1.7 Mb CNVs at 2q13 were found to be significantly associated with schizophrenia for the first time, compared with the prevalence in 23,838 population-based controls (42.9-fold increase, $p=0.0002$). Additional novel findings that will facilitate the future clinical interpretation of smaller CNVs in schizophrenia include: (i) a greater proportion of individuals with two or more rare exonic CNVs >10 kb in size (1.5-fold increase, $p=0.0134$) in schizophrenia; (ii) the systematic discovery of new candidate genes for schizophrenia; and (iii) functional gene enrichment mapping highlighting a differential impact in schizophrenia of rare exonic deletions involving diverse functions, including neurodevelopmental and synaptic processes (4.7-fold increase, $p=0.0060$). These findings suggest consideration of a potential role for clinical microarray testing in schizophrenia, as is now the suggested standard of care for related developmental disorders like autism.

208

Genetic assessment of congenital brain malformations. *U. Hehr[1,2], T. Rödl[1], S.M. Herbst[1,2], S. Schirmer[1], T. Geis[3], B. Kasper[4], G. Schuierer[5], J. Winkler[6], G. Uyanik[7].* 1) Center for Human Genetics, Regensburg, Germany; 2) Department of Human Genetics, University of Regensburg, Regensburg, Germany; 3) Department of Pediatrics, University Hospital Regensburg, Regensburg, Germany; 4) Department of Neurology, University Hospital Erlangen, Erlangen, Germany; 5) Division of Neuroradiology, University Hospital Regensburg, Regensburg, Germany; 6) Division of Molecular Neurology, University Hospital Erlangen, Erlangen, Germany; 7) Center for Medical Genetics, Hanusch Hospital, Vienna, Austria.

Congenital brain malformations are a clinically and genetically important cause of early developmental delay, cognitive impairment and seizures. They are identified and classified based on specific structural cerebral abnormalities observed in cerebral imaging. We report our genetic results for a cohort of more than 950 independent patients with holoprosencephaly (HPE) and neuronal migration disorders (periventricular nodular heterotopia, subcortical band heterotopia, classical lissencephaly, polymicrogyria and cobblestone lissencephaly) over 13 years. Genetic testing was individually assigned considering available clinical information including family history as well as cerebral imaging. Individual testing strategies included linkage analysis for suitable families, CNV analysis by MLPA, Sanger sequencing and most recently the introduction of massive parallel sequencing, allowing the identification of the underlying genetic alterations for approximately 25% of analyzed samples. In the HPE cohort mutations were identified in SHH (16 patients), SIX3 (10), ZIC2 (14), TGIF (3), Gli2 (4). In patients with neuronal migration disorders identified mutations affected LIS1 (30), DCX (57), ARX (13), TUBA1A (5), TUBB2B (3), GPR56 (8), FLNA (32), POMT1 (16), POMGnT1 (15), FKTN (4), FKRP (7), ISPD (1), LARGE (2). Genotype phenotype correlations in particular were observed for DCX, ARX and POMT1, where truncating mutations more frequently resulted in more severe clinical manifestations. Genetic findings support the differential diagnosis, prognostic evaluation as well as genetic counseling including the option of prenatal genetic testing or even preimplantation genetic diagnosis. Our data draw special attention to autosomal dominant or X-linked genetic forms, associated with a high recurrence risk for further offspring and transmitted by apparently "healthy" parents. At closer examination a substantial proportion of heterozygous mutation carriers for X-linked conditions associated with mutations in FLNA, ARX or DCX themselves presents with a wide range of clinical manifestations, which may significantly impair social interactions as well as the perception of health and genetic risks and are currently further addressed in an ongoing study. Genome wide genetic testing strategies will further improve the number of identified genes and genetic alterations, but continue to require their critical interpretation in the context of clinical findings and cerebral imaging.

209

Investigation of CASK gene aberrations in 38 patients with severe intellectual disability, microcephaly and disproportionate pontine and cerebellar hypoplasia (MICPCH). *S. Hayashi[1], O. Nobuhiko[2], J. Takanashi[3], J. Inazawa[1].* 1) Dept Molec Cytogenetics, Tokyo Med & Dental Univ, Tokyo, Japan; 2) Department of Planning and Research, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan; 3) Department of Pediatrics, Kameda Medical Center, Chiba, Japan.

The CASK gene [OMIM: *300172] at Xp11.4, 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 it was reported to play several roles in neural development and synaptic function. While loss of function of CASK raised by mutation or genomic copy-number variant (CNV) causes intellectual disability and microcephaly with pontine and cerebellar hypoplasia (MICPCH) [OMIM: #300749] in females, insufficiency of CASK probably leads to lethality in males. We reported a first case of MICPCH with heterozygous deletion at Xp11.4p11.3 including CASK in 2008, and thereafter we have recruited patients presenting with MICPCH in order to investigate CASK aberrations. So far, we have detected various types of CASK aberrations in 24 of 38 patients with MICPCH: large deletions in 6 patients, intragenic duplication or complex rearrangement in 3 patients and point mutations in 15 patients. Through this screening, we were able to detect other factors probably responsible for MICPCH, for example, a mosaic mutation of CASK in a male MICPCH patient, another type of genomic aberration involving in CASK or a deletion of another candidate gene. These results seem to suggest a novel etiology for MICPCH, thus we are currently examining a correlation between these genotypes and phenotypes. Moreover, we are also screening other candidate genes for CASK-negative MICPCH patients by target re-sequencing to detect another locus of MICPCH. Our research demonstrates a comprehensive etiology of MICPCH.

210

A homozygous PDE6D mutation in Joubert syndrome impairs targeting of farnesylated INPP5E protein to the primary cilium. *S. Thomas[1,2], K. J. Wright[3], S. Le Corre[4], A. Micalizzi[5,6], M. Romani[5], A. Abhyankar[7], J. Saada[8], I. Perrault[1,2], J. Amiel[1,2,9], J. Litzler[9], E. Filhol[2,10], N. Elkhartoufi[2,10], M. Kwong[3], J.L. Casanova[2,7,11], N. Boddaert[2,12], W. Baehr[13], S. Lyonnet[1,2,9], A. Munnich[1,2,9], L. Burglen[14], N. Chassaing[15], F. Encha-Ravazi[1,2,9], M. Vekemans[1,2,9], J. G. Gleeson[16], E.M. Valente[5], P. K. Jackson[3], I. A. Drummond[4,17], S. Saunier[2,11], T. Attié-Bitach[1,2,9]*. 1) INSERM U781, Paris, France; 2) Université Paris Descartes, Paris Sorbonne, France; 3) Genentech Inc., South San Francisco, California 94080, USA; 4) Nephrology Division, Massachusetts General Hospital, Boston, MA 02129; 5) Mendel Laboratory, Istituto di Ricovero e Cura a Carattere Scientifico Casa Sollievo della Sofferenza San Giovanni Rotondo, Italy; 6) Department of Medical and Surgical Paediatric Sciences, University of Messina, Messina, Italy; 7) St. Giles Laboratory of Human Genetics of Infectious Diseases, The Rockefeller University, New York; 8) Service de Gynécologie obstétrique, Hôpital Antoine-Béclère, Assistance Publique - Hôpitaux de Paris (AP-HP), Clamart, France; 9) Département de Génétique, Hôpital Necker-Enfants Malades, AP-HP, Paris, France; 10) INSERM U983, Hôpital Necker-Enfants Malades, Paris, France; 11) Laboratory of Human Genetics of Infectious Diseases INSERM U980, Necker Medical School, Paris, France; 12) - Service de radiologie Pédiatrique, Hôpital Necker-Enfants Malades, AP-HP, Paris, France; 13) University of Utah Health Science Center, Salt Lake City, UT 84132; 14) AP-HP, Hôpital Trousseau, Centre de référence des malformations et maladies congénitales du cervelet et Service de génétique, Paris, 75012, France; 15) Service de génétique médicale, CHU de Toulouse ; EA-4555 UPSIII, Toulouse, France; 16) Neurogenetics Laboratory, Institute for Genomic Medicine, Department of Neurosciences and Pediatrics, Howard Hughes Medical Institute, University of California, San Diego, California, USA; 17) Department of Genetics, Harvard Medical School, Boston, MA 02115.

Joubert syndrome (JS) is characterized by a distinctive cerebellar structural defect, namely the « molar tooth sign ». JS is genetically heterogeneous, involving 18 genes identified to date, which are all required for cilia biogenesis and/or function. In a consanguineous family with JS associated with optic nerve coloboma, kidney hypoplasia and polydactyly, combined exome sequencing and mapping identified a homozygous splice site mutation in PDE6D, encoding a prenyl-binding protein. We found that pde6d depletion in zebrafish leads to renal and retinal developmental anomalies and wild-type but not mutant PDE6D is able to rescue this phenotype. Proteomic analysis identified INPP5E, whose mutations also lead to JS or MORM syndromes, as novel prenyl-dependent cargo of PDE6D. Mutant PDE6D shows reduced binding to INPP5E, which fails to localize to primary cilia in patient fibroblasts and tissues. Furthermore, mutant PDE6D is unable to bind to GTP-bound ARL3, which acts as a cargo-release factor for PDE6D-bound INPP5E. Altogether, these results indicate that PDE6D is required for INPP5E ciliary targeting and suggest a broader role for PDE6D in targeting other prenylated proteins to the cilia. This study identifies PDE6D as a novel JS disease gene and provides the first evidence of prenyl-binding dependent trafficking in ciliopathies.

211

Assessment of incidental findings in whole-exome sequences from the Baylor-Hopkins Center for Mendelian Genomics. *J. Jurgens[1], N. Sobreira[1], H. Ling[2], E. Pugh[2], E. Cirulli[3], F. Schiettecatte[4], K. Doheny[2], A. Hamosh[1], D. Valle[1]*. 1) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Center for Inherited Disease Research (CIDR), Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Center for Human Genome Variation, Duke University School of Medicine, Durham, NC; 4) FS Consulting, Salem, MA.

As WES and WGS are increasingly utilized, the need to assess the significance of incidental variants is ever more pressing. The ACMG recently published a list of 57 genes implicated in common monogenic disorders with the recommendation that pathogenic and likely pathogenic variants in these genes detected by clinical sequencing be reported. As an initial step in determining the practical consequences of this recommendation, we searched for variants in these genes in the WES data of 55 unrelated probands sequenced in the Baylor-Hopkins Center for Mendelian Genomics. We identified nonsynonymous exonic and splicing variants and indels with MAF < 1% in the 1000 genomes, EVS, and our internal control databases and determined if any of these variants had been described in the HGMD, Emory or ClinVar databases. Our analysis generated a list of 119 variants in 36 (63%) of the 57 genes that passed IGV inspection analysis: 99 missense; 3 nonsense; 15 nonframeshifting indels; and 2 frameshifting indels. 59 variants (50%) were novel. There was a mean of 2.16 variants per individual (range 0 -7); 5 probands had no variants. Of the 119 variants, 44 (37.0%) were listed in HGMD with the prediction that 31 were disease mutations (DM), 10 were uncertain disease mutations (DM?), 2 were disease polymorphisms (DP), and 1 was likely benign (FP). Of these same 44 variants, 2 were in the Emory (with discordant classifications), and 10 were in ClinVar (5 with discordant classifications). 4 variants were listed only in ClinVar. Of the 36 variant-containing genes on the ACMG list, *APOB*, *BRCA2*, *CACNA1S*, *DSP*, *PCSK9*, and *RYR1* had the most variants (range 7-12/gene) in our 55 probands. Among 1189 controls sequenced at Duke, these genes were found to have rare (<0.5%) functional variants in 2.8%, 29.75%, 21.35%, 33.3%, 12.89%, and 56.34% of the individuals, respectively. Our analysis shows frequent variation in at least some of the 57 genes on the ACMG list with the majority being novel missense mutations not classified in any of the 3 reference databases. Some of these genes are highly mutated in controls. Our result indicates the challenge of assessing the phenotypic consequences of these incidental variants. Analysis of additional WES/WGS data will make interpretation easier, but a single database where variants are uniformly classified with phenotypic information and ongoing review is urgently needed for better interpretation of variants and more accurate counseling of patients.

212

Deep Sequencing in Extended Pedigrees Reveals a Major Rare Non-Synonymous Variant Influencing the *De Novo* Ceramide Synthesis Pathway. *J.E. Curran[1], P.J. Meikle[2], J.M. Weir[2], J.B. Jowett[2], T.M. Teslovich[3], G. Jun[3], S. Kumar[1], M. Almeida[1], J.M. Peralta[1], C. Fuchsberger[3], A.R. Wood[4], A. Manning[5], T.M. Frayling[4], P. Gingolani[6], R. Sladek[7], D.M. Lehman[8], G. Abecasis[3], M.C. Mahaney[1], T.D. Dyer[1], L. Almasy[1], R. Duggirala[1], J. Blangero[1], T2D-GENES Consortium.* 1) Texas Biomedical Research Institute, San Antonio, TX; 2) Baker IDI Heart and Diabetes Institute, Melbourne, Vic, AU; 3) University of Michigan, Ann Arbor, MI; 4) University of Exeter, Exeter, UK; 5) Broad Institute, Boston, MA; 6) McGill University, Montreal, CA; 7) Montreal Diabetes Research Institute, Montreal, CA; 8) University of Texas Health Science Center at San Antonio, San Antonio, TX.

Lipids play critical roles in many cellular processes; and a major role in health and disease. The biologically simple nature of lipid species suggests their determinants may reside closer to the action of genes than the classical lipid measures. Thus their potential for a simpler causal web makes them valuable endophenotypes for identifying genes in lipid metabolism. Using targeted lipidomic profiles and whole genome sequence (WGS) in 605 Mexican Americans from large pedigrees, we searched for causal variants influencing the *de novo* ceramide synthesis pathway. This pathway plays a role in obesity, diabetes and CVD. The core of this pathway is the generation of biologically active ceramide from its inactive precursor, dihydroceramide. We focused our initial search on ~9,000 missense and nonsense mutations predicted to be highly deleterious by PolyPhen-2 (score ≥ 0.80). We identified a rare functional variant (L175Q) in the *DEGS1* gene on chr 1, observing 17 total copies in our pedigrees; it was only seen in 3 individuals in the 1000 Genomes Project. This rare variant is significantly associated with an increase in dihydroceramide levels ($p=1.57\times10^{-6}$) by a biologically significant amount of 1.2 standard deviation units (SDU). Pleiotropy analyses revealed that this increase in dihydroceramide is accompanied by significant decreases in both ceramide (p=0.027, -0.45 SDU) and spingolipids including sphinogmyelin (p=0.0009, -0.75 SDU) and trihexosylceramide (p=0.003, -0.70 SDU). It was also associated with a significant decrease in waist/hip ratio (p=0.014, -0.45 SDU) and cholesterol esters (p=0.020, -0.53 SDU). *DEGS1* is the fatty acid desaturase responsible for the desaturation of dihydroceramide to generate ceramide *de novo*. Studies have shown that silencing *DEGS1* in cells focused on specific processes such as metabolic stress, oxidative stress, apoptosis and cancer, results in a decrease in ceramide and sphingomyelin synthesis, while increasing cell levels of dihydroceramide and sphinganine species. Additionally, *DEGS1* knockdown results in down-regulation in almost all metabolic biosynthesis pathways, including cell-cycle growth. Our association results support these findings in that the L175Q variant results in a significant increase in dihydroceramide levels and a significant reduction in biologically active ceramide/sphingomyelin levels. Given the diverse role of ceramide in cellular processes, *DEGS1* has excellent therapeutic potential for many diseases.

213

From embryonic lethal to no phenotype: What autozygome can teach us about loss of function in the human genome. *F.S. Alkuraya.* Genetics, KFSHRC, Riyadh, Riyadh, Saudi Arabia.

Autozygous inheritance of loss of function alleles in the human genome has greatly contributed to our ability to annotate the human genome in a clinically relevant manner. Tracking down these alleles by virtue of their existence within the autozygome has been key to this process both before and after the advent of next-generation sequencing. However, the advent of next-generation sequencing made it possible to fully unlock the potential of the autozygome in the context of unmasking phenotypic aspects that have historically evaded the application of classical autozygosity mapping, particularly at the extremes of the phenotypic spectrum. On one end of the spectrum is embryonic lethality, which we know from mouse studies is likely to be caused in some instances by biallelic loss of function of genes but these genes were difficult to identify in humans. We show here that careful selection of cases can maximize the power of combined exome/autozygome analysis to map novel human embryonic lethal genes. The other end of the phenotypic spectrum is when no phenotype can be assigned to the biallelic loss of function of some genes. Such genes may be truly dispensable as in genes that are on their way of turning into pseudogenes, dispensable in a context-dependent manner as in genes that are only required for the metabolism of certain substrates that are no widely available in contemporary human diet, or may be associated with subtle phenotypes that we are unable to quantify such as genes that contribute to the external appearance. We show that exome sequencing of well-phenotyped individuals enriched for the size of autozygome i.e. offspring of first cousin parents, can be an effective way of mapping such genes particularly when the loss of function alleles are very rare. Thus, our data demonstrate that next-generation sequencing can unlock the full potential of the autozygome to add to the global effort to assign clinically relevant information to the human genome by expanding the repertoire of phenotypes that can be interrogated.

214

The evolutionary dynamics of regulatory DNA in the mouse and human genomes. *J. Vierstra[1], E. Rynes[1], R. Sandstrom[1], R.E. Thurman[1], J.A. Stamatoyannopoulos[1,2].* 1) Genome Sciences, University of Washington, Seattle, WA, 98195; 2) Department of Medicine, Division of Oncology, University of Washington, Seattle, WA, 98195.

We used DNaseI hypersensitivity mapping to systematically delineate mouse regulatory DNA across 44 diverse cell types and primary tissues. In total we identified >1.3 million mouse DNaseI hypersensitive sites (DHSs), of which >450,000 are also detected in human tissues at orthologous sequence positions. Of these, ~50% show tissue-selective activity patterns similar to their human counterparts. We find that this small regulatory compartment is densely populated by virtually known classes of lineage-specifying transcriptional regulators. However, when compared with the mouse genome, the human genome has undergone extensive rewiring of its cis-regulatory architecture along several axes, including sequence remodeling of individual regulatory DNA regions and changes in their patterns of tissue selectivity. The transcription factor composition of individual regulatory DNA elements has undergone extensive turnover, with <15% of transcription factor recognition sequences positionally constrained, corresponding to ~30% of regulatory elements with functionally conserved chromatin accessibility. These changes coincide with pervasive 'repurposing' (lineage switching) of regulatory DNA, due chiefly to evolutionary innovation of specific transcription factor recognition sequences that drive tissue specificity. Analysis of species-specific DHSs in both mouse and human exposes a potential critical role for repetitive DNA in evolution of regulatory DNA. We find that ~50% of species-specific regulatory DNA has arisen due to the massive expansion of repetitive elements that these expansions have occurred asymmetrically in mouse and human. Furthermore, these repeat expanded regulatory regions contain the recognition sequences for many transcriptional regulators that in many cases account for up to 20% of the total binding sites within the accessible genome. Taken together, our findings delineate a core mammalian regulon and provide extensive insights into the genesis, extinction, and evolutionary perpetuation of mammalian regulatory DNA.

215

Widespread exonic transcription factor binding directs codon usage and protein evolution. *A.B. Stergachis[1], E. Haugen[1], A. Shafer[1], W. Fu[1], B. Vernot[1], J.M. Akey[1], J.A. Stamatoyannopoulos[1,2].* 1) Genome Sciences, University of Washington, Seattle, WA; 2) Department of Medicine, University of Washington, Seattle, WA.

The human genome contains two codes that have long been assumed to operate independently of one another -- the genetic code that specifies the sequence of amino acids in a protein, and a regulatory code that specifies recognition sites for >1,000 sequence-specific transcription factors (TFs) that collectively control gene expression. We used genomic DNaseI footprinting to systematically map transcription factor occupancy at nucleotide resolution across the human exome in 81 diverse cell types. Here we show that >14% of codons in human exons simultaneously specify both amino acids and regulatory information in the form of transcription factor recognition sites. Such dual-use codons ('duons') are highly evolutionarily conserved, and exhibit systematic constraint of both degenerate and non-degenerate codon positions that is directly attributable to overlying binding of a sequence-specific TF. This constraint has widely impacted codon choice, and acts as the major driver of codon usage biases in the human genome. Duons have also widely impacted protein evolution by constraining possible nonsynonymous changes. We show further that the genetic code has reciprocally affected the regulatory code, which is selectively depleted of recognition sites with the potential to recognize (and therefore ectopically introduce) stop codons. Finally, we show that at least 17% of human coding variants (including synonymous, nonsynonymous, and disease-associated variants) that lie within duons directly impact overlying TF binding. In summary, our results show that transcription factors have systematically shaped human codon choice and protein evolution, and that interpretation of genetic variation within coding sequence must account for overlying regulatory codes.

216

Short tandem repeat polymorphisms create an abundant source of expression variability. *M. Gymrek[1,2,3,4], S. Georgiev[5], B. Markus[1], J. Chen[2], P. Villarreal[1], J. Pritchard[5,6], Y. Erlich[1].* 1) Whitehead Institute, Cambridge, MA; 2) Harvard-MIT Division of Health Sciences and Technology, MIT, Cambridge, MA; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 4) Department of Molecular Biology and Diabetes Unit, Massachusetts General Hospital, Boston, MA; 5) Department of Human Genetics, University of Chicago, Chicago, IL; 6) Howard Hughes Medical Institute, Chevy Chase, MD.

A central goal in genomics is to elucidate the genetic architecture of complex traits. So far, efforts to discover eQTLs (expression Quantitative Trait Loci) have been mainly focused on the contribution of SNPs and CNVs to gene expression. A few dozen single gene studies in human and model organisms have suggested that short tandem repeat (STR) variations can modulate expression of nearby transcripts. Here, we report the first genome-wide survey to identify STRs that affect gene expression. We analyzed STR variations across hundreds of samples from the 1000 Genomes Project across nine populations using a custom algorithm and performed association tests with the expression levels of nearby transcripts. This process identified significant associations between STR variations and expression profiles (eSTRs) of more than 3,000 genes. These associations were replicated across populations and on orthogonal expression assays (RNA-sequencing and expression array). Predicted eSTRs are found near transcription start sites, are strongly enriched in annotated promoter and enhancer elements, and fall into regions predicted by chromatin marks to be active regulatory regions. Moreover, fine-mapping techniques indicate that the eSTR association signals are unlikely to stem from tagging nearby SNPs or other variations in linkage disequilibrium. These loci may provide a novel set of regulatory variants that can help explain the genetic architecture of gene expression and contribute to the heritability of complex traits.

217

RNA-DNA Sequence Differences Occur within Seconds following RNA exit PolII Active Sites and Are Responsive to Cellular Stress. *V.G. Cheung[1,3,4], I.X. Wang[3], L. Core[2], H. Kwak[2], L. Brady[5], A. Bruzel[1,3], A.L. Richards[5], M. Wu[1,3], J.T. Lis[2].* 1) Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA; 2) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA; 3) Department of Genetics, University of Pennsylvania, Philadelphia, PA 19104, USA; 4) Department of Pediatrics, University of Pennsylvania, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 5) Cell and Molecular Biology Graduate Program, University of Pennsylvania, Philadelphia, PA 19104, USA.

RNA and protein sequences are expected to be identical to the corresponding DNA sequences. However, there are known exceptions such as ADAR mediated RNA editing, and other types of RNA-DNA differences, RDDs, which we and others have reported. To understand how these RDDs arise and their biological implications, we are studying the mechanisms by which the RDDs are formed and their role in cellular responses to stress.

First, we isolated nascent RNAs by PRO-seq and GRO-seq which are nuclear run-on assays followed by deep sequencing. We compared the sequences of these RNA with their corresponding DNA sequences and found all 12 types of RDDs, including transversions. We validated the presence of RDDs in nascent transcripts by examining chromatin-associated RNAs. The single nucleotide resolution of PRO-seq allows us to determine that RDDs are formed about 50 to 60 bases from the RNA polymerase II active sites. They are made after 5' capping and before splicing. These RDDs are found throughout the transcripts including the untranslated regions (3' and 5' UTRs), antisense RNAs associated with transcription start sites, introns and exons. The exonic sites are exported into the cytoplasm and translated into proteins.

Second, to study the biological effects of RNA editing and RDDs, we found that their levels change in response to cellular stress. For example, the A-to-G editing levels in the SEC genes required for vesicle formation in the endoplasmic reticulum (ER) change following ER stress. In addition, ARAP1 that was found in genetic studies of type II diabetes to be significantly associated with fasting proinsulin level showed only modest changes in gene expression in response to insulin; however, the RDD level in ARAP1 increased significantly. Together these results show that RNA editing and RDDs play a previously unrecognized role in the maintenance of cellular homeostasis.

In this presentation, I will describe our findings on when during transcription RDDs are formed and how RDD levels change in response to cellular stress.

218

RNA-seq transcriptome profiling uncovers how structural variants influence alternative splicing. *E. Ait Yahya Graison[1], A. Necsulea[2], A. Reymond[1].* 1) CIG, University of Lausanne, Lausanne, Switzerland; 2) Swiss Institute for Experimental Cancer Research (ISREC) Ecole Polytechnique Fédérale de Lausanne (EPFL) CH-1015 Lausanne Switzerland.

Structural variants (SVs) are a major contributor to genetic diversity but a comprehensive understanding of their functional impact remains to be established. We previously showed that they impact tissue transcriptome by modifying the level and timing of expression of genes that localize within and on their flanks. Here we used mouse inbred strains as a model to extensively gauge by RNA-seq the influence of structural variants on the transcriptome complexity and regulation. Towards this aim, we generated extensive RNA-seq data from liver and brain of C3H/HeJ, 129S2, DBA/2J and C57BL/6J strains using the Illumina HiSeq2000 and Genome Analyzer IIx. These high-throughput sequencing expression data were intersected with the catalog of structural variants encompassing insertions, inversions, deletions, copy number gains and complex SVs produced within the frame of the Mouse Genomes project to assess simultaneously the impact of genome structural changes on both gene expression and alternative splicing. While large SVs (mostly copy number gains or deletions) have a direct impact on transcript expression levels as embedded gene generally show a direct correlation between number of copies and expression levels, smaller SVs significantly influence splicing diversity in several manners. First, the presence of either a deletion or an insertion in an exon significantly favors the emergence of alternative splice donor sites. Second and conversely, a deletion or insertion within an intron significantly increases the number of alternative acceptor sites in downstream exon. This effect is independent of maintenance or disruption of the open reading-frame. Third, the splicing machinery more rarely skips out the SV-containing exons. Fourth, on the contrary, exons that lie just upstream or downstream of an SV-containing intron are more likely to be skipped out by splicing. This latter effect is not dependent on the size of the SV suggesting that it is a true property of the rearrangement rather than the consequence of an increased or decreased size of the flanking intron. To conclude , we show that SVs do impact tissue transcriptome on a global scale also by altering its complexity and diversity through alternative splicing.

219

Exploring regulatory and loss-of-function variation in personalized multi-tissue transcriptomes using allele-specific expression. *T. Lappalainen[1,2,3,4], M.A. Rivas[5], M. Lek[6,7], M. Pirinen[5,8], J. Maller[6,7], K. Kukurba[1], E. Tsang[9], D. DeLuca[7], M. Sammeth[10,11,12], . Geuvadis Consortium[2,11], M.I. McCarthy[5], C.D. Bustamante[1], S.B. Montgomery[1,13], K. Ardlie[7], D.G. MacArthur[6,7], E.T. Dermitzakis[2,3,4], GTEx Consortium.* 1) Department of Genetics, Stanford University, Stanford, CA; 2) Department of Genetic Medicine and Development, University of Geneva Medical School, Switzerland; 3) Institute for Genetics and Genomics in Geneva (1G3), University of Geneva, Switzerland; 4) Swiss Institute of Bioinformatics, Geneva, Switzerland; 5) Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 6) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 7) Broad Institute of Harvard and MIT, Cambridge, MA; 8) Institute for Molecular Medicine Finland, University of Helsinki, Finland; 9) Program in Biomedical Informatics, Stanford University School of Medicine, Stanford, CA; 10) Centro Nacional d'Anàlisi Genòmica, Barcelona, Spain; 11) Center for Genomic Regulation CRG, Barcelona, Spain; 12) National Center for Scientific Computation, Petropolis, Rio de Janeiro, Brazil; 13) Department of Pathology, Stanford University, Stanford, CA.

Analysis of genetic regulatory variation is advancing towards comprehensive analysis of common and rare regulatory and loss-of-function variants across a wide variety of tissues. A powerful approach to characterize these effects both in populations and at the level of an individual is analysis of allelic imbalance between the two haplotypes of an individual. In this study, we analyzed allele-specific expression (ASE) and transcript structure (ASTS) in 1582 samples of the GTEx pilot data from 171 individuals and 45 tissues with RNA-seq, genotype and partial exome-seq data. First, we measured the activity of eQTLs discovered from 9 major tissues across all the 45 tissues, and quantified how well these eQTLs capture systemic regulatory effects and assigned the best eQTL proxy for diverse tissues. Quantifying common and rare regulatory variation genome-wide with ASE analysis shows that 9% of effects in one tissue are captured by analyzing another tissue. However, the low sharing is mostly due to tissue-specificity of gene expression, as allelic imbalance alone is highly shared (median 43%) between different tissues within an individual, and ASTS analysis captures a similar pattern for splicing variation. We quantified transcriptome effects of a total of 11,518 putative loss-of function variants - stop-gained SNPs, splice variants, and frameshift indels - from both GTEx and the Geuvadis project (RNA-seq data of 462 individuals from 1000 Genomes). We predict the likelihood for a stop-gained variant to trigger nonsense-mediated decay based on its properties, which will be crucial for interpreting loss-of-function variants discovered in future studies. Additionally, as much as 22% of variants show variable NMD between tissues. Taking population-level assessment of functional variants to the level of an individual is a challenge in personalized genetics applications. With ASE analysis we uncover substantial variation in eQTL effects between individuals, suggesting that modifiers of common variants are widespread even at the cellular level. Furthermore, transcriptome sequencing of carriers of rare putative loss-of-function variants allows interpretation of their functional effects across different tissues. Altogether, this study demonstrates the power of transcriptome sequencing to understand systemic effects of functional variation and to interpret personalized genomes.

220

Analysis of the Genetic Variation and Age Interplay on Gene Expression Using RNA-seq Data from Multiple Tissues. *A. Viñuela[1], M.N. Davies[1], A. Buil[2,3,4], A.A. Brown[5], H.F. Zheng[6], J.B. Richards[1,6], K.S. Small[1], R. Durbin[5], E.T. Dermitzakis[2,3,4], T.D. Spector[1].* 1) Dep. of Twin Research & Genetic Epidemiology, King's College London, London, United Kingdom; 2) Dep. Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 3) Institute of Genetics and Genomics in Geneva, University of Geneva, Geneva, Switzerland; 4) Swiss Institute of Bioinformatics, Switzerland; 5) Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 6) Dep. of Medicine, Human Genetics, Epidemiology and Biostatistics McGill University, Canada.

Global gene expression becomes noisier with age, but it is not clear whether this is due to changes in the genetic architecture of regulation (gene-environment interactions, GxE) or to environmental/stochastic factors. In particular, tissue specificity of age-related sources of variation in expression is largely unknown. We analysed RNA-seq data from adipose, skin, whole blood, and lymphoblastoid cell lines (LCLs) samples from ~800 female adult twins from the TwinsUK cohort (39-85 years old). Using this data, we first identified genes with an age-related component in expression. By looking at discordance and heritability we then investigated whether these changes were due to GxE or environmental factors. Finally, we investigated genetic variants whose effect changes with age (GxA). Analysing multiple tissues allows us to contrast the rate and nature of the ageing process in different tissues and environments. Analysis of gene expression identified 1172 (adipose), 3748 (skin), 458 (blood), and 22 (LCLs) genes whose expression was associated with age (FDR < 0.05). Of those, 398 genes were associated in two tissues; with 2 genes (*PDE4D* and *SCL9A9*) significantly associated in skin, adipose and blood. This suggests tissue specificity for the ageing process and for genetic regulation. We are currently investigating differential splicing events with age, another aspect of gene expression that may become deregulated with age. Changes in transcriptional noise with age (variance in gene expression) were identified in 16 genes in adipose (e.g: *BCL6* or *PRL1*) and 61 in skin (e.g: *MUC1* or *SOD2*). Of those, 14% (adipose) and 16% (skin) increased in variance with age. To understand the sources of variation we exploited the twin design by looking at changes in level of discordance in expression within MZ pairs. Within skin, 129 genes showed altered discordance in expression with age (BH corrected p-value < 0.05). Since MZ twins are genetically identical, discordance in expression points to environmental or GxA effects on expression. We are currently investigating changes in heritability to characterize the changing contributions of genotype and environment to variation in gene expression with age. Finally, we are performing a genome-wide scan of genetic variants with GxA interactions affecting expression (gxa-eQTLs) to identify relevant genes for aging. These interactions provide concrete examples of how genetic control of expression is modified over time.

221

Transcriptomes of individual cells. *C. Borel, P.G. Ferreira, E. Falconnet, P. Ribaux, S.E. Antonarakis, E.T. Dermitzakis.* Dept Genetic Medicine, Univ Geneva Medical Sch, Geneva, Switzerland.

We sequenced hundreds of single-cell transcriptomes to decipher significant and stochastic cell-to-cell transcriptional variation. Here, we aim to uncover the extent of cell specific alternative splicing, allele specific gene expression (ASE) and the dynamic of gene expression in the transcriptome of individual cells. Starting from a homogeneous cell population of human female primary fibroblasts from one individual, we used the C1 Single Cell Auto Prep System to capture individual cells and to generate high quality of individual pre-amplified cDNA for next-generation sequencing. Single-cell mRNA-seq librairies were deep sequenced (PE, 100bp) on Illumina HiSeq 2000 sequencer. On average, we obtained 30 millions of reads per single cell, of which more than 60% mapped uniquely to GENCODE annotated exons with GEM aligner. In total, ~ 130,000 transcripts were detected per single cell (RPKM>=0.5), representing ~40% of the total transcripts expressed by the bulk sample containing millions of cells. First, we noted a wide spectrum of transcriptional heterogeneity. Although most of the same transcripts are expressed in all individual cells, we observed a signature of expressed genes that are cell specific. The analysis of their biological function by gene ontology analysis revealed a clustering into different biological processes. Second, we identified cell-specific novel exons, multitude of alternative spliced isoforms and 3'UTR isoforms due to alternative polyadenylation. To further assess the differential allelic expression at the single cell level, whole genome sequencing has been performed on this sample. We are currently identifying genome-wide the number of transcripts with detectable SNPs displaying differential ASE. We are expecting that genes with the highest allelic imbalance are located on the X-chromosome. The data also provide a comprehensive survey of X inactivation and escape. Insight gained from this study is the unprecedented understanding of genetic variability and gene expression at single-cell level. Single cell transcriptomic is likely to become an important tool in Genetics, Cell Biology, Development, Immunity and Cancer. S.E.A and E.T.D. laboratories contributed equally.

222

A low-frequency variant in a lincRNA doubles the risk of pneumococcal bacteraemia in Kenyan children. *A. Rautanen[1], M. Pirinen[1], T.C. Mills[1], S.J. Chapman[1], V. Naranbhai[1], J.A. Scott[2,3,4], T.N. Williams[2,3,4], P. Donnelly[1,5], A.V.S. Hill[1,6], C.A. Spencer[1], The Wellcome Trust Case Control Consortium 2.* 1) WTCHG, University of Oxford, Oxford, United Kingdom; 2) Kenya Medical Research Institute, Wellcome Trust Research Programme, Kilifi, Kenya; 3) Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom; 4) INDEPTH Network, Accra, Ghana; 5) Department of Statistics, University of Oxford, Oxford, United Kingdom; 6) The Jenner Institute, University of Oxford, Old Road Campus Research Building, Oxford, United Kingdom.

Bacteraemia (bacterial bloodstream infection) is a major cause of illness and death globally, and especially in sub-Saharan Africa, but little is known about the genetic basis of individual susceptibility. We conducted a genome-wide association study of bacteraemia susceptibility using 893,635 genotyped autosomal markers in 1536 children with blood culture-proven bacteraemia and 2677 healthy infants living in Kilifi District in Kenya. We performed whole-genome imputation resulting in nearly 11 million SNPs that passed our QC filters. These were analysed both with bacteraemia overall and with the largest subphenotype, bacteraemia caused by Streptococcus pneumoniae (pneumococcus). We verified the genotypes at the most associating SNPs by direct genotyping, and replicated the results in a further 434 cases and 1366 controls. This confirmed a previously reported strong association in the HBB locus including the sickle cell disease-causing polymorphism rs334 (P discovery = 4.09×10^{-11}; P replication = 7.7×10^{-4}); the association was driven by a strong homozygote risk (P combined = 2.66×10^{-12}; OR = 4.93, 95% CI = 3.15-7.70) and relatively modest heterozygote protection (P combined = 4.67×10^{-3}; OR = 0.77, 95% CI = 0.65-0.92). More importantly, we identified a novel association in a lincRNA gene (P discovery = 3.58×10^{-7}; P replication = 1.16×10^{-3}; P combined = 1.69×10^{-9}; OR combined = 2.47, 95% CI = 1.84-3.31) with pneumococcal bacteraemia. Although the function of this gene is unknown, it has been reported to be expressed solely in placenta and white blood cells. We further studied its mRNA expression in different white blood cell types, and detected expression exclusively in neutrophils that are known to be important in killing the invading pneumococcus. Interestingly the susceptibility allele is derived rather than ancestral, occurs at low frequency (2.7% in controls and 6.4% in cases), and shows evidence for variation in the susceptibility it confers to different bacterial infections. Understanding the molecular mechanisms leading to the doubled risk of pneumococcal bacteraemia associated with this allele could provide new clues in the pressing search for new therapeutic targets. The importance of lincRNAs as key regulators of gene expression has only recently been recognised and our finding will further focus research on the role of lincRNAs in human disease.

223

Identifying multiple causative genes at a single GWAS locus. *M. Flister[1,2], S. Tsaih[1], B. Endres[1,2], A. Geurts[1,2], J. Lazar[1,3], M. Dwinell[1,2], C. Moreno[1], H. Jacob[1,4].* 1) Human and Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, WI; 2) Department of Physiology, Medical College of Wisconsin, Milwaukee, WI; 3) Department of Dermatology, Medical College of Wisconsin, Milwaukee, WI; 4) Department of Pediatrics, Medical College of Wisconsin, Milwaukee, WI.

Background: Genome-wide association studies (GWAS) can nominate candidate genes, but are largely unable to establish disease causality or differentiate between variants in linkage disequilibrium (LD). Moreover, some GWAS loci might contain multiple causative variants that contribute to the overall disease susceptibility at a single locus. However, most GWAS lack the statistical power to detect multigenic interactions, prompting us to adopt an alternative approach to testing multiple GWAS genes empirically. Here we present a new strategy using inbred rat strains to rapidly dissect the (+), (-), or (=) phenotypic effects of multiple genes at a single hypertension GWAS locus. **Methods:** *Agtrap*, *Mthfr*, *Clcn6*, *Nppa*, *Nppb*, and *Plod1* were individually mutated in the Dahl SS rat by zinc finger nuclease injections (ZFN). ZFN-mutated and wild type (WT) control SS rats (n=8-25 per group) were assessed for BP and renal damage after 10 days on a 4% NaCl diet. Human haplotypes were functionally annotated using HapMap, 1000 Genomes, and ENCODE data. **Results:** Compared with WT, *Nppa* (Δ+27mmHg, P<0.001) and *Plod1* (Δ+17mmHg, P<0.01) mutations increased BP in response to high salt diet (4% NaCl), whereas *Clcn6* mutation decreased BP (Δ-22mmHg, P<0.001). Compared with WT (91±5 mg/day), *Plod1* (147±12 mg/day, P<0.001) and *Mthfr* (132±23 mg/day, P<0.05) mutations increased proteinuria, whereas *Clcn6* (56±5 mg/day, P<0.05) and *Agtrap* (54±9 mg/day, P<0.05) mutation decreased protein excretion. Using HapMap and 1000 Genomes, we identified 226 SNPs in LD (r2<0.6) with disease associated haplotype blocks. Out of 9 disease-associated haplotypes, 6 had multiple SNPs in LD that caused nonsynonymous changes or were associated with differential expression of multiple genes, demonstrating that genetic interactions within the *AGTRAP-PLOD1* locus were not only possible, but rather were highly plausible. **Conclusions:** Using rat and human data, we were able to draw previously unforeseen conclusions. Foremost, that multiple genes at a single GWAS locus can influence clinically relevant phenotypes. We present the first evidence that *Nppa*, *Clcn6*, *Mthfr*, *Plod1*, and *Agtrap* mutations cause divergent CVD phenotypes and some have the ability to modify renal phenotypes independently of BP. Combined, these data suggest that the current "best-fit" candidate gene interpretations of GWAS are in some cases only a simplified view of the complex genetic architecture underlying individually associated loci.

224

Gene Silencing and Haploinsufficiency of Csk in GWAS Locus 15q24 Increase Blood Pressure. *B. Oh[1,2,3], H. Lee[2], S. Ji[2], S. Park[2], M. Kim[3], B. Jigden[1], J. Lim[1], Y. Lee[4].* 1) Biomedical Engineering, Sch Med, Kyung Hee Univ, Seoul, South Korea; 2) Department of Biomedical Science, Graduate School, Kyung Hee University, Seoul, Korea; 3) Department of Medicine, Graduate School, Kyung Hee University, Seoul, Korea; 4) Department of Physiology, College of Medicine, BK 21 Project for Medical Sciences, Yonsei University, Seoul, Korea.

Recent genomewide association studies have identified 33 genetic loci that influence blood pressure. The 15q24 locus is one such locus that has been confirmed in Asians and Europeans. There are 21 genes in the locus within a 1-Mb boundary, but a functional link of these genes to blood pressure has not been reported. We aimed to identify a causative gene in the 15q24 locus by siRNA in vivo delivery in mice. *CYP1A1*, *CYP1A2, CSK*, and *ULK3* were selected as candidate genes, based on their functions and proximity to the lead SNP (rs1378942). The siRNA for each gene was injected into mouse tail vein. Of the 4 genes, *Csk* and *Cyp1a2* siRNA reduced their mRNA levels in injected mice. Thus, the blood pressure of *Csk* and *Cyp1a2* siRNA-injected mice was measured in the carotid artery. Whereas *Cyp1a2* siRNA did not change blood pressure, *Csk* siRNA increased it. Further, blood pressure in $Csk^{+/-}$ heterozygotes was higher than in wild-type, consistent with what we observed in *Csk* siRNA-injected mice. These results suggest that *CSK* is a causative gene in the 15q24 locus. After *ATP2B1* in 12q21, *CSK* is the second gene in a blood pressure GWAS locus to be identified as a causative gene. *CSK*, c-src tyrosine kinase, is a negative regulator of *SRC* that is involved in angiogenesis and the *CSK* knockout embryo has an abnormal vascular network. This study may provide a new genetic pathway that underlies the development of hypertension and a novel therapeutic target for the treatment of hypertension.

225

A novel genetic basis for systemic vasculitis: Systemic and cutaneous Polyarteritis Nodosa (PAN) are caused by recessive mutations in an immune-related gene. *R. Segel[1], S.B. Pierce[2], P. Elkan-Navon[1], T. Walsh[2], S. Padeh[3], J. Barash[4], A. Zlotogorski[5], Y.Y. Berkun[5], J.J. Press[6], M. Mukamel[6], P.J. Hashkes[1], E. Ling[7], L.L. Harel[6], M. Tekin[8,9], F. Yalcinkaya[8], O. Kasapcopur[10], E.F. Emirogullari[9], M.K. Lee[2], R.E. Klevit[2], P.F. Renbaum[1], A. Weinberg-Shukron[1], S. Zeligson[1], D. Marek-Yagel[3], M. Shohat[6], A. Singer[11], E. Pras[3], A.A. Rubinow[5], Y. Anikster[3], M.C. King[2], E. Levy-Lahad[1].* 1) Medical Genetics Institute, Shaare Zedek Medical center, Jerusalem, Israel; 2) University of Washington, Seattle, United States; 3) Sheba Medical Center, Tel Hashomer, Israel; 4) Kaplan Medical Center, Rehovoth, Israel; 5) Hadassah Medical Center, Jerusalem, Israel; 6) Schneider Children's Medical Center, Petach Tikvah, Israel; 7) Soroka Medical Center, Beer Sheva, Israel; 8) Ankara University, Ankara, Turkey; 9) University of Miami, Miami, United States; 10) Istanbul University, Istanbul, Turkey; 11) Barzilai Medical Center, Ashkelon, Israel.

Background: Polyarteritis nodosa (PAN) is a systemic necrotizing vasculitis of middle-sized arteries, found in adults and children. Disease pathogenesis is poorly understood. We identified multiple cases of PAN and cutaneous PAN in families and individuals of Georgian-Jewish ancestry, a historically endogamous isolate, consistent with autosomal recessive inheritance. While most cases had childhood onset, those with cutaneous PAN initiated in middle age. Methods: Exome sequencing was carried out for 4 affected individuals from 2 families and targeted sequencing in 16 additional Georgian-Jewish cases, and in 9 Turkish pediatric PAN cases. Mutations were assayed by protein structure analysis, expression in mammalian cells, biophysical analysis of purified protein, and protein activity in patient sera. Results: A founder mutation was identified in an immune-related gene, the only rare, damaging variant homozygous in all 4 exomes. Mutations in this gene have not been previously associated with any human phenotype. All Georgian-Jewish patients were homozygous for this mutation, except for one, who was compound heterozygous for this mutation and another missense mutation. There was a wide range of expressivity and of severity of the vasculitis phenotype in individuals who were homozygous for the same allele. One Turkish case was compound heterozygous for a mutation at the same codon and another missense mutation. In the Georgian Jewish population, the allele frequency was 0.05, reflecting the high prevalence of PAN in this endogamous community. The other mutations were absent both in the general population and in ethnically matched controls. The missense allele leads to loss of function of the protein both in patient sera and in vitro. Protein activity was significantly reduced in patient sera. Expression of mutant proteins in HEK293T cells revealed a significant reduction of secreted protein in the media. Conclusions: We report mutations in a gene not previously associated with any human phenotype as the first genetic cause of systemic vasculitis. Both population genetics and functional approaches were applied to demonstrate the causality of the founder allele. The mutations identified suggest blood vessels may be particularly vulnerable to loss of this serum protein's activity. They are consistent with PAN being an immune disorder, suggest possible new treatment and offer new opportunities to elucidate the pathogenesis of PAN and other forms of vasculitis.

226

***BMP9* mutations cause a vascular anomaly syndrome with phenotypic overlap with hereditary hemorrhagic telangiectasia.** *J. McDonald[1,2], W. Wooderchak-Donahue[3], B. O'Fallon[3], P. Upton[4], W. Li[4], B. Roman[5], S. Young[5], P. Plant[3], G. Fülöp[6], C. Langa[6], N. Morrell[4], L. Botella[6], C. Bernabeu[6], D. Stevenson[7], J. Runo[8], P. Bayrak-Toydemir[1,3].* 1) Department of Pathology, University of Utah, Salt Lake City, UT; 2) Department of Radiology, University of Utah, Salt Lake City, UT; 3) ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT; 4) Department of Medicine, University of Cambridge School of Clinical Medicine, Addenbrooke's Hospital, Cambridge, United Kingdom; 5) Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA; 6) Centro de Investigaciones Biologicas, CSIC and Centro de Investigacion Biomedica en Red de Enfermedades Raras (CIBERER), Madrid, Spain; 7) Division of Medical Genetics, Department of Pediatrics, University of Utah, Salt Lake City, UT; 8) Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI.

Hereditary hemorrhagic telangiectasia (HHT) is a vascular disorder characterized by telangiectases and arteriovenous malformations (AVMs). Hallmark features are recurrent nosebleeds due to telangiectases of the nasal mucosa; dermal, oral and gastrointestinal tract; and solid organ AVMs, particularly of the lungs, liver, and brain. The dermal telangiectases are typically pinpoint to pinhead size, specifically concentrated on the hands, face and lips, and not diffuse. Telangiectases on the limbs and trunk are not characteristic. Three genes in the transforming growth factor beta (TGF-β) signaling pathway (ENG, ACVRL1, and SMAD4) are currently known to cause HHT, but approximately 15% of individuals with clinical features of HHT do not have a mutation in one of these genes. We performed exome and/or Sanger sequencing on 191 unrelated individuals suspected to have HHT, who had tested negative for a mutation in ENG, ACVRL1, and SMAD4. A mutation predicted to be deleterious was identified in three unrelated probands in the BMP9 gene. These three individuals had nosebleeds and dermal lesions described as telangiectases, but which bore resemblance to lesions described in patients with RASA1-related disorders/CM-AVM (capillary malformation-arteriovenous malformation syndrome). Also, vascular liver findings similar to those seen in HHT were reported in one proband. CM-AVM, like HHT, is also characterized by telangiectases and AVMs, but recurrent nosebleeds are not described, and the typical dermal telangiectases differ from HHT in location and appearance. Telangiectases in CM-AVM include the tiny punctate lesions seen in HHT, as well as larger telangiectases referred to as capillary malformations. These punctate telangiectases are commonly diffuse and tend to cluster in a region of the trunk or limbs. Expression analyses were performed which suggest the identified mutations alter BMP9 protein processing and activity. Bmp9 knockdown experiments were performed in zebrafish and suggest BMP9 is involved in angiogenesis, but support the idea that BMP9 mutations might cause a novel vascular phenotype distinct from HHT. We identify a new genetic cause of a hereditary telangiectasia disorder which may overlap clinically with HHT and CM-AVM.

227

Identification of a Novel Cause of X-linked Heterotaxy. *M. Tariq[1], A.E. Cast[1], J.W. Belmont[2], S.M. Ware[1,3].* 1) Division of Molecular Cardiovascular Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

Heterotaxy is characterized by severe cardiovascular malformations resulting from abnormalities of the proper specification of left-right (LR) asymmetry during embryonic development with a prevalence of ~1 in 10,000 newborns. As a result of abnormal LR development, an unusual spatial arrangement of thoracic and/or abdominal organs may occur. The high relative risk of heterotaxy within families highlights its genetic basis. However, heterotaxy is challenging for traditional genetic analyses because of clinical and genetic heterogeneity, variable expressivity, and non-penetrance. Although the genetic basis is most frequently multifactorial, Mendelian inheritance does occur, and X-linked inheritance resulting from mutations in the transcription factor *ZIC3* was initially described in 1997. However, some X-linked pedigrees do not contain mutations in coding regions of *ZIC3* despite being closely linked to the *ZIC3* locus. In this study we recruited 4 X-linked heterotaxy pedigrees, previously found to be negative for *ZIC3* mutations, and investigated genetic causation using whole-exome or custom X-exome sequencing and segregation analysis of DNA variants. In one pedigree with 4 affected males including two with asplenia, gut malrotation, L-transposition of the great arteries, and abdominal situs inversus, a rare missense hemizygous variant in *GPR101* (p.Val409Met) was identified, predicted to be damaging for protein function by bioinformatic tools. This variant falls in the sixth transmembrane domain (TM1-7) of GPR101, an orphan G-protein coupled receptor with no known ligand. This candidate, which is located just 534 Kb upstream of *ZIC3*, was further investigated by expression analyses in mouse, including RT-PCR and whole mount in-situ hybridization, with results suggesting a possible role in LR pattering and early heart development. Morpholino-induced knockdown of *Gpr101* in *Xenopus laevis* resulted in heterotaxy with randomization of heart and gut looping, suggesting *Gpr101* is a novel gene involved in embryonic asymmetric laterality. This study identifies *GPR101* as a novel X-linked genetic cause of heterotaxy and provides opportunity to elucidate new mechanisms controlling embryonic LR asymmetry.

228

Novel and Recurrent Gain-of-Function Mutation in PRKG1 Causes Thoracic Aortic Aneurysms and Acute Aortic Dissections. *D. Guo[1], E. Regalado[1], D.E. Casteel[2], R.L. Santos-Cortez[3], L. Gong[1], J.J. Kim[4], S. Dyack[5], S.G. Horne[6], G. Chang[1], G. Jondeau[7], C. Boileau[7], J.S Collelli[8], Z. Li[9], S.M. Leal[3], J. Shendure[10], M.J. Rieder[10], M.J. Bamshad[10], D.A. Nickerson[10], C. Kim[11], D.M. Milewicz[1,12], GenTAC Registry Consortium; NHLBI-Go Exome Sequencing Project.* 1) Dept Internal Med, Univ Texas/ Houston Med Sch, Houston, TX; 2) Dept Medicine and Cancer Center, Univ California at San Diego, San Diego, CA; 3) Dept Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 4) Dept Pharmacology, Baylor College of Medicine, Houston, TX; 5) Dept Pediatrics, Dalhousie University, Halifax, Nova Scotia, Canada; 6) Dept Medicine, Dalhousie University, Halifax, Nova Scotia, Canada; 7) Hopital Bichat, Paris, France; 8) Michael E. DeBakey Dept Surgery, Baylor College of Medicine, Houston, TX; 9) Dept Internal Med, Univ Kentucky, Lexington, KY; 10) Dept Genome Sciences, Univ Washington, Seattle, WA; 11) Dept Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX; 12) Texas Heart Institute at St. Luke's Episcopal Hospital, Houston, TX.

Ascending thoracic aortic aneurysms can lead to life-threatening acute aortic dissections (TAAD). Previous studies have identified gene mutations that lead to decreased contraction of vascular smooth muscle cells (SMCs) causing familial TAAD (FTAAD), such as loss-of-function mutations in myosin light chain kinase (MLCK), which are predicted to decrease phosphorylation of the regulatory light chain (pRLC) and decrease SMC contraction. To identify further genes that cause FTAAD, exome sequencing of distant relatives from large families, followed by Sanger sequencing of unrelated probands with FTAAD, identified the same rare variant, PRKG1 c.530G>A (p.Arg177Gln), that segregated with TAAD in four families with a combined two-point LOD score of 7.88. Majority of affected individuals had aortic dissections (63%) at relatively young ages (mean 31 years, range 17 - 51 years) and TAAD was fully penetrant in both men and women. PRKG1 encodes a type I cGMP-dependent protein kinase (PKG I) that is activated upon binding of cGMP. PKG I exists as two splice variants (α and β), each with a unique set of substrates. PKG Iα phosphorylates regulatory myosin-binding subunit leading to de-phosphorylation of RLC and SMC relaxation. Protein structure analysis showed that Arg-177 is located at the cGMP binding domain A and critical for cGMP binding. Fluorescence polarization assay demonstrated that p.Arg177Gln abolished cGMP binding activity. However, kinase activity assay showed the mutant PKG Iα was highly active even without cGMP, while WT PKG Iα required cGMP for activation. These results suggest that the p.Arg177Gln mutation structurally perturbed PKG Iα and abolished allosteric inhibition of kinase activity resulting in a kinase whose activity is no longer modulated by cGMP. Immunoblot assays also showed reduced levels of pRLC in mutant fibroblasts from patients when exposed to low doses of 8-Fluo-cGMP compared with control fibroblasts. Thus, we have identified loss-of-function mutations in MLCK controlling SMC contraction and gain-of-function mutations in the kinase controlling relaxation. Finally, these data illustrate the complexity of associating rare variants with complex, adult onset diseases. Association studies based on a burden test of rare variants in PRKG1 with TAAD would not demonstrate an association, and studies with the specific PRKG1 p.Arg177Gln variant would only show an association in a large cohort or if family-based studies were pursued.

229

Mutations in the DCHS1 Gene Cause Mitral Valve Prolapse In Humans. *R. Durst[1,2], DS. Peal[3], A. deVlaming[4], M. Leyne[1], M. Talkowski[1], M. Perrocheau[5], C. Jett[1], C. Simpson[1], MR. Stone[1], F. Charles[1], C. Chiang[1], JA. Rosenfeld[11], X. Jeunemaitre[5,6,7], A. Hagege[6,7,8], N. Bouatia-Naji[5,6], FN. Delling[12], LA. Freed[13], C. Dina[9, 14, 15], JJ. Schott[9, 14, 15], KD. Irvine[10], Y. Mao[10], K. Sauls[4], A. Wessels[4], T. Motiwala[4], K. Williams[4], RR. Markwald[4], RA. Levine[3,16], DJ. Milan[3], RA. Norris[4], SA. Slaugenhaupt[1].* 1) Center for Human Genetic Research, Massachusetts General Hospital and Department of Neurology, Harvard Medical School, 185 Cambridge St., Boston, MA 02114 USA; 2) Cardiology Division, Hadassah Hebrew University Medical Center, Jerusalem, Israel, POB 12000; 3) Cardiovascular Research Center, Cardiology Division, Harvard Medical School and Massachusetts General Hospital, 55 Fruit Street, Boston, Massachusetts 02114 USA; 4) Department of Regenerative Medicine and Cell Biology, School of Medicine, Cardiovascular Developmental Biology Center, Children's Research Institute, Medical University of South Carolina, 171 Ashley Avenue, Charleston, SC 29425, USA; 5) 10INSERM, UMR-970, Paris Cardiovascular Research Center, 75015 Paris, France; 6) University Paris Descartes, Sorbonne Paris Cité, Faculty of Medicine, 75006 Paris, France; 7) Assistance Publique - Hôpitaux de Paris, Cardiology Department, Hopital EuropeenEuropéen Georges Pompidou, 75015, Paris, France; 8) INSERM, UMR-970 633, Paris Cardiovascular Research Center, 75015 Paris, France; 9) INSERM, UMR1087, l'Institut du Thorax, Nantes, France; CNRS, UMR6291, Nantes, France; 10) Howard Hughes Medical Institute, Waksman Institute and Department of Molecular Biology and Biochemistry, Rutgers, the State University of New Jersey, Piscataway NJ 08854 USA; 11) Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA 99207, USA; 12) Department of Medicine (Cardiovascular Division), Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA; 13) Yale-New Haven Hospital Heart and Vascular Center, Yale School of Medicine, 20 York Street, New Haven, CT 06510 USA; 14) Université de Nantes, Nantes, France; 15) CHU Nantes, l'Institut du Thorax, France; 16) Cardiac Ultrasound Laboratory, Cardiology Division, Massachusetts General Hospital, Harvard Medical School, 55 Fruit Street, Boston, Massachusetts 02114 USA.

Mitral valve prolapse (MVP) is a common cardiac valve disease that affects nearly 1 in 40 individuals. MVP can manifest clinically as mitral regurgitation and lead to arrhythmia, heart failure, and sudden death. Despite a clear heritable component, the genetic etiology and the defective developmental pathways that lead to MVP have remained elusive. A large multigenerational family segregating non-syndromic MVP was linked to chromosome 11 and DNA sequencing of the candidate region was performed on four affected family members. We report a genetic variant in the DCSH1 gene as a genetic risk factor for non-syndromic MVP in humans. Morpholino knock-down of the zebrafish homologue of DCHS1 results in a disruption of the atrioventricular constriction that is not rescued by the human mutation. Loss of function studies in Dchs1 knockout mice result in a mitral valve-specific valvulopathy characterized by a myxomatous phenotype with leaflet elongation and thickening in adults. This adult pathology was traced back to developmental errors in interstitial cell alignment during valve morphogenesis. DCHS1 haploinsufficiency, resulting from decreased stability of the mutant protein, leads to non-syndromic MVP in a large pedigree. Evaluation of mice heterozygous for Dchs1 loss demonstrates that disruption of interstitial cell alignment during valve morphogenesis can result in a myxomatous valvulopathy in adults, implicating a previously unrecognized paradigm in valve development in the long-term structural integrity of the mitral valve.

230

Robust epistasis between the genes encoding a TGFβ effector and its regulatory microRNA governs modification of cardiovascular phenotypes in TGFβ vasculopathies. *J. Calderon[1], H. Dietz[1,2].* 1) Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Howard Hughes Medical Institute, Chevy Chase, MD.

The Loeys-Dietz Syndrome (LDS) spectrum disorders share a predisposition for widespread and aggressive vascular lesions (tortuosity and aneurysms) in association with primary heterozygous mutations in genes encoding critical TGFβ signaling effectors (*TGFBR1, TGFBR2, SMAD3* and *TGFB2*). Many studies document that the C57BL/6J (B6) background exacerbates TGFβ deficiency phenotypes when compared to 129SvE (SV129), but the source of modification is unknown. While postnatal aneurysm progression is blunted upon crossing LDS mutations onto B6, LDS mice show near-complete penetrance of perinatal death due to truncus arteriosus (TA), a congenital heart defect that has been associated with loss of TGFβ signaling in the neural crest, once the B6 contribution becomes substantial. This discrete readout in LDS was utilized to map the relevant modifier alleles. Pure SV129 *Tgfbr2*$^{G357W/+}$ mice were bred to F2 WT mice with extensive recombination between B6 and SV129 chromosomes and E17.5 fetuses were phenotyped for TA. GWA analysis revealed a single major B6 locus associated with TA on chr 9 (LOD=9); a minor linkage signal was apparent on chr X (LOD=2.75). Combinatorial analyses revealed that TA was dependent on B6 contribution at the chr 9 locus, but was thereafter dramatically influenced by B6 contribution at the chr X locus in a dosage-dependent manner. *Tgfbr2* emerged as a promising candidate gene in the region of association on chr 9, both by virtue of its known function and the presence of strain-specific variation in its 3'UTR that is predicted to alter mRNA conformation. In an unbiased search for miRNAs that target *Tgfbr2*, *miR-106a* emerged as a prominent candidate because it is located within the suggestive peak of association on chr X. A luciferase reporter allele harboring the B6 *Tgfbr2* 3'UTR showed dramatically reduced activity when compared to its 129 counterpart (p<10E-9) in transfection studies, with a disproportionately increased activity upon addition of a *miR-106a* antagonist. These results suggest greater sensitivity of the B6 version of *Tgfbr2* to the repressive activity of *miR-106a*. We further showed that vascular smooth muscle cells from pure B6 animals express significantly less Tgfbr2 protein and exhibit blunted activation of intracellular signaling mediators in response to TGFβ, when compared to 129. These data reveal a complex architecture of modification that may prove relevant to other Mendelian disorders and inform therapeutic strategies for LDS.

231

ERK Activation Unifies Deleterious Gene-by-Gene and Gene-by-Environment Interactions in Marfan Syndrome. *J.J. Doyle[1,2], A.J. Doyle*[1], N. Wilson[1], D. Bedja[3], J. Pardo-Habashi[1,4], L. Myers[1], K. Braunstein[5], N. Huso[1], S. Bachir[1], O. Squires[1], B. Rusholme[1], A. George[1], M. Lindsay[1], D. Huso[3], C. Thomas[6], D. Judge[7], H.C. Dietz[1,7,8].* 1) Inst Gen Med, Johns Hopkins, Baltimore, MD; 2) Good Samaritan Hospital, Baltimore, MD; 3) Dept of Pathobiology, Johns Hopkins, Baltimore, MD; 4) Dept of Pediatrics, Johns Hopkins, Baltimore, MD; 5) Dept of Pathology, Johns Hopkins, Baltimore, MD; 6) National Institutes of Health Chemical Genomics Center, Rockville, MD; 7) Dept of Medicine, Johns Hopkins, Baltimore, MD; 8) Howard Hughes Medical Institute, Baltimore, MD.

Many of the manifestations of Marfan syndrome (MFS), including aortic aneurysm, associate with a clear signature of increased TGFβ signaling and can be attenuated by TGFβ neutralizing antibody or the AT1R blocker losartan. While TGFβ can signal through both canonical (Smad2/3) and noncanonical (ERK) pathways, selective ERK inhibition using RDEA119 fully rescues aortic aneurysm growth in MFS mice. To further inform our understanding of ERK activation in MFS, we analyzed its role in mediating both genetic and environmental modifications of aortic disease in MFS mice. Compared to MFS mice on a pure BL6 background, MFS mice on a pure 129 background show a 4-fold increase in aortic aneurysm growth and unique lethality from aortic dissection. Calcium channel blockers (CCBs) are used in MFS patients due to their antihypertensive effects, despite limited empirical evidence for their utility. Unexpectedly, we find that BL6 MFS mice treated with the CCBs amlodipine or verapamil show hyperacute aneurysm expansion, aortic rupture and premature lethality. Compared to BL6 animals, 129 MFS mice show an overt increase in Smad2/3 and ERK activation, as do CCB-treated MFS mice compared to placebo-treated littermates. Despite the severity of these provocations, both losartan and RDEA119 fully rescue aortic aneurysm growth in 129 MFS mice, while dual treatment with CCB and either losartan or RDEA119 fully rescues this deleterious gene-by-environment interaction. Taken together, these data instill confidence that these conditional experimental provocations inform the underlying pathogenesis of MFS. We next sought to identify a central nodal point that integrates these observations. AT1R, TGFβ and CCBs are known to drive IP3-mediated PKC activation, a critical mediator of ERK activation, while AT2R signaling (which is protective in MFS) can antagonize it. MFS mice show greatly increased PKC activation compared to WT animals, which is further enhanced by either the 129 background or CCB therapy. Furthermore, hydralazine (which can inhibit IP3-mediated PKC activation) fully rescues aortic aneurysm growth in MFS mice, in conjunction with reduced PKC and ERK activation. These data definitively implicate ERK in driving aortic disease in MFS mice, implicate PKC as a critical integration point in the pathway, challenge prevailing assumptions about the effect of CCBs in MFS, and provide rationale and incentive for a clinical trial of ERK and PKC antagonists and hydralazine in MFS.

232

Reasons why patients decline whole genome sequencing in the MedSeq Project. *D.M. Lautenbach[1], J.L. Vassy[2], K.D. Christensen[3], A.L. McGuire[4], H.L. Rehm[5], M.F. Murray[6], C.Y. Ho[7], C.A. MacRae[3,7], C.E. Seidman[3,7], R.M. Miller[1], C. Liu[1], R.C. Green[1,3].* 1) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston , MA; 2) VA Boston Healthcare System, Boston, MA; 3) Harvard Medical School, Boston, MA; 4) Baylor College of Medicine Center for Ethics and Health Policy, Houston, TX; 5) Laboratory for Molecular Medicine, Partners Center for Personalized Genetic Medicine, Boston, MA; 6) Genomic Medicine Institute, Geisinger Health System, Danville, PA; 7) Division of Cardiology, Department of Medicine, Brigham and Women's Hospital, Boston, MA.

Background: As whole genome sequencing (WGS) becomes more accessible, it is important to understand why some patients who are interested in WGS ultimately decide not to pursue it. The MedSeq Project is a randomized clinical trial exploring how physicians and patients interpret, understand and utilize WGS results in primary care and cardiology settings. MedSeq Project patients are randomized to receive either usual care or usual care plus WGS. Reports with the genome results are placed in the patient's electronic medical record. **Methods:** Enrolled physician-participants from primary care and cardiology recruit patients from their own practice and refer interested patients to the study staff who review the risks, benefits and limitations of WGS first by phone and then in-person during the informed consent. Reasons why these well-informed and initially-interested patients declined participation were recorded by study staff and coded by consensus. **Results:** During the first 6 months of recruitment, 113 patients were referred from primary care and 38 patients from cardiology. Twenty patients from primary care (17.7%) and 6 from cardiology (15.8%) declined participation. In total, 24 patients provided 25 reasons for declining study participation (1 patient gave 2 reasons and 2 gave no reason). Nineteen of 25 reasons (76%) were related to a potential negative impact: 12 patients were concerned about potential insurance discrimination with their results placed in their medical record, 6 did not wish to learn specific results, and 1 was concerned about privacy in regard to his/her sequencing data being shared with the scientific community. Six of 25 reasons (24%) were logistical: 6 patients were concerned about the time commitments of participating. Three of these 6 desired WGS but did not want to commit because the randomized study design meant that they might not receive WGS. **Conclusions:** Preliminary data suggests that some patients interested in WGS decline such testing due to concerns about a potential negative personal impact of having these results available and placed in their medical record. Others declined WGS simply because they did not want to commit their time to the research visits, surveys and interviews. Some were so enthusiastic about WGS that they did not wish to contribute their time to the study if not guaranteed WGS. We will continue to collect reasons for decline as recruitment progresses.

233

How do research participants perceive "uncertainty" in genomic sequencing? *B. Biesecker[1], L. Biesecker[2], P. Han[3].* 1) Soc Behavioral Res Branch, NHGRI/NIH, Bethesda, MD; 2) Genetic Disease Rsearch Branch, NHGRI/NIH, Bethesda, MD; 3) Maine Medical Center, 81 Research Dr Scarborough, ME 04074.

How research participants and patients interpret uncertainty in health care is not well understood. When investigators and health care providers describe health information as uncertain, they often assume that the recipient shares their intended meaning. No data exists to support or refute this assumption. Further, little is known about the consequences of communicating uncertainty to patients facing a health care decision. Politi and colleagues have identified conceptual differences in how uncertainty is defined and measured and call for exploratory research. We present data from adults participating in an NIH genome sequencing cohort study where uncertainty is hypothesized to be a key variable in predicting decisions to learn health-related information from one's genome sequence. We conducted six focus groups with 39 ClinSeq® participants that elicited perceptions of the uncertainty associated with genomic sequencing. Participants perceived uncertainty as pertaining to the type and quality of information generated, describing it as ambiguous, of questionable accuracy/validity, limited, probabilistic, fluid, and infinite. These perceptions lead to concerns that the uncertainty may undermine their faith in the information, or lead to confusion, uneasiness, hopelessness or further questions. As such, participants requested that investigators define the degree of uncertainty associated with any information and estimate the probabilities for disease. Some only want to learn "valid" information. One prefers a high level of certainty when making health related decisions. The majority also stated that uncertainty was expected of genomic information, that it was reality, unsurprising, fair, accepted and honest. They further described how uncertainty means that more will be learned in the future, knowledge is unfolding and that they are curious to see "how far it will go." These varied descriptions of uncertainty need to be replicated in larger studies, but suggest that in consenting participants to genomic sequencing, different dimensions of uncertainty should be raised. And serve to remind investigators that uncertainty is not uniformly interpreted as negative but at times seen as an opportunity.

234

Factors Influencing Healthcare Utilization in Response to Personal Genetic Testing. *S.S. Kalia[1], K.D. Christensen[1], C.A. Chen[2], J.L. Mountain[3], T.A. Moreno[4], J.S. Roberts[5], R.C. Green[1,6] for the PGen Study group.* 1) Department of Medicine, Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2) Data Coordinating Center, Boston University School of Public Health, Boston, MA; 3) 23andMe, Inc., Mountain View, CA; 4) Pathway Genomics, San Diego, CA; 5) Department of Health Behavior and Health Education, University of Michigan School of Public Health, Ann Arbor, MI; 6) Partners Center for Personalized Genetic Medicine, Boston, MA.

Background: The expansion of personal genetic testing has raised questions about whether consumers will alter their use of healthcare resources after viewing their results. In the Impact of Personal Genomics (PGen) Study, we prospectively surveyed personal genetics consumers about their use of tests, exams, or procedures after seeing their genetic risks. Hypothesis: We hypothesized that pretest intention to discuss results with physicians, post-test perceptions of results, and demographic factors would be associated with having tests, exams, or procedures after receiving personal genetic information. Methods: Participants were new customers of Pathway Genomics, recruited through the health-based social network Patients Like Me, and new customers of 23andMe. Subjects completed an anonymous online survey after purchasing the testing but before receiving results. Follow-up surveys were completed at 2-3 weeks and 6 months after receiving results. Results: Of 986 respondents to the 6-month follow-up survey, 104 (10.7%) reported having tests, medical exams, or procedures as a result of seeing their genetic risk information for at least 25 diseases, including heart disease, diabetes, arthritis and macular degeneration. Self-reported tests, exams and procedures included blood tests (e.g. glucose, lipids, rheumatoid factor), cancer screening procedures (e.g. mammography, colonoscopy), MRI, ultrasound, and ophthalmology exams, among others. In bivariate analyses, older age was associated with increased likelihood of having tests, exams, or procedures ($F=13.3$, $p<0.001$). In logistic regression, after controlling for demographic factors, having tests, exams, or procedures was associated with poorer self-reported health (OR=2.4, $p<0.001$), higher pretest anxiety (OR=2.4, $p=0.002$), pretest plans to discuss results with one's PCP (OR=1.8, $p=0.008$) or another medical professional (OR=1.8, $p=0.010$), and perceiving 'many' or 'all' results as interesting (OR=2.2, $p=0.005$), useful for improving one's health (OR=3.4, $p<0.001$) or above average risk (OR=2.0, $p=0.010$). Conclusions: Pretest intentions, self-reported health, and psychological states were associated with healthcare utilization after receiving personal genetic information, as were post-test risk perceptions and perceptions about test utility. Results suggest ways to identify customers who are more likely to pursue additional tests or exams through their physicians in response to personal genetic testing.

235

Does Personal Genome Testing drive service utilization in an adult preventive medicine clinic? *N. Hoang[1], R. Hayeems[1, 4], J. Davies[2], L. Velsher[2], J. Aw[2], S. Pu[1], S. Wodak[1], S. Chenier[3], J. Stavropoulos[1], R. Babul-Hirji[1], R. Weksberg[1], C. Shuman[1].* 1) Clinical and Metabolic Genetics, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Genetics, Medcan, Toronto, Ontario, Canada; 3) Département de pédiatrie, Centre Hospitalier Universitaire de Sherbrooke, Sherbrooke, Quebec, Canada; 4) Institute of Health Policy Management & Evaluation, University of Toronto, Toronto, Ontario, Canada.

Background: Genome-wide single nucleotide polymorphism arrays that assess individual genetic risk for common polygenic diseases can guide the use of preventive health care services, but outcome data are limited. We aimed to understand how personal genome testing (PGT) risk results relate to subsequent health service utilization in a preventive medicine clinic. **Method:** We conducted a retrospective medical record review for two groups of clients at Medcan, an executive health clinic in Toronto. Ascertainment of cases over a 1 year period included those who pursued PGT at their comprehensive health assessment (CHA) (denoted CHA1); controls underwent CHA but with no PGT. We measured condition specific services used post CHA1 up to CHA3 (typically a 2 year period). PGT risk estimates for nine conditions were examined including: abdominal aneurysm, atrial fibrillation, celiac disease, colon cancer, type 2 diabetes, glaucoma, heart attack, melanoma, and prostate cancer. Client data were collected on: age, sex, family history, health risk status at CHA1 and CHA2, environmental modifiers (exercise level, alcohol intake, and smoking) and Medcan membership. Of 448 PGT ordered, 388 (86.6%) met inclusion criteria; a random sample of 388 matched Medcan clients were selected as controls. Analysis for all 9 conditions will be presented but the data here pertain to risk for heart attack. **Results:** Binomial log link regression analyses of relative risk (RR) indicate that being identified as at risk for heart attack on CHA1 and/or CHA2 is strongly associated with increased heart related services used compared to not being at risk [RR=4.44, 95% CI (3.57, 5.53)]. Similarly, age over 40 is associated with increased heart related services used compared to age under 40 [RRage41-50=2.03, 95% CI (1.33, 3.10); RRage50+=2.85, 95% CI (1.92, 4.23)]. Pursuing PGT itself was associated with increased services used [RR=1.48, 95% CI (1.20, 1.82)]. With respect to risk conferred by PGT, having an average risk for heart attack was associated with increased services used compared to the control group [RR=1.53, 95% CI (1.24,1.89)], but there was no difference between the increased risk group and the control group [RR=1.16, 95% CI (0.76, 1.79)]. **Conclusions:** For heart attack, the strongest drivers of related services used were CHA health risk status and age. It would appear that individuals who pursue PGT are likely to use more health related services regardless of the actual PGT test result.

236

Opportunity and cost of clinical whole genome sequencing. *F. Dewey[1,2], M. Grove[1,2], C. Pan[2,3], B. Goldstein[4], J. Bernstein[5], H. Chaib[2,3], R. Goldfeder[6], K. Ormond[3,7], C. Caleshu[1,2,5], K. Kingham[8], T. Klein[3], M. Whirl-Carrillo[3], K. Sakamoto[3,7], M. Wheeler[1,2], A. Butte[9], J. Ford[8], L. Boxer[4], J. Ioannidis[4,10,11,12], A. Yeung[1], R. Altman[3,13], T. Assimes[1], M. Snyder[2,3], E. Ashley[1,2,3], T. Quertermous[2,3].* 1) Stanford Center for Inherited Cardiovascular Disease, Stanford, CA; 2) Stanford Center for Genomics and Personalized Medicine, Stanford, CA; 3) Stanford Department of Genetics, Stanford, CA; 4) Stanford Department of Medicine, Stanford, CA; 5) Stanford Department of Pediatrics, Stanford, CA; 6) Stanford Biomedical Informatics Training Program, Stanford, CA; 7) Stanford Center for Biomedical Ethics, Stanford, CA; 8) Stanford Division of Medical Oncology, Stanford, CA; 9) Stanford Division of Systems Medicine, Stanford, CA; 10) Stanford Department of Statistics, Stanford, CA; 11) Stanford Prevention Research Center, Stanford, CA; 12) Stanford Department of Health Research and Policy, Stanford, CA; 13) Stanford Department of Bioengineering, Stanford, CA.

Background Exome and whole-genome sequencing (WGS) are increasingly applied in clinical genetics and clinical research. WGS is expected to uncover potentially clinically significant findings regardless of the primary indication for sequencing. We investigated the burden of clinically reportable findings in WGS data, resources required for their discovery and interpretation, and costs of associated secondary evaluation. Methods We performed WGS on two platforms and clinical genome interpretation for adult participants in a primary care setting. A multi-disciplinary team reviewed all potentially reportable genetic findings and generated medical genomics reports for discussion with the study participant, genetic counselor, primary physician and medical geneticist. These physicians proposed secondary evaluations based on the report. Results Review of 90-127 Mendelian disease risk candidates in each participant (n=10, median age 53, 6 female) required a median of 54 (range 5-223) minutes per genetic variant, resulted in modestly good inter-rater classification agreement between genomic professionals (Gross' kappa 0.61, 95% confidence interval 0.47-0.74), and reclassified 70% of genetic variants cataloged as "disease causing" in mutation databases to variants of uncertain or lesser significance. A median of 5 (range 2-6) reportable personal disease risk findings were discovered in each individual, including a frameshift deletion in BRCA1 strongly associated with hereditary breast and ovarian cancer that was found in a participant without family or personal history of the disease. Four of ten participants harbored variants cataloged by mutation databases as "disease causing" in genes recommended by the ACMG for return of incidental findings, and seven harbored apparent "expected pathogenic" novel variants in these genes. Cross-platform concordance of these latter variants was low (0-10%). Physician review of medical genomics reports prompted consideration of a median of 3 (range 0-10) follow-up diagnostic tests and referrals at an estimated total cost of $594 (range $362-1427) per individual. Conclusions WGS-based genomics involves considerable manual interpretation of potentially clinically significant genetic variants, uncertainty among genomics professionals and knowledge bases about these findings, and a moderate financial cost of secondary clinical evaluation. In select cases WGS will reveal genetic findings warranting early medical intervention.

237

Cost-effectiveness analysis (CEA) of next generation sequencing (NGS) in etiologic evaluations for prelingual hearing loss. *M.E. Nunes[1], N.T. Manzano[1], J.L. Natoli[2], K.A. Wendt[3].* 1) Dept Genetics/Pediatrics, Kaiser Permanente, San Diego, CA; 2) Department of Clinical Analysis Technology Assessment and Guidelines Unit Kaiser Permanente, Pasadena, CA; 3) Medical Genetics and Genomics Laboratories, Kaiser Permanente, Los Angeles, CA.

We used cost-effectiveness analysis (CEA) to examine the estimated cost, diagnostic impact, and efficiency afforded by incorporating Next Generation Sequencing (NGS) into hearing loss evaluations performed during the first year of life. Newborn hearing screen (NBHS) identifies hearing loss in 1.2 per 1,000 infants. Half of the congenital cases are ultimately found to be genetic, of which 30% are syndromic, requiring additional diagnostic evaluation. Establishing etiology may identify comorbidity, determine management and prognosis, and facilitate genetic counseling. A multidisciplinary systematic approach includes appropriate genetic testing, until recently limited within our managed care institution to *Connexin 26* and *30* (*Cx 26/30*) genes and locus specific testing. Using the evaluation strategy proposed by the British Association of Audiological Physicians (2007), we evaluated the theoretical impact on diagnosis and cost of three genetic testing schemes: 1) *Cx26/30*, 2) *Cx 26/30* as first tier with reflex to NGS if negative, and 3) NGS only. Genetic testing would occur after confirmation of congenital hearing loss and clinical evaluation by an Otolaryngologist, prior to additional work-up for genetic and non-genetic etiology. The cost of NGS was $1500 per sample, assumed to incorporate > 50 loci for non-syndromic and syndromic hearing loss. We calculated total, average, and marginal cost per diagnosis established. Cytomegalovirus (CMV) testing ($286,970) and Ophthalmology evaluation ($443,182) had the largest marginal cost per diagnosis made, *Connexin 26/30* gene testing ($3494) and Otolaryngology clinical evaluation ($5720) the lowest. Scheme 1 identified a diagnosis in 33% by 1 year of age, with average cost per diagnosis $6168. Scheme 2 identified 41%, with average cost $7456. Scheme 3 likewise identified 41%, but became the most effective method with average cost $6679 and NGS marginal cost per diagnosis $5249. Although the individual cost of DNA testing for *Connexin 26/30* is considerably less than NGS, introducing NGS early in the evaluation of congenital hearing loss reduced both diagnoses made by clinical evaluations and need for imaging studies, resulting in cost saving. This analysis suggests that NGS could successfully be incorporated early in hearing loss evaluations with significant savings in time and money. Similar CEA are possible to justify the expense of NGS platforms in other medical conditions with well established diagnostic pathways.

238

Documentation of medical decision-making for genetic testing in the health record. *M.T. Scheuner[1,2], J. Peredo[1], T.J. Sale[1], B.T. Tran[1], A.T. Jones[1], A.B. Hamilton[1,3], L. Hilborne[4,5], I.M. Lubin[6].* 1) VA Greater Los Angeles Healthcare System, Los Angeles, California; 2) Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, Califonria; 3) Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine at UCLA, los Angeles, California; 4) Quest Diagnostics, West Hills, California; 5) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, Los Angeles, California; 6) Division of Laboratory Sciences and Standards, Centers for Disease Control and Prevention, Atlanta, Georgia.

Objective: Errors in the genetic testing process most likely occur in the pre-analytic (ordering) and post-analytic (results interpretation) phases. We assessed documentation of medical decision-making for relatively common genetic tests (cystic fibrosis, HLA-B27, hemochromatosis, and thrombophilia). Methods: Health records were selected from 122 patients who had a genetic test of interest ordered by a non-geneticist (Apr-Sep 2011). Each record included progress notes from a 6-month period prior to and after the test collection date. Outcomes abstracted from the records included documentation of four pre-analytic (indication for testing, test name, relevant family history, and informed consent) and three post-analytic outcomes (test result, implications of the result, and management recommendations). Results: Mean age of patients was 50.5 years (SD 16.3), 36.1% were white, and 83.6% were male. Documentation of test indication was found in 98.4% of records, test name in 87.7%, family history in 52.4%, and informed consent in only 31.1%. Abnormal results were found in 18.0% of records. Documentation of the result was found in only 49.2%; more often when abnormal, 86% vs. 40%. Management recommendations were found in 50%, usually limited to the condition of concern and rarely specific to the result. Management was documented for abnormal results twice as often than normal results, 86% vs. 41%. Only 1 of 9 records had documentation of the implications of the result with respect to diagnosis, disease and reproductive risks, or the validity and limitations of the testing. Implications were only documented for a subset of abnormal results. Conclusions: Medical decision-making relating to genetic testing was documented more often for the pre-analytic than post-analytic outcomes. Family history was often lacking, which could compromise appropriate test selection and accurate interpretation of results. Documentation of informed consent was found in a minority of records, suggesting this may not be viewed as necessary by most clinicians. Lack of documentation of normal results and their implications raise concerns that clinicians may not recognize the limitations of the testing performed. There is a need for education and clinical decision support designed to improve clinician behavior and documentation during the pre- and post-analytic phases of genetic testing, which could result in improved utilization of genetic testing resources and health outcomes.

239

Medical Genetics and Genomics: Parallel Revolutions in Science and Undergraduate Medical Education. *S. Dasgupta[1], K. Hyland[2], K. Garber[3], J-A. Gold[4], H. Toriello[5], K. Weissbecker[6], D. Waggoner[7].* 1) Department of Medicine, Biomedical Genetics, Boston University School of Medicine, Boston, MA; 2) Department of Biochemistry and Biophysics, University of California, San Francisco, School of Medicine, San Francisco, CA; 3) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 4) Pediatric Division of Genetics, Loma Linda University Medical School, Loma Linda, CA; 5) College of Human Medicine, Michigan State University, Grand Rapids, MI; 6) Hayward Genetics Center, Tulane University Medical School, New Orleans, LA; 7) Department of Human Genetics, University of Chicago, Pritzker School of Medicine , Chicago, IL.

PURPOSE

Our knowledge of the influence of genetic mechanisms on health and disease has grown exponentially. Much of these advances are attributable to the Human Genome Project and the development of genomic-based technologies. As a result, today's physicians need a comprehensive understanding of the principles of genetics and genomics, from basic science to clinical application, in order to make informed clinical decisions.

METHODS

In response to these advances, the Association of Professors of Human and Medical Genetics (APHMG) developed and updated undergraduate medical education core curriculum in genetics[1]. We also mapped these updated learning objectives to the ACGME competency domains. In parallel, we used a survey administered to medical genetics curriculum directors at US and Canadian medical schools to examine curricular trends in terms of student demographics, curriculum structure and oversight, as well as educational content.

RESULTS

Recognizing a renewed movement toward competency-based education, we constructed a set of broad competencies required of all graduating medical students. Both basic scientists and clinicians contributed to this document, which has been approved and endorsed by the APHMG. Through the survey, we identified topics of emerging importance to be included in a medical curriculum (e.g., use of microarrays, direct-to-consumer testing and the Genetics Information Nondiscrimination Act) and topics of declining importance (e.g., linkage studies). The survey also queried hours dedicated to genetics instruction and structural policies that impact the methods of instruction. In particular, we have described the context of medical genetics with respect to the students' year of study and whether the settings were mixed classrooms including students from multiple professional programs. We also examined who the instructors are, what their training is, what teaching modalities they employ, and how they are supported for these efforts.

CONCLUSIONS

The competency document serves as a definitive guide for curriculum directors and teachers of genetics content in medical schools, and it provides flexibility for individual curricular models. The survey results complement the competencies in giving medical genetics curriculum directors benchmarks with which to demonstrate the need for further development of medical genetics education at their home institutions.

REFERENCES

1. AJHG (1995, 56:535-537), with revisions in 2001, 2010, and 2013.

240

Genomic medicine in primary care: views of Ontario family physicians.

S. Morrison[1], J. Allanson[1,3], F.A. Miller[4], J.A. Permaul[5], B.J. Wilson[6], J.C. Carroll[2]. 1) Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 2) Department of Family & Community Medicine, Mount Sinai Hospital, University of Toronto, Toronto, ON, Canada; 3) Department of Pediatrics, University of Ottawa, Ottawa, ON, Canada; 4) Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada; 5) Mount Sinai Hospital, Toronto, ON, Canada; 6) Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, ON, Canada.

Context: Primary care providers (PCPs) will be critical to the effective and appropriate implementation of genomic medicine (GM), which is expected to profoundly affect medical practice. Little is known about what role family physicians (FPs) see for themselves, including attitudes toward genomics and resources required. Objective: To determine FPs' current involvement in GM, attitudes regarding its importance, confidence in knowledge, and beliefs about resources needed to deliver GM. Design: Mailed self-complete questionnaire. Setting and Participants: Random sample of 2000 Ontario (Canada) FPs from Scott's Directory. Eligibility included practicing family medicine in Ontario. Intervention: Questionnaire including current practice, confidence in genetic competencies, attitudes to GM, knowledge, education, resource needs and demographics. Modified Dillman Method used. Outcomes: Role, confidence, attitudes, educational resources needed. Results: Adjusted response rate: 358/1363 eligible FPs (26%). Mean age 51 years, 53% male. FPs reported high involvement in some aspects of genetics (e.g., eliciting family history, 92%; deciding to whom to offer genetic counseling, 91%), and a majority expected that advances in genomic medicine would improve patient health outcomes (59%). Yet involvement in emerging genetic practices was more limited (e.g., evaluating usefulness of genetic tests, 51%; discussing pharmacogenetics, 29%), and FPs expressed reduced enthusiasm regarding newer developments (e.g., incorporating GM into practice, 36%; importance of learning about personalized medicine based on targeted or whole genome sequencing, 44%). FPs valued educational resources in genetics (e.g., genetic disorder summaries, 86%; care pathways for common genetic disorders, 81%; genetic testing guidelines, 96%; website 59%), and preferred access to a local genetic counselor (66%) to aid in integrating GM. Conclusion: FPs have an established role in genomic medicine, but are cautious about newer developments. They have limited confidence in genomic medicine competencies but high interest in educational resources to enable practice. These results can be used to guide both residency and continuing education.

241

Phenotype and genotype in 17 patients with succinate-CoA ligase deficiency caused by mutations in *SUCLA2* and *SUCLG1*.

E. Oestergaard[1], M. Rasmussen[2], H. Amartino[3], IF. de Coo[4], DC. Buhas[5], S. Mesli[6], K. Naess[7], M. Tulinius[8], N. Darin[8], M. Duno[1], P. Jouvencel[9], I. Redonnet-Vernhet[6], F. Wibrand[1], E. Holme[10]. 1) Clinical Genetics, Copenhagen University Hospital, Copenhagen, Denmark; 2) Neuropediatric Department, Oslo University Hospital Rikshospitalet, Oslo, Norway; 3) Servicio de Neurología Infantil, Hospital Universitario Austral, Buenos Aires, Argentina; 4) Department of Neurology, Erasmus Medical Centre, Rotterdam, the Netherlands; 5) Department of Genetics, Sainte-Justine University Hospital, Montreal, QC, Canada; 6) Biochemistry, CHU de Bordeaux, Bordeaux, France; 7) Department of Laboratory Medicine and Centre for Inherited Metabolic Diseases, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden; 8) Department of Pediatrics, University of Gothenburg, The Queen Silvia's Children Hospital, Gothenburg, Sweden; 9) Neonatal and Pediatric Intensive Care Unit, Children's Hospital, Bordeaux, France; 10) Department of Clinical Chemistry, Institute of Biomedicine, The Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden.

Background: The encephalomyopathic mtDNA depletion syndrome with methylmalonic acidura is associated with deficiency of the tricarboxylic acid cycle enzyme succinate-CoA ligase, and is caused by mutations in *SUCLA2* or *SUCLG1*. Previously, 24 patients with *SUCLA2* mutations have been reported, and 16 patients with *SUCLG1* mutations. We here report the clinical and molecular genetic findings in 17 additional patients with succinate-CoA ligase deficiency. **Patients and results:** Of the 17 patients, 13 had *SUCLA2* mutations and four had *SUCLG1* mutations. In *SUCLA2*, we found seven mutations, of which six were novel. One was a whole gene deletion, one was a 1 bp duplication, and four were missense mutations. Three of four *SUCLG1* mutations were novel, all missense. Onset of symptoms was from birth to 1-2 years of age. The presenting symptom in most patients was muscle hypotonia, with onset from birth to 5½ months of age. In one patient, onset of symptoms was at age 1-2 years with sensorineural hearing impairment. Most patients had dystonia, whereas only a few had epilepsy. Nearly all patients had severe hearing impairment; the age at diagnosis of hearing impairment was from 8 months to 4 years. In five of the patients, the disorder was associated with unusually long survival, beyond 20 years. No significant clinical differences between the patient groups with *SUCLA2* and *SUCLG1* mutations were identified. Brain MRI/CT was done in 11 patients; abnormal basal ganglia were found in nine. Five patients had central and cortical atrophy, and in three patients abnormalities of the white matter was found. Analysis of respiratory chain enzyme activity in muscle generally showed a combined deficiency of complexes I and IV, and in some cases also of complex II + III. One patient had normal enzyme activity in muscle. Urine organic acids showed elevated excretion of methylmalonic acid except in one patient, although this patient did have elevated methylcitrate. **Conclusions:** In the 17 patients with succinate-CoA ligase deficiency we found nine novel mutations in *SUCLA2* and *SUCLG1*. Five patients had long survival, beyond 20 years, which has only been reported in a few patients previously. Normal respiratory chain enzyme activity in muscle or urine excretion of normal amounts of methylmalonic acid does not exclude a succinate-CoA ligase deficiency.

242

Restoration of the mitochondrial citrate transporter by overexpression of SLC25A1 in primary deficient fibroblasts of patients with combined D-2- and L-2-hydroxyglutaric aciduria. *G.S. Salomons[1], E.A. Struys[1], A. Pop[1], E.E. Jansen[1], M.R. Fernandez Ojeda[1], W.A. Kanhai[1], M. Kranendijk[1], S.J.M. van Dooren[1], M.R. Bevova[2], E.A. Sistermans[2], A.W.M. Nieuwint[2], M. Barth[3], T. Ben-Omran[4], G.F. Hoffmann[5], P. de Lonlay[6], M.T. McDonald[7], A. Meberg[8], A.C. Mühlhausen[9], C. Muntau[10], J-M. Nuoffer[11], R. Parini[12], M-H. Read[13], A. Renneberg[14], R. Santer[9], T. Strahleck[16], E. van Schaftingen[15], M.S. van der Knaap[17], C. Jakobs[1], B. Nota[1].* 1) Metabolic Unit, VU University Med Center, Amsterdam, Netherlands; 2) Dept of Clin Genetics, VUmc, Amsterdam, Netherlands; 3) Génétique CHU Angers, Angers, France; 4) Clinical and Metabolic Genetics,HMC WCMC, Doha, Qatar; 5) Dept Kinderheilkunde I, Heidelberg, Germany; 6) Inh Metabolic Diseases, Hôpital Necker, Paris, France; 7) Div of Med Genetics, Duke University, Durham, United States; 8) Neonatal Unit, Vestfold Hospital Trust, Txnsberg, Norway; 9) Ped Metabolic Med, Univ Childs Hospital, Hamburg, Germany; 10) Dept of Mol Paediatrics,Ludwig-Max. Univ, Munich, Germany; 11) Div of Ped Endocrinology, Univ. of Bern, Bern, Switzerland; 12) Dept of Pediatrics, Fondazione "MBBM", Monza, Italy; 13) Laboratoire de biochimie, CHU de Caen, Caen, France; 14) Dept of Neuropediatrics, Klin Bürgerpark, Bremerhaven, Germany; 15) Lab of Physiol Chem, de Duve Institute, Brussels, Belgium; 16) Dept of Neonatology, Olgahospital, Stuttgart, Germany; 17) Dept of Child Neurology, VUmc, Amsterdam, Netherlands.

The Krebs cycle has a fundamental role in the ATP production in prokaryotic and eukaryotic cells. Both enantiomers of 2-hydroxyglutarate are directly linked to this fundamental energy production pathway and both are found to be elevated in certain neoplastic malignancies. D-2- and L-2-hydroxyglutaric acidurias are rare neurometabolic disorders. We showed that one variant of 2-hydroxyglutaric aciduria - D-2-hydroxyglutarate type I - (Kranenedijk et al, Science 2010) is caused by cancer-associated IDH2 germline mutations. Complementary to this finding, we recently reported pathogenic mutations in the mitochondrial citrate transporter gene (SLC25A1) using exome sequencing in 12 individuals with the combined D-2- and L-2-hydroxyglutaric aciduria (Nota et al, april 2013, Am J Hum Genet). We demonstrated impairment of mitochondrial citrate efflux by stable isotope labeled experiments in cultured skin fibroblasts. The absence of SLC25A1 protein in fibroblasts harboring null alleles (SLC25A1-/-) was shown by western blotting and could be restored by transfection of wild type SLC25A1. Overexpression of SLC25A1 also restored citrate efflux and decreased intracellular D-2- and L-2 hydroxyglutarate levels. These results prove that the mitochondrial citrate transporter is the cause of combined D-2- and L-2 hydroxyglutaric aciduria.

243

Unprocessed RNA intermediates interfere with mitochondrial translation and cause respiratory chain deficiency. *R. Kopajtich[1], T.B. Haack[1,2], P. Freisinger[3], T. Wieland[1,2], J. Rorbach[4], T.J. Nicholls[4], E. Baruffini[5], A. Walther[1,2], K. Danhauser[1], F.A. Zimmermann[6], R.A. Husain[7], H. Mundy[8], I. Ferrero[5], T.M. Strom[1,2], T. Meitinger[1,2], R.W. Taylor[9], M. Minczuk[4], J.A. Mayr[6], H. Prokisch[1,2].* 1) Institute of Human Genetics, Helmholtz Zentrum München, Munich / Neuherberg, Germany; 2) Institute of Human Genetics, Techniche Universität München, Munich, Germany; 3) Department of Pediatrics, Klinikum Reutlingen, Reutlingen, Germany; 4) MRC Mitochondrial Biology Unit, Cambridge, United Kingdom; 5) Department of Life Sciences, University of Parma, Parma, Italy; 6) Department of Pediatrics, Paracelsus Medical University Salzburg, Salzburg, Austria; 7) Department of Neuropediatrics, Jena University Hospital, Jena, Germany; 8) Centre for Inherited Metabolic Disease, Evelina Children's Hospital, Guys and St Thomas' NHS Foundation Trust, London, United Kingdom; 9) Wellcome Trust Centre for Mitochondrial Research, Institute for Ageing and Health, The Medical School, Newcastle University, Newcastle upon Tyne, United Kingdom.

Mitochondria harbor their own genome, transcription and translation machinery. The human mitochondrial genome encodes 13 mRNAs, 22 tRNAs and 2 rRNAs to produce 13 subunits of the respiratory chain. Transcription of the mitochondrial DNA produces large polycistronic precursor transcripts that have to be precisely processed since most genes are directly adjacent or separated by only a few non-coding nucleotides. Most human mRNA and rRNA genes are flanked by tRNA genes and, according to the 't-RNA punctuation model', excision of these tRNAs leads to the liberation of the corresponding mRNAs and rRNAs. The primary transcript is processed by subsequent cleavage of RNase P and RNase Z, encoded by MRPP1,2,3 and ELAC2 respectively. By exome sequencing and subsequent mutation screening, we now identified 5 individuals from 3 different families carrying either compound heterozygous or homozygous mutations in ELAC2. All patients presented with infancy-onset hypertrophic cardiomyopathy, lactic acidosis and respiratory chain defects. Quantitative PCR demonstrated the accumulation of unprocessed RNA intermediates up to 400-fold in in skeletal muscle samples and up to 30-fold in fibroblasts. However, Northern blot analysis showed no change in steady state levels of rRNAs, mRNAs or tRNAs, and tRNAs were found to be existing in their mature form including the trinucleotide CCA 3'end. Nevertheless, the patients' cell lines displayed an impaired mitochondrial translation. In contrast to prior studies the translation defect could not be explained by changes in the abundance of individual RNA species. Finally, transcriptome analysis by RNAseq in fibroblasts showed that approximately 10% of the mitochondrial mRNAs still have a tRNA attached to their 5'-ends. Therefore we propose a model in which aberrantly processed mRNA species (with an extended 5'-terminus) fail to be effectively eliminated by the RNA surveillance machinery interfere with translation. This constitutes a new pathomechanism for mitochondrial disorders. This study highlights that next generation sequencing not only provided a molecular diagnosis for these patients but also helped to understand the underlying pathomechanism.

244

Mutations in the cytochrome c1 subunit of respiratory chain complex III cause insulin-responsive hyperglycemia and recurrent ketoacidosis.

J. Christodoulou[1,2,3,4], P. Gaignard[5], M. Menezes[3,4], M. Schiff[5,6,7], A. Bayot[6,7], M. Rak[6,7], H. Ogier de Baulny[8], C.-H. Su[9], M. Gilleron[10,11], A. Lombes[10,11], H. Abida[9], A. Tzagoloff[9], L. Riley[4], S.T. Cooper[3,12], K. Mina[13], M.R. Davis[14], R.J.N. Allcock[15,16], N. Kresoje[15], N.G. Laing[17], D.R. Thorburn[18,19], A. Slama[5], P. Rustin[6,7]. 1) Western Sydney Genetics Program, Children's Hospital at Westmead, Westmead, NSW, Australia; 2) Discipline of Genetic Medicine, Sydney Medical School, University of Sydney, Australia; 3) Discipline of Paediatrics and Child Health, Sydney Medical School, University of Sydney, Australia; 4) Genetic Metabolic Disorders Research Unit, Research Institute, Children's Hospital at Westmead, Sydney, NSW, Australia; 5) Laboratoire de Biochimie, APHP Hôpital de Bicêtre, 94275 Le Kremlin Bicêtre cedex, France; 6) INSERM UMR 676, Hôpital Robert Debré, 75019, Paris, France; 7) Université Paris 7, Faculté de Médecine Denis Diderot, Paris, France; 8) Reference Center for Inherited Metabolic Diseases, Hôpital Robert Debré, APHP, 75019, Paris, France; 9) Department of Biological Sciences, Columbia University, New York, NY 10027, USA; 10) INSERM, UMRS 1016, Institut Cochin, F-75014 Paris, France; 11) APHP, Hôpital de La Salpêtrière, Service de Biochimie Métabolique et Centre de Génétique Moléculaire et Chromosomique, 75651 Paris, France; 12) Institute for Neuroscience and Muscle Research, Children's Hospital at Westmead, Sydney, NSW, Australia,; 13) Department of Molecular Genetics, PathWest and School of Pathology and Laboratory Medicine, The University of Western Australia, Nedlands, Perth, Western Australia; 14) Neurogenetic Laboratory, Department of Anatomical Pathology, Pathwest Laboratory Medicine WA, Royal Perth Hospital, Perth, Western Australia,; 15) Lotterywest State Biomedical Facility Genomics and School of Pathology and Laboratory Medicine, University of Western Australia, Nedlands, Perth, Western Australia; 16) Department of Clinical Immunology, Pathwest Laboratory Medicine WA, Royal Perth Hospital, Perth, Western Australia; 17) Centre for Medical Research, University of Western Australia & Western Australian Institute for Medical Research, QEII Medical Centre, Nedlands, Perth, Western Australia; 18) Murdoch Childrens Research Institute and Victorian Clinical Genetics Services, Royal Children's Hospital, Flemington Road, Parkville, Melbourne Melbourne, VIC, Australia; 19) Department of Paediatrics, University of Melbourne, Melbourne, VIC, Australia.

Primary complex III (CIII) deficiency is an uncommon and clinically highly variable mitochondrial respiratory chain disorder. Using a combination of whole exome and candidate gene sequencing, we have identified mutations in the CYC1 gene (c.643C>T; p.Leu215Phe and c.288G>T; p.Trp96Cys) encoding the cytochrome c1 subunit of CIII, in two unrelated patients (Sri Lankan and Lebanese, both the offspring of consanguineous unions), presenting with recurrent episodes of severe keto- and lactic acidosis and insulin-responsive hyperglycemia. A functional deficiency of CIII in skeletal muscle was associated with reduced levels of CYC1 protein in skeletal muscle and fibroblasts from the patients, suggesting that the mutation make the CYC1 protein unstable. In addition, there were reduced levels of assembly-dependent CIII subunits (core 2), and of assembly-dependent subunits of CI (subunit NDUFB8) and CIV (subunit1), suggesting that CYC1 mutations affect the in vivo stability or assembly of the respirasome. We then examined the effect of the two CYC1 mutations on mitochondrial function in Saccharomyces cerevisiae CYT1. Transformants harboring either mutant gene in single copy failed to grow on non-fermentable carbon sources, and had only 6% of NADH-cytochrome c reductase activity compared to wild-type, confirming that the two mutations severely affect respiratory chain function. Finally, lentiviral rescue experiments in patient fibroblasts showed that wild-type CYC1 corrected complex III activity (patient 1), and resulted in significant increases in protein levels of CYC1, CIII subunits (core 2) and assembly-dependent subunits of CI and CIV (patient 2). Together, our findings demonstrate that the mutations we have identified have a deleterious effect on cytochrome c1 stability and complex III assembly, and highlight that whole exome sequencing has matured to the point where it should be considered as a first line molecular screening when one of multiple genes could be causative for a particular phenotype.

245

Mutations in *FBXL4* cause mitochondrial encephalopathy and a disorder of mitochondrial DNA maintenance.

R.W. Taylor[1], J.W. Yarham[1], A. Besse[2], P. Wu[2], E.A. Faqeih[3], A.M. Al-Asmari[3], M.A.M. Saleh[3], W. Eyaid[4], A. Hadeel[4], L. He[1], F. Smith[5], S. Yau[5], E.M. Simcox[1], S. Miwa[6], T. Donti[2], K.K. Ab-Amero[7], L-J. Wong[2], W.J. Craigen[2], B.H. Graham[2], K.L. Scott[2], R. McFarland[1], P.E. Bonnen[2,8]. 1) Wellcome Trust Centre for Mitochondrial Research, Newcastle University, Newcastle upon Tyne, United Kingdom; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 3) Section of Medical Genetics, Children's Hospital, King Fahad Medical City, Riyadh, Saudi Arabia; 4) Department of Pediatrics, King Abdulaziz Medical City, King Saud Bin Abdulaziz University for Health & Science, Riyadh, Saudi Arabia; 5) DNA Laboratory, Guy's and St Thomas' Serco Pathology, Guy's Hospital, London, United Kingdom; 6) Institute for Ageing and Health, Newcastle University, Newcastle upon Tyne, United Kingdom; 7) Ophthalmic Genetics Laboratory, Department of Ophthalmology, College of Medicine, King Saud University, Riyadh, Saudi Arabia; 8) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA.

Mitochondria play a key role in a variety of fundamental cellular processes including oxidative phosphorylation (OXPHOS), and due to their unique bigenomic heritage, faithful organellar functioning is dependent upon the coordinated expression and interaction of both nuclear and mitochondrial-encoded gene products. Paediatric-onset mitochondrial disease has an estimated incidence of 1 in 5,000 but despite this common occurrence the molecular aetiology in individual cases often remains unknown, largely reflecting the large number of nuclear genes (~1400) which comprise the mitochondrial proteome. Mutations have been described in a number of nuclear genes involved in the maintenance of mitochondrial DNA (mtDNA), ultimately leading to secondary mtDNA abnormalities in clinically-affected tissues. Nuclear genetic disorders causing mitochondrial DNA (mtDNA) depletion syndrome - early-onset childhood disorders characterised by a quantitative loss of mtDNA copy number in clinically-relevant tissues - show phenotypic and genetic heterogeneity. Through whole exome sequencing, we have identified recessive nonsense and splicing mutations in the *FBXL4* gene segregating in unrelated consanguineous kindreds in which affected children present with a fatal encephalopathy, lactic acidosis and severe mtDNA depletion in muscle. We show that FBXL4 is an F-box protein that co-localises with mitochondria, and that loss-of-function and splice mutations in this protein result in a severe respiratory chain deficiency, loss of mitochondrial membrane potential and a disturbance of the dynamic mitochondrial network and nucleoid distribution in patient fibroblasts. Expression of the wild-type *FBXL4* transcript in patient cell lines fully rescued the levels of mtDNA copy number, leading to a correction of the mitochondrial biochemical deficit. Together these data establish *FBXL4* as a new mitochondrial disease gene with a possible role in maintaining mtDNA integrity and stability.

246

A lipomatosis endophenotype in methylmalonic acidemia: evidence from patients and mice. *I. Manoli[1], J.R. Sysol[1], M.K. Crocker[2], G. Niu[3], J. Storrar[1], S. Mendelson[1,2], J.L. Sloan[1], C. Wang[1], Y. Ktena[1], P.M. Zerfas[4], V. Hoffman[4], H.J. Vernon[5], A. Hamosh[5], J.C. Reynolds[6], X. Chen[3], O. Gavrilova[7], J.A. Yanovski[2], C.P. Venditti[1].* 1) Genetics and Molecular Biology Branch, NHGRI, NIH, Bethesda, MD; 2) Program in Developmental Endocrinology and Genetics, Eunice Kennedy Shriver NICHD, NIH, Bethesda, MD; 3) Laboratory of Molecular Imaging and Nanomedicine, NIBIB, NIH, Bethesda, MD; 4) Office of Research Services, Division of Veterinary Resources, NIH, Bethesda, MD; 5) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 6) Radiology and Imaging Sciences Department, Clinical Center, NIH, Bethesda, MD; 7) Mouse Metabolism Core, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD.

We have discerned that a subset of patients with isolated methylmalonic acidemia (MMA) develop centripetal obesity, yet display a paucity of visceral fat, with wasting of distal upper and lower extremities. Measurements of whole and regional body composition, expressed as fat- (FM) and fat-free body mass (FFM), were obtained by dual energy X-ray absorptiometry (DXA) in 46 individuals with MMA (32 *mut*, 8 *cblA* and 6 *cblB* subtype) and 99 age-, gender-, race/ethnicity-, and BMI-matched controls. Compared to controls, mut^0 patients had decreased FFM (P=0.005) and increased proximal/total upper and lower extremity FM (73.6 ± 7.1 *vs.* 68.6 ± 4.8% in controls, P= 0.001 for both extremities). While serum leptin concentrations correlated with FM in both groups, the ratio of proximal/total extremity FM correlated positively with serum adiponectin levels in mut^0 patients (r=0.425, P=0.04 for upper and r=0.601, P=0.003 for lower extremity), reminiscent of the hyperadiponectinemia and proximal accumulation of subcutaneous fat observed in the multiple lipomatosis syndromes previously reported in alcoholic subjects and, rarely, in patients with primary mitochondrial disease. To model extrahepatic disease manifestations including brown/white fat pathophysiology in MMA, we generated mice that express the methylmalonyl-CoA mutase (*Mut*) cDNA ($Tg^{INS\text{-}Alb\text{-}Mut}$) under the control of a liver-specific promoter on the $Mut^{-/-}$ background. $Mut^{-/-}$;$Tg^{INS\text{-}Alb\text{-}Mut}$ are rescued from lethality and, similar to MMA patients, display decreased FFM and resting energy expenditure (P=0.05), as well as depressed thermogenesis after β3 adrenergic stimulation (P=0.001), and blunted induction of *Ucp1* and *Dio2* after 4h of cold exposure. When challenged with a high protein diet (HPD), $Mut^{-/-}$;$Tg^{INS\text{-}Alb\text{-}Mut}$ mice had suppressed ^{18}F-FDG-glucose uptake in brown fat measured by positron emission tomography (PET), associated with abnormal brown fat mitochondrial ultrastructure. The inclusion of 0.5% bezafibrate in the HPD resulted in improved survival (P=0.042), reversal of the PET findings, and increased *Ucp1* expression in the subcutaneous shoulder and inguinal fat depots of treated $Mut^{-/-}$;$Tg^{INS\text{-}Alb\text{-}Mut}$ mice. Our findings delineate a novel obesity phenotype in MMA distinct from other lipodystrophy syndromes associated with mitochondrial dysfunction, identify regions of inducible subcutaneous fat in mice and patients and suggest new therapeutic targets for this disease.

247

MOONLIGHTING IN MITOCHONDRIA: ACAD9 PLAYS A DUAL ROLE IN ENERGY METABOLISM. *M. Schiff[1], B.M. Haberberger[2], E.S. Goetzman[1], A.W. Mohsen[1], H. Prokisch[2], J. Vockley[1].* 1) Pediatrics, University of Pittsburgh, Pittsburgh, PA; 2) Human Genetics, Technical University and Helmholtz Zentrum, University of Munich, Germany.

Background: Liver Acyl-CoA dehydrogenase-9 (ACAD9) deficiency was first described in 2 patients with liver failure. Subsequently, point mutations in ACAD9 were shown to cause respiratory chain complex I (CI) deficiency due to a moonlighting function of ACAD9 as a CI assembly factor. However, the impact of these mutations on fatty acid oxidation (FAO) activity and the physiologic contribution of ACAD9 in eukaryotes to FAO remain unknown. Methods: Prokaryotic expression mutagenesis studies were performed for 13 missense ACAD9 mutations. Oxidation of long chain fatty acids was analyzed in very long-chain acyl-CoA dehydrogenase (VLCAD) and long-chain acyl-CoA dehydrogenase (LCAD) double knockout mouse fibroblasts. ACAD9 gene knockout in HEK293 cells was performed using Transcription Activator-like Effector Nucleases (TALEN) technology. Results: All but two ACAD9 mutants (R127Q, A326P) had a stable ACAD9 antigen. Three (R469W, R518H, R518C) had mildly (75% of wild type) decreased ACAD9 enzymatic activity while one had normal activity (R532W); activity was partially (20% of wild type) decreased in 2 mutants (L98S and R433Q) and undetectable in 7 (I87N, R127Q, R266Q, A326P, E413K, R417C and D418G). ACAD9 protein stability in patient fibroblasts largely mirrored these findings though some discordance was apparent. Doubly deficient VLCAD/LCAD knockout mouse fibroblasts retained ACAD9 activity and supported significant capacity for long-chain FAO. TALEN-mediated knockout HEK293 cells exhibited a total absence of ACAD9 antigen as well as defective FAO. Conclusions: The moonlighting function of ACAD9 as a CI assembly factor is independent of its FAO enzymatic activity. Mutations in ACAD9 can affect either or both functions. We also demonstrated that ACAD9 plays a significant role in eukaryotic FAO. Further characterization of genotype/phenotype correlations in ACAD9 deficient patients will be complicated by this dual function.

248

Primary ovarian insufficiency is caused by recessive partial loss-of-function mutations in genes for mitochondrial protein homeostasis. *S.B. Pierce[1], T. Walsh[1], S. Gulsuner[1], M.K. Lee[1], M.-C. King[1,2].* 1) Medicine (Medical Genetics), University of Washington, Seattle, WA; 2) Genome Sciences, University of Washington, Seattle, WA.

Primary ovarian insufficiency (POI), a form of hypergonadotropic hypogonadism, causes infertility in approximately 1% of women of normal 46,XX karyotype by 40 years of age. Both genetic and non-genetic causes of POI have been identified, but the mechanism of ovarian dysfunction in most cases is unknown. Heretofore, mutations in 14 genes have been reported to account for about 15% of cases. In a search for additional genes responsible for POI, we identified families with both POI in young women and sensorineural hearing loss (SNHL) in both females and males, described clinically as Perrault syndrome. Our participants are 15 families with between one and six affected relatives. In 5 of the 15 families, the POI and SNHL phenotypes are explained by recessive mutations in *HARS2*, *LARS2*, or *HSD17B4*. *HARS2* and *LARS2* encode mitochondrial aminoacyl-tRNA synthetases, which are required for mitochondrial protein translation. Together with *CLPP*, which encodes a mitochondrial ATP-dependent protease and also causes POI and SNHL, mutations in *HARS2* and *LARS2* reveal a role for mitochondria in the maintenance of ovarian function. Although mutations in genes encoding mitochondrial proteins have not previously been implicated in POI, our findings are consistent with increasing evidence that mitochondrial integrity is critical for oocyte competence and subsequent embryonic development. Thus, our focus on POI in association with another highly vulnerable phenotype, in this case hearing loss, reveals a specific functional class of genes as critical for fertility. Among individuals in our series carrying mutations in *HARS2*, *LARS2*, and *HSD17B4*, none carry two completely null alleles. This is consistent with the essential activities of these genes, for which complete loss of function is likely to be incompatible with life. Indeed, more severe recessive mutations in *HSD17B4* are known to cause DBP deficiency, a disorder of peroxisomal fatty acid β-oxidation that is usually fatal in early childhood. Similarly, mutations in *CHD7* have been implicated in POI, whereas more severe alleles cause CHARGE syndrome. Our approach of evaluating families affected by both POI and SNHL has resulted in the identification of new genes harboring mutations for POI and has revealed mutations in mitochondrial protein homeostasis as critical for maintenance of fertility. Gene discovery for the remaining families is in progress.

249

mtDNA mutations variously impact mtDNA maintenance throughout the human embryofetal development. *S. RONDEAU[1], S. MONNOT[1], P. VACHIN[1], E. HERZOG[1], B. BESSIERES[1], N. GIGAREL[1], D. SAMUELS[2], L. HESTERS[3], N. FRYDMAN[3], G. CHALOUHI[4], S. GOBIN LIMBALLE[1], M. RIO[1], A. ROTIG[1], A-S. LEBRE[1], A. BENACHI[5], L. SALOMON[4], A. MUNNICH[1], J-P. BONNEFONT[1], J. STEFFANN[1].* 1) Genetics, INSERM U781, Necker-Enfants Malades Hospital, Paris, France; 2) Center for Human Genetics Research, Department of Molecular Physiology and Biophysics, Vanderbilt University Medical Center, Nashville, USA; 3) Reproductive Genetics, Antoine-Beclere Hospital, Clamart, France; 4) Obstetrics, Necker-Enfants Malades Hospital, Paris, France; 5) Obstetrics, Antoine-Beclere Hospital, Clamart, France.

Mitochondrial DNA (mtDNA) mutations cause serious disorders with maternal inheritance and a high transmission risk, resulting in common requests for preimplantation or prenatal diagnoses. These procedures are hampered by the lack of data on the pathophysiology of mtDNA mutations during human development. Specifically, it is not known whether and how mtDNA mutations impact the overall mtDNA content throughout human embryofetogenesis. We collected oocytes, preimplantation embryos, whole placentas and fetal tissues at various stages of development, from controls and carriers of m.3243A>G (MTTL1), m.8344A>G (MTTK) and m.8993T>G (MTATP6), responsible for MELAS, MERRF, and NARP syndromes, respectively. We devised a test assessing simultaneously mtDNA copy number (CN) and mutant load in single cells. mtDNA CN i)was identical in control oocytes and embryos, in agreement with mtDNA replication silencing during the first embryonic cleavages; ii)gradually increased from the germinal vesicle to the blastocyst stage in m.3243A>G cells ($p<0.01$), suggestive of a mutation-dependent induction of mtDNA replication; iii)correlated with the m.3243A>G mutant load ($p<0.001$), suggesting some compensation for the respiratory chain dysfunction; iv)was identical in m.8344A>G vs control embryos, indicating that modulation of the mtDNA CN depends on the mutation type. Analyses of placentas (multiple biopsies) showed that mtDNA CN i)did not vary within a placenta, ii)significantly increased in m.3243A>G vs control at 11-GW ($p <0.01$), iii)decreased thereafter, becoming identical to controls at delivery. These data could be accounted for by a placental energy demand maximal at the end of the 1st trimester. Analyses of 7 fetal tissues (12-22 GW) showed that mtDNA CN was i)similar in all control tissues apart from the heart ($p<0.01$); ii)similar in m.3243A>G vs control tissues apart from the lung ($p<0.01$); iii)lower in m.8993T>G muscle, heart, and liver vs control ($p=0.01$, 0.01, 0.001, respectively). Mutant loads were identical in all tissues from a given fetus, indicating an absence of mutant load/mtDNA CN correlation. These data highlight the complex relationships between mtDNA mutations and mtDNA content, depending on mutation types, mutant loads, cell types and development stages. Transcriptome studies and measurement of mtDNA replication should help us in unravelling the molecular bases of these observations, of particular relevance for therapeutic approaches in mtDNA disorders.

250

Selecting likely causal genes, pathways and relevant tissues from genome-wide association studies of complex traits by data-driven expression-prioritized integration. *TH. Pers[1,2,3], J. Karjalainen[4], JN. Hirschhorn[1,2,5], L. Franke[4], the Genetic Investigation of ANthropometric Traits (GIANT) Consortium.* 1) Division of Endocrinology, Children's Hospital Boston, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, USA; 3) Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark; 4) University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands; 5) Harvard Medical School, Boston, USA.

Genome-wide association studies (GWAS) continue to identify thousands of loci where common variants are associated with complex traits. Many of these loci have no single obviously causal gene; therefore the main challenge for gaining novel biological insight is to identify which gene at each locus most likely explains the association. Because functional follow-up studies are often intensive, a key first step is to use computational approaches to prioritize genes with respect to their biological relevance. Previous computational approaches have shown some success but often focus on single types of data, limiting their discriminatory power. We have developed an approach called DEPICT (Data-driven Expression-Prioritized Integration for Complex Traits) that integrates complementary data types (including 77,840 expression microarrays, 169,810 protein-protein interactions, 211,882 gene-phenotype pairs from mouse knock-out studies, and 6,004 gene sets from pathway databases) to systematically identify the most likely (1) causal gene at a given locus, (2) pathways that are enriched in genetic associations, and (3) tissues in which genes from associated loci are highly expressed. We applied DEPICT to multiple GWAS data sets, including data from the GIANT consortium for height, body-mass index (BMI) and waist-hip ratio adjusted for BMI. The method identifies enrichment of associated genes expressed in different relevant tissues corresponding to the different traits/diseases (e.g., cartilage for height, central nervous system tissues for BMI, adipose tissues for waist-hip ratio; and lymphoid tissue for inflammatory bowel disease, IBD). For the anthropometric traits, DEPICT also identifies more statistically significantly enriched pathways than MAGENTA, another gene set enrichment tool, and many of them overlap with relevant biology. We benchmarked DEPICT further using height and IBD results along with receiver operating statistics area under the curve statistics and show that the method outperforms DAPPLE and GRAIL, two commonly used GWAS data gene prioritization methods. As unbiased benchmarks, we tested for enrichment of genes that were differentially expressed in murine growth plates (DEPICT=0.79, GRAIL=0.67, DAPPLE=0.62) and genes that were transcriptionally regulated by IBD-associated markers in blood based on expression quantitatively trait locus data meta-analysis of 5,311 individuals (DEPICT=0.74, GRAIL=0.66, DAPPLE=0.64).

251

Non-targeted metabolite profiling in large human population-based studies: a new data analysis workflow and metabolome-wide association study of C-reactive protein. *A. Ganna[1], T. Fall[2], W. Lee[3], C.D. Broeckling[4], J. Kumar[2], S. Hägg[1,2], P.K.E. Magnusson[1], J.E. Prenni[4], L. Lind[5], Y. Pawitan[1], E. Ingelsson[2].* 1) Department of medical epidemiology and biostatistics, Karolinska institutet, Stockholm, Sweden; 2) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 3) Department of Statistics, Inha University, Incheon, Korea; 4) Proteomics and Metabolomics Facility, Colorado State University, Fort Collins, Colorado, U.S.A; 5) Department of Medical Sciences, Uppsala University, Uppsala, Sweden.

Background: Recently, the potential of metabolomics in medical and pharmacological research has been illustrated through the investigation of genotype-metabolite or metabolite-phenotype associations in several population-based studies. However, the majority of these studies have been performed with small sample sizes and/or in a targeted manner involving a biased analysis of a pre-determined panel of metabolites. Objectives: Our aims were:(1) To illustrate a new data analysis workflow for detection and annotation of metabolites in large human population-based studies;(2) To illustrate a real application of the described workflow to serum samples from 2,380 fasting individuals and to conduct a non-targeted metabolome-wide association study of high-sensitive C-reactive protein (hsCRP) levels. Methods: Samples were analyzed using ultra-performance liquid chromatography coupled with mass spectrometry (UPLC-MS). The workflow is comprised of four modules:(1) Peaks from each chromatogram are detected, aligned and grouped across samples. Each peak group is called a 'feature'.(2) Feature intensities are log-transformed and normalized. Outliers are excluded. Factors of unwanted variation are identified and removed.(3) Features associated with the outcome are identified through univariate statistical analysis. False discovery rate is controlled to select features for replication in an independent validation study.(4) MS and MS/MS spectra are generated using an indiscriminate data acquisition workflow coupled with correlational grouping and used to annotate significant features through spectral matching against both private and public spectral libraries. Confidence levels are assigned to define the quality of the metabolite annotation. Results: Using the described workflow, we identified > 8,000 molecular features in serum samples from two population-based studies of 2,380 participants. We performed a non-targeted metabolome-wide association analysis of hsCRP and identified 439 features corresponding to 101 unique metabolites that could be replicated in an external population. Ten metabolites were annotated with high confidence. Our results revealed unexpected biological associations, such as metabolites annotated as monoacylphosphorylcholines (LysoPC) being negatively associated with hsCRP. Conclusions: The workflow and results presented illustrate the viability and potential of non-targeted metabolite profiling in large population-based studies.

252

Transcription factor and chromatin features predict genes associated with eQTLs. *D. Wang[1], A. Rendon[2], L. Wernisch[3].* 1) Ontario Cancer Institute, Princess Margaret Hospital, Toronto, Ontario, Canada; 2) Department of Haematology, University of Cambridge, Cambridge, UK; 3) MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK.

Cell type-specific gene expression in humans involves complex interactions between regulatory factors and DNA at enhancers and promoters. Mapping studies for expression quantitative trait loci (eQTLs), transcription factors (TFs) and chromatin markers have become widely used tools for identifying gene regulatory elements, but prediction of target genes remains a major challenge. Here, we integrate genome-wide data on TF-binding sites, chromatin markers and functional annotations to predict genes associated with human eQTLs. Using the random forest classifier, we found that genomic proximity plus five TF and chromatin features are able to predict greater than 90% of target genes within 1 megabase of eQTLs. Despite being regularly used to map target genes, proximity is not a good indicator of eQTL targets for genes 150 kilobases away, but insulators, TF co-occurrence, open chromatin and functional similarities between TFs and genes are better indicators. Using all six features in the classifier achieved an area under the specificity and sensitivity curve of 0.91, much better compared with at most 0.75 for using any single feature. We hope this study will not only provide validation of eQTL-mapping studies, but also provide insight into the molecular mechanisms explaining how genetic variation can influence gene expression.

253

Combining regulatory domain and genetic variation information to identify cell types, regulatory elements, and causal genetic variants that influence human disease. *E. Schmidt[1], J. Chen[1], C. Willer[1,2], Metabochip GIANT-BMI and ICBP.* 1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Cardiovascular Medicine, University of Michigan, Ann Arbor, MI.

Examining trait-associated variants from genome wide association studies relative to genomic regions of functional importance can give insight into the molecular mechanisms leading to disease phenotypes. We hypothesize that there are particular cell types in which trait-associated variants impact transcriptional regulation and investigate whether these variants are enriched in regulatory domains in relevant cell types. In addition, we aim to prioritize GWAS variants or their proxies as functional candidates based on overlap with regulatory domains.

We evaluate enrichment of GWAS SNPs for blood pressure, body mass index, coronary artery disease, lipids, and type 2 diabetes. For each trait, we group variants in LD with trait-associated index SNPs and determine overlap of these SNPs with DNase hypersensitivity sites from 213 cell types obtained from the ENCODE and Roadmap Epigenomics Projects. BED files containing regions of chromatin accessibility identified by DNase-seq are used to identify overlap. We compare the observed overlap with permuted sets of SNPs from the 1000 Genomes data, which match index SNPs for: i) number of SNPs in high LD (r^2>0.7), ii) exact minor allele frequency, and iii) exact distance to nearest gene. For cell types in which we see significant enrichment in DNase HS sites, we further investigate enrichment using functional elements such as histone methylation marks, FAIRE and ChIP-seq TF binding from ENCODE as well as functional chromatin states defined by Ernst 2011. Lastly, we annotate individual SNP overlap with significant regulatory marks examined above as well as expression quantitative trait loci in relevant tissues.

We find evidence of significant enrichment in DNase HS sites for each set of trait-associated variants tested: lipid variants in leukemia cells (P=2×10^{-14}); BMI variants in olfactory neurosphere-derived cells (P=2×10^{-5}); BP variants in osteoblasts (P=2×10^{-7}); CAD variants in hepatocellular carcinoma cells (P=6×10^{-7}); and T2D variants in colorectal carcinoma cells (P=1×10^{-13}). Our method can be applied to other traits to: i) characterize cell type where GWAS variants may exert their effect; ii) identify regulatory elements or TFs that may be impacted; and iii) fine-map or prioritize functional variants at specific loci.

254

Genome-wide Expression Quantitative Trait Loci: Results from the NHLBI's SABRe CVD Initiative. *R. Joehanes[1,2], T. Huan[1], C. Yao[1], X. Zhang[1], S. Ying[2], M. Feolo[3], N. Sharopova[3], T. Przytycka[3], A. Sturcke[3], A.A. Schaffer[3], N. Heard-Costa[4], H. Chen[6], P. Liu[5], R. Wang[5], K.A. Woodhouse[5], N. Raghavachari[5], J. Dupuis[4,6], A.D. Johnson[1], C.J. O'Donnell[1,7], P.J. Munson[2], D. Levy[1].* 1) Division of Intramural Research, National Heart, Lung and Blood Institute; the NHLBI's Framingham Heart Study, National Institutes of Health, Bethesda, MD, USA; 2) Mathematical and Statistical Computing Laboratory, Center of Information Technology, National Institutes of Health, Bethesda, MD, USA; 3) National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA; 4) School of Public Health, Boston University, Boston, MA, USA; 5) DNA Sequencing and Genomics Core, National Institutes of Health, Bethesda, MD, USA; 6) Department of Biostatistics, Boston University School of Public Health; 7) Division of Cardiology, Massachusetts General Hospital, Boston, MA, USA.

Introduction: Expression quantitative trait loci (eQTL) analysis has helped validate genotype-trait associations identified through genome-wide association studies (GWAS). Several eQTL studies have been performed, but many are derived from relatively small sample sizes. Methods: In this study, we performed an eQTL analysis using genotype and gene transcriptomic data from 5,257 Framingham Heart Study participants, the largest eQTL study to date. In total, we analyzed 39,315,185 1000G-imputed single nucleotide polymorphisms (SNPs) in association with whole blood-derived expression levels of 283,805 exons and 17,873 genes, measured in the Affymetrix Human Exon Array 1.0 ST platform. We used an additive regression model, adjusted for sex, age, family structure, complete blood counts, and technical covariates, such as batch. Cis eQTL is defined as eQTL 1 MB up and downstream of the corresponding transcript location. Results: At a false discovery rate <0.05, at the exon level, we identified 125,219,884 cis-eQTL and 64,066,764 trans-eQTL-exon pairs (21,694,034 unique SNPs and 282,730 unique exons). At the gene level, we detected 3,423,257 cis-eQTL and 36,218,111 trans-eQTL-gene pairs (9,435,973 unique SNPs and 17,873 unique genes). Of these, 3,447 exon-level and 2,172 gene-level eQTLs are reported as trait-associated SNPs in the NHGRI GWAS catalog, giving functional support to the GWAS results of various phenotypes. Among such phenotypes, platelet count and mean platelet volume are associated with several of the strongest eQTLs, such as rs1354034, rs16971217, rs10512472, rs12485738, rs505404, rs11602954, and rs17655730. Of these, rs1354034 in ARHGEF3 is a strong trans-eQTL that is associated with expression of at least 55 other transcripts across the genome. This finding lends support to the hypothesis that ARHGEF3 mRNA interactions are important in thrombopoiesis. In addition, we found strong eQTLs for 567 other phenotypes. For example, rs2247056 is a strong eQTL for triglycerides, rs3131379 for systemic lupus erythematosus, rs3916765 for type 2 diabetes, and rs3117582 for lung cancer. Our results will be available in NCBI GTEx database. Conclusion: Our eQTL results can be used to suggest functional elements relevant to numerous phenotypes. Our results also suggest that many of the SNPs identified by GWAS studies exert their influence to their respective phenotypes through mRNA expression levels of genes that may be distant from the SNP locations.

255

A hierarchical multiscale model to infer transcription factor occupancy from chromatin accessibility data. *A. Raj[1], H. Shim[1], Y. Gilad[1], M. Stephens[1,2], J. Pritchard[1,3].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Statistics, University of Chicago, Chicago, IL; 3) Howard Hughes Medical Institute, University of Chicago, Chicago, IL.

Understanding global gene regulation critically depends on accurate annotation of regulatory elements that are functional in a given cell type. CENTIPEDE, a powerful, probabilistic framework for identifying transcription factor binding sites from tissue-specific DNase I cleavage patterns and genomic sequence content, leverages the hypersensitivity of DNA-bound sites and the information in the DNase I footprint characteristic of each DNA binding protein to accurately infer functional factor binding sites. However, this framework assumes that conditional on binding, the DNase hypersensitivity at a pair of genomic locations around the binding motif are independent; an assumption that is biologically unrealistic and is not supported by DNase I data at motifs with high Chip-Seq read depth for several different factors. In this work, we adapt a Bayesian multiscale modeling framework for Poisson processes to better model the underlying spatially structured DNase I cleavage pattern induced by the binding of a particular transcription factor. In comparison to results from CENTIPEDE, the factor-specific footprint inferred using this hierarchical model tends to be smoother and the confidence of factor binding at putative binding motifs shows improved correlation with the occupancy of that factor quantified by its Chip-Seq signal. Furthermore, we demonstrate the improved area under Receiver Operating Characteristics of this model for several transcription factors by comparing against the Chip-Seq peaks for those factors identified using MACS. Finally, we show that a straightforward extension of these models to genomic locations containing motifs with low position-weight matrix scores identifies several high-confidence binding sites, increasing the precision-recall characteristics of the learning algorithm.

256

Development of a methods-based proficiency test for next generation sequencing (NGS). *N. Aziz[1], J. Durtschi[2], Q. Zhao[1], L. Bry[12], D. Driscoll[1], J. Gibson[15], W. Grody[14], M. Hedge[10], G. Hoeltge[16, 17, 18], D. Leonard[13], J. Merker[6], L. Palicki[1], R.S. Robetorye[7], I. Schrijver[9], K. Weck[8], T. Hambuch[4], T. Harkins[5], D. Ballinger[3], K. Voelkerding[2, 11].* 1) Transformation Program Office, College of American Pathologists (CAP); 2) ARUP Laboratories; 3) Complete Genomics; 4) Illumina; 5) Life Technologies; 6) Department of Pathology, Stanford University School of Medicine; 7) Department of Laboratory Medicine & Pathology, Mayo Clinic Hospital; 8) Departments of Pathology & Laboratory Medicine and Genetics, University of North Carolina at Chapel Hill; 9) Stanford University School of Medicine, Departments of Pathology and Pediatrics; 10) Emory University School of Medicine; 11) Department of Pathology, University of Utah Department of Medicine; 12) Dept. Pathology, Brigham & Women's Hospital, Harvard Medical School; 13) Department of Pathology, University of Vermont College of Medicine; 14) UCLA; 15) University of Central Florida College of Medicine;; 16) Robert J Tomsich Pathology & Laboratory Medicine Institute; 17) Cleveland Clinic; 18) Cleveland Ohio.

The higher throughput and lower cost of NGS has led to its rapid adoption in clinical testing. However, NGS-based tests are of much higher complexity than Sanger sequencing-based tests because of the requirements for extensive data analysis. In addition, there are challenges related to the acquisition and storage of data that far exceed those commonly generated in clinical laboratories. NGS tests are currently only offered as laboratory developed tests (LDTs) and require the development of clinical lab standards. The College of American Pathologists (CAP) recently developed accreditation requirements for labs offering NGS-based tests. Another mechanism of quality management mandated by CLIA is the requirement for routine proficiency testing. Here, we describe the development of a methods-based proficiency test (PT) developed at CAP that will be used to the monitor the proficiency of clinical labs offering NGS-based tests. The PT was designed to be suitable for all clinical labs using a variety of sequencing platforms and test applications and is appropriate for those offering either genome, exome, or gene panel tests. In order to create this PT product, we obtained genomic DNA from an established cell line derived from a healthy, appropriately consented individual. This “CAP genome” was sequenced at high depth by using three different technologies: Illumina (IL), Complete Genomics (CGI), and Ion Torrent (IT). There were 2 runs at the genome level using IL Hiseq (mean coverage [MC] 54X) and CGI (MC 44X) and 2 runs at the exome level using the IT PGM (MC 64X) at Life Technologies and IL HiSeq (MC 144X) at ARUP. The variants obtained using each center’s internally developed pipeline are being analyzed for concordance to create a master list of variants that will consist of SNVs and indels. The pilot PT will query 200 concordant loci and will be a mix of SNVs, indels and wild type sequences. Participants will be sent genomic DNA to test their wet bench and bioinformatics pipelines and ability to correctly call variants. The result of our first pass analysis for characterizing the CAP genome for the 4 sequencing runs for exonic regions (coding, 5’&3’UTRs) showed 3827 (CGI), 1835 (IL-genome), 3495 (IL-exome), and 713 (IT-exome) SNVs that were unique for each run, with a total number of 21,575 concordant SNVs for all four runs. As expected, there is a lower concordance rate for indels. This methods-based PT will help ensure quality in clinical testing by NGS.

257

An integrated nexus of >12,000 genome sequences and analysis tools facilitates novel gene discovery. *J. Reid[1], A. Carroll[2], N. Veeraraghavan[1], C. Gonzaga-Jauregui[3], A. Morrison[4], T. Gambin[3], A. Sundquist[2], M. Bainbridge[1], M. Dahdouli[1], Z. Huang[1], A. Li[4], F. Yu[1], R. Daly[2], J. Lupski[3], G. Duyk[2], R. Gibbs[1,3], E. Boerwinkle[1,4].* 1) Human Genome Sequencing Ctr, Baylor College Med, Houston, TX; 2) DNAnexus, Mountain View, CA; 3) Dept of Molecular and Human Genetics, Baylor College Med, Houston, TX; 4) Human Genetics Ctr, Univ of Texas Health Science Ctr at Houston, Houston, TX.

Discovery of the genetic causes of human disease is a first step toward personalized medicine, predictive diagnostics, and drug development. Accomplishing this goal requires genome sequence from an informative sample of patients with the same diagnosis, a large sample of healthy individuals to serve as a comparison or filtering group, and a suite of informatics tools to take the data from raw sequence to study results. Creating such a nexus of data and analysis tools requires considerable resources, which may be inefficient to repeatedly recreate in a multitude of academic or clinical laboratories.

To address the analysis requirements of both the high-throughput research environment and rapidly growing clinical sequencing efforts, we have developed the Mercury data processing pipeline -- an automated, flexible, and extensible analysis workflow designed to provide accuracy and reproducibility to a sequencing pipeline. Using Mercury, we have created a resource of 2,000 patients and 10,000 healthy individuals with exome and whole genome sequence data and a suite of easy-to-use analysis tools for the purpose of promoting biomedical research, particularly novel gene discovery. By leveraging cloud computing technologies via the DNAnexus platform, this collaborative resource is scalable, extensible, and compliant with clinical security standards (including ISO 27001 certification). Inter-operability, data standards, and an intuitive interface facilitate efficient data and tool sharing.

We have used such a resource for discovery of multiple disease genes for Mendelian disorders and complex disease risk factors. Three use cases will be presented in detail. In the first case, a diagnosis was made in a patient with Bohring-Opitz-like syndrome that was only possible because of having a large database of similarly affected individuals and an informative control/filtering set. In the second case, we were able to create a consortium of Charcot-Marie Tooth investigators to evaluate the spectrum of phenotypic expression for patients having the same or similar disease mutations. In the last case, we demonstrate that detailed performance metrics available in the cloud make these data resources an ideal platform for comparing and benchmarking analysis tools.

258

Pulling out the 1%: Whole-Genome In-Solution (WISC) capture for the targeted enrichment of ancient DNA sequencing libraries. *C.D. Bustamante[1], M.L. Carpenter[1], J.D. Buenrostro[1], C. Valdiosera[2,3], H. Schroeder[2], M.E. Allentoft[2], M. Sikora[1], M. Rasmussen[2], S. Guillén[4], G. Nekhrizov[5], K. Leshkatov[6], D. Dimitrova[5], N. Theodossiev[6], D. Pettener[7], D. Luiselli[7], A.E. Moreno[1], S. Gravel[8], Y. Li[8], J. Wang[8,9,10,11], M.T.P. Gilbert[2,12], E. Willerslev[2], W.J. Greenleaf[1].* 1) Genetics, Stanford University School of Medicine, Stanford, CA; 2) Natural History Museum of Denmark, Centre for GeoGenetics, Copenhagen, DK-1350 Denmark; 3) Department of Archaeology, Environment, and Community Planning, Faculty of Humanities and Social Sciences, La Trobe University, Melbourne, VIC 3086, Melbourne, Australia; 4) Centro Mallqui, Calle Ugarte y Moscoso 165, San Isidro, Lima 27, Peru; 5) Bulgarian Academy of Sciences, National Institute of Archaeology, Sofia, 1000, Bulgaria; 6) Department of Archaeology, Sofia University, Sofia, 1504, Bulgaria; 7) Department of Biological, Geological, and Environmental Sciences, University of Bologna, Bologna, Italy; 8) BGI-Shenzhen, Shenzhen 518083, China; 9) King Abdulaziz University, Jeddah, Saudi Arabia; 10) Department of Biology, University of Copenhagen, Copenhagen, Denmark; 11) Macau University of Science and Technology, Macau, China; 12) Ancient DNA Laboratory, Murdoch University, South St, 6150 Western Australia, Australia.

The very low levels of endogenous DNA remaining in most ancient specimens has precluded the shotgun sequencing of many interesting samples due to cost. For example, ancient DNA (aDNA) libraries derived from bones and teeth often contain <1% endogenous DNA, meaning that the majority of sequencing capacity is taken up by environmental DNA. We will present a method for the targeted enrichment of the endogenous component of human aDNA sequencing libraries. Using biotinylated RNA baits transcribed from genomic DNA libraries, we are able to significantly enrich for human-derived DNA fragments. This approach, which we call whole-genome in-solution capture (WISC), allows us to obtain genome-wide ancestral information from ancient samples with very low endogenous DNA contents. We demonstrate WISC on libraries created from four Iron Age and Bronze Age human teeth from Bulgaria, as well as bone samples from seven Peruvian mummies and a Bronze Age hair sample from Denmark. Prior to capture, shotgun sequencing of these libraries yielded an average of 1.2% of reads mapping to the human genome (including duplicates). After capture, this fraction increased dramatically, with up to 59% of reads mapped to human and folds enrichment ranging from 5X to 139X. Furthermore, we maintained coverage of the majority of fragments present in the pre-capture library. Intersection with the 1000 Genomes Project reference panel yielded an average of 50,723 SNPs (range 3,062-147,243) for the post-capture libraries sequenced with 1 million reads, compared with 13,280 SNPs (range 217-73,266) for the pre-capture libraries, increasing resolution in population genetic analyses. We will also present the results of performing WISC on other aDNA libraries from both archaic human and non-human samples, including ancient domestic dog samples. Our capture approach is flexible and cost-effective, allowing researchers to access aDNA from many specimens that were previously unsuitable for sequencing. Furthermore, this method has applications in other contexts, such as the enrichment of target human DNA in forensic samples.

259

Network Analysis of Mutations Across Cancer Types. *M.DM. Leiserson[1], F. Vandin[1], H-T. Wu[1], J. Dobson[1,2], A. Gonzalez-Perez[3], D. Tamborero[3], N. Lopez-Bigas[3], B. Raphael[1].* 1) Department of Computer Science, and Center for Computational Molecular Biology, Brown University, Providence, RI; 2) Department of Molecular Biology, Cell Biology & Biochemistry, Brown University, Providence, RI; 3) Experimental and Health Science Department, University Pompeu Fabra, Barcelona, Spain.

Recent large-scale cancer sequencing studies have revealed extensive mutational heterogeneity across samples from the same cancer type with relatively few genes recurrently mutated across many samples and a “long tail” of many genes, each mutated in a small number of samples. A major reason for this heterogeneity is that cancer mutations target cellular signaling and regulatory pathways, and different combinations of mutations may perturb these pathways in different individuals. This complicates efforts to distinguish driver from passenger mutations according to their observed frequencies.

We introduce a new algorithm, HotNet2, to identify subnetworks of a protein-protein (or protein-DNA) interaction network that are mutated in a statistically significant number of samples. HotNet2 uses an “insulated” heat diffusion model to simultaneously assess both the significance of mutations in individual proteins and the local topology of a protein’s interactions. Thus, HotNet2 is not restricted to analysis of known gene sets, but can discover novel combinations of interacting genes. We score the significance of individual proteins using: (1) the frequency of mutation; (2) the accumulation of mutations with high functional impact across tumor samples computed by Oncodrive-FM (Gonzalez-Perez and Lopez-Bigas 2012); (3) bias towards the gene misregulation due to copy number aberrations computed by Oncodrive-CIS (Tamborero et al. 2013).

We used HotNet2 to perform a pan-cancer analysis of mutated networks using whole-exome sequencing and copy number aberration data from 2866 samples from The Cancer Genome Atlas (TCGA) across twelve different cancer types. We used two networks: one consisting of high quality protein-protein interactions (Das and Yu 2012) and a second consisting of a variety of interaction types (Khurana et al. 2013). We identified >10 significantly mutated subnetworks ($P < 0.01$) in each case. These subnetworks overlap well-known cancer signaling pathways (e.g. p53, RTK, and RB), but also include subnetworks with less characterized roles in cancer; e.g. the cohesin and condensin complexes and the SLIT-ROBO pathway, the latter involved in cell migration. Several of these subnetworks are significantly enriched for mutations in a specific cancer type (e.g. chromatin related genes including ARID1A and PBRM1 in renal cell carcinoma).

260

The landscape of tumor suppressors in primary tumors. *P. Van Loo[1,2,3], J. Cheng[4,5], H.K.M. Vollan[6,7,8].* 1) Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 2) Center for the Biology of Disease, VIB, Leuven, Belgium; 3) Department of Human Genetics, University of Leuven, Leuven, Belgium; 4) SCD-SISTA, Department of Electrical Engineering - ESAT, University of Leuven, Leuven, Belgium; 5) iMinds Future Health Department, University of Leuven, Leuven, Belgium; 6) Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway; 7) The K.G. Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway; 8) Department of Oncology, Division of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital, Oslo, Norway.

Homozygous deletions are rare in cancers and often target tumor suppressors. We are performing a systematic screen of homozygous deletions in a compendium of 14,000 human cancers, aiming to identify known and novel tumor suppressors. Such a systematic screen has so far been performed only in cell lines, as non-aberrant cell infiltration has historically hindered the reliable identification of homozygous deletions in primary tumors. Results on a pilot series of 2218 primary tumors across 12 human cancers identified 89 genomic regions recurrently targeted by homozygous deletions. These recurrent homozygous deletions occurred either over tumor suppressors or over regions of increased genomic instability called fragile sites. We constructed a statistical model that separates fragile sites from regions showing signatures of positive selection for homozygous deletions that therefore are likely to contain tumor suppressors. From the patterns of the homozygous deletions, the patterns of mutations in COSMIC and the literature, we subsequently identified candidate tumor suppressors.

Our pilot screen identified 11 established tumor suppressors and discovered 25 candidate tumor suppressors. Our results add to the emerging evidence of several genes recently proposed in the literature as tumor suppressors (including *FAT1*, *BIRC2*/*BIRC3*, *TET1* and *MGMT*), and bring forward several novel candidate tumor suppressors (including *CASP3*, *CASP9*, *RAD17*, *BAZ1A*, *CPEB3* and *SETD1B*). Interestingly, we observed four peaks of homozygous deletions around the ubiquitin specific proteases *USP34*, *USP29*, *USP44* and the neighboring genes *USP8* and *USP50*, predominantly in lung cancer. These are all family members of a large group of deubiquitinating enzymes that prevent protein degradation by the proteasome. Interestingly, p53 has been described as a target of USP29 and USP34 has been linked to Wnt signaling. In addition, USP44 regulates the mitotic cell cycle checkpoint and *Usp44* knockout mice spontaneously formed lung tumors. Therefore, our results further establish *USP44* as a tumor suppressor in lung cancer, and implicate other ubiquitin specific proteases in oncogenesis as well.

In summary, through copy-number meta-analysis of large-genomic datasets, our study is significantly advancing the landscape of tumor suppressors.

261

Recurrent somatic mutation altering DNA-binding motif of transcription factor YY1 explains pathogenesis of insulin-producing adenomas. *M.K. Cromer[1,2], M. Choi[1,2], C. Nelson-Williams[1,2], A.L. Fonseca[3], J.W. Kunstman[3], R.M. Korah[3], J.O. Overton[1,4], S. Mane[1,4], B. Kenney[5], C.D. Malchoff[6], P. Stalberg[7], G. Akerström[7], G. Westin[7], P. Hellman[7], T. Carling[3], P. Björklund[7], R.P. Lifton[1,2,4,8].* 1) Department of Genetics, Yale University School of Medicine, New Haven, CT; 2) Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, CT; 3) Department of Surgery, Yale Endocrine Neoplasia Laboratory and Yale Cancer Center, Yale University School of Medicine, New Haven, CT; 4) Yale Center for Genome Analysis, Yale University, New Haven, CT; 5) Department of Pathology, Yale University School of Medicine, New Haven, CT; 6) Division of Endocrinology and Neag Cancer Center, University of Connecticut Health Center, Farmington, CT; 7) Department of Surgical Sciences, Uppsala University, Uppsala, Sweden; 8) Department of Internal Medicine, Yale University School of Medicine, New Haven, CT.

Insulinomas are benign or malignant pancreatic tumors that inappropriately secrete insulin, producing hypoglycemia. Despite an average of less than 4 protein-altering somatic mutations per tumor, 11 of 33 benign insulinomas harbored the identical somatic mutation in the DNA-binding zinc finger of the transcription factor YY1. ChIP-Seq showed that this mutation, YY1T372R, changes the DNA motif bound by YY1. Gene expression analysis distinguished tumors with and without YY1T372R mutations and identified genes with new YY1 binding sites and significantly altered expression. These included two, ADCY1 (an adenylyl cyclase) and CACNA2D2 (a Ca2+ channel pore-forming subunit), involved in key pathways regulating insulin secretion. Expression of these genes in the rat INS-1 cell line resulted in markedly increased insulin secretion. These findings implicate YY1T372R in insulinoma pathogenesis and intracellular Ca2+ as a common mechanism underlying endocrine neoplasias.

262

Somatic L1 retrotransposition occurs early during colorectal tumorigenesis. *S. Solyom[1], A.D. Ewing[2], N. Baker[2], A. Gacita[2], L.D. Wood[3], S.J. Meltzer[4], B. Vogelstein[5], K.W. Kinzler[5], H.H. Kazazian[1].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Center for Biomolecular Science and Engineering, University of California, Santa Cruz, CA; 3) Gastrointestinal & Liver Pathology, Johns Hopkins Hospital, Baltimore, MD; 4) Johns Hopkins University School of Medicine & Sidney Kimmel Comprehensive Cancer Center, Baltimore, MD; 5) The Ludwig Center and The Howard Hughes Medical Institute at Johns Hopkins Kimmel Cancer Center, Baltimore, MD.

Long INterspersed Element-1 (L1) retrotransposons are autonomous mobile elements that comprise 17% of the human genome. L1s retrotranspose by a 'copy and paste' mechanism via an RNA intermediate. Although insertional mutagenesis by these elements is a known cause of various human Mendelian diseases, their somatic mobilization in the cancer genome has only recently been established as a new mutational phenomenon. Here we investigate the timing of insertional events and the extent of tumor heterogeneity conferred by retrotransposons, as well as their impact on malignancy. We studied DNA from 4 colon cancer patients who had been previously diagnosed with colonic polyps (3 adenomas and one hyperplastic), as well as from 5 additional patients with colorectal dysplasia arising in inflammatory bowel disease (IBD). In contrast to the paired polyp-cancer samples, the IBD cancers were immediately adjacent to, and likely originated from, their matched dysplasias. Two of the 4 patients with both colon adenomas and carcinomas also had metastases. After dissection of abnormal from normal tissue, next-generation L1-targeted resequencing (L1-seq) was carried out on DNA from these tissues. After PCR-validation and Sanger sequencing of putative somatic L1 insertions, we found for the first time that certain pre-cancerous lesions were mutagenized by somatic L1 insertions. We validated 56 somatic insertions in cancerous or pre-cancerous lesions, of which 31 occurred in pre-malignant lesions. We found no verifiable insertions in normal colon. Surprisingly, 2 adenomas contained more than 9 insertions each. Two IBD dysplasias contained insertions as well, and 6 L1 insertions were present not only in these lesions, but also in their paired cancers. Finally, we validated 2 metastasis-specific insertions and 6 insertions found in both primary colon cancers and their matched metastases. Since 6 of 7 insertions in colon cancers were present in their matched metastases, and 6 of 9 insertions in IBD dysplasias were also present in their paired carcinomas, many insertions may be present in every cell of the cancer or the dysplasia. These data suggest that L1s may serve as novel biomarkers of malignant progression. Numerous genes - some with unknown function - were targeted by L1s and may play a role in malignancy. To conclude, somatic retrotransposition is involved early in the pathogenesis of some colon cancers and may provide useful biomarkers of neoplastic progression.

263

Transcriptome sequence analysis of human colorectal cancer samples reveals cancer functional attributes. *H. Ongen[1,2,3], T.F. Orntoft[4], J.B. Bramsen[4], B. Oster[4], L. Romano[1,2,3], A. Planchon[1,2,3], C.L. Andersen[4], E.T. Dermitzakis[1,2,3].* 1) Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 2) SIB, Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) iGE3, Institute of Genetics and Genomics in Geneva, Geneva, Switzerland; 4) Department of Molecular Medicine, Aarhus University Hospital, Aarhus, Denmark.

In developed countries colorectal cancer (CRC) is the second leading cause of cancer death with a million new cases worldwide yearly and a mortality rate of ~50%. Among cancers, the incidence of CRC ranks fourth in men and third in women and will grow with the increase of westernized lifestyles. Here we will report the RNA-seq analysis of matched tumor and normal mucosa from 103 CRC patients and 20 other normal tissues from unrelated healthy donors. The normal tissues are used to assess how CRC differentiates not only from its normal counterpart but also from other normal tissue types. RNA was purified from microdissected samples (tumour percentage ≥ 60%) and sequenced on the Illumina HiSeq (34-80 million reads). Furthermore 90 of the CRC patients' germline genomes are genotyped. We observe 1626 differentially expressed genes (FDR = 5%, fold change ≥ 2) between normal colon and cancer. Tumors, normal colons, and other tissue samples form three distinct clusters based on gene expression, and tumors cluster in between the other two, suggesting that cancer increases the variance of the transcriptome but in a predictable manner. On average there are 688 significant allele-specific expression (ASE) signals (FDR = 1%) per sample. The proportion of heterozygous sites that have an ASE effect is significantly more in tumors. About 34% of the ASE sites are tumor specific and ~10% of shared ASE sites between normal and tumor exhibit reversal of the effect suggestive of loss of heterozygosity in these genes. Unlike gene expression when ASE is considered, genetic effects on gene expression are more similar between tumors and their match controls than between tumors and other tumors. We identify multiple regions on nearly all chromosomes where the correlation of expression for proximal genes is significantly increased in the tumors when compared to normals. There are 1693 and 948 eQTLs (permutation $P < 0.01$) in normal and tumors respectively. Approximately 60% of the eQTLs are shared between normal and tumor, with tumor specific eQTLs numbering over 300. We are investigating whether the tumor specific eQTL genes accumulate somatic mutations making them likely candidates for being driver genes. Preliminary analysis indicates that there are 124 gene fusions in the tumor population studied although none are recurrent. Altogether these results will greatly benefit our understanding of colorectal tumorigenesis.

264

Comparative Whole Genome Sequencing to Identify Candidate Somatic Driver Mutations of Li-Fraumeni Syndrome Sarcomagenesis in Humans and Mice. *J. Wong[1,2], L.C. Strong[1,2], G. Lozano[1,3], L.L. Bachinski[1], R. Krahe[1,2,3].* 1) Genetics, MD Anderson Cancer Center, Houston, TX; 2) Human & Molecular Genetics Program, University of Texas-Houston Graduate School in Biomedical Sciences, Houston, TX; 3) Genes & Development Program, University of Texas-Houston Graduate School in Biomedical Sciences, Houston, TX.

Li-Fraumeni Syndrome (LFS) is a rare, clinically heterogeneous, variably penetrant cancer susceptibility syndrome that has largely been attributed to germline mutations in the tumor suppressor gene *TP53* (*p53*; >70% of all cases). The tumor spectrum for LFS is considerably broad, characterized by a high incidence of sarcomas considered a hallmark of LFS. *p53* LFS carriers develop sarcomas at a wide range of ages and sites, suggesting that while the p53 mutation confers cancer predisposition, additional mutations are necessary for sarcomagenesis. To evaluate potential somatic drivers, we performed whole genome sequencing (WGS) across two sarcomas and matched (PBL) normal tissue from known *p53* LFS carriers *(M133T, R175H)* to pinpoint recurrent acquired changes. However, driver identification is complicated by the acquisition of driver mutations along with numerous passenger events, which is potentially confounded by genome architecture. To distinguish drivers from passengers in human sarcomas, we performed comparative WGS on tumor and normal tissues from a "humanized" LFS mouse model with a germline *p53+/R172H* missense mutation (analogous to the human hot-spot *TP53 R175H* LFS mutation) that predisposes to a similarly broad tumor spectrum, including sarcomas. Two mouse sarcomas were evaluated against a pool of five normal mice for recurrent somatic changes. In several genes with known human orthologs, the mouse sarcomas had somatic variants predicted to have functional consequences and demonstrated roles in human cancers. Integration of human and mouse data revealed the PTEN pathway as a potential key cancer pathway shared between human and mouse sarcomas. Several genes, such as *TRIM16* and *PCDHB8*, showed potential driver mutations across multiple samples. Our overall results indicate the utility of comparative WGS of human cancers and their mouse models for the identification of somatic driver mutations.

265

Comprehensive transcriptome and epigenome sequencing of hypoxic breast cancer reveals non-coding RNAs associated with clinicopathological features. *H. Choudhry[1,2], J. Schodel[3], S. Oikonomopoulos[1], C. Camps[1], P. Ratcliffe[3], A. Harris[4], D. Mole[3], I. Ragoussis[1].* 1) Nuffield Department of Clinical Medicine, Wellcome Trust Centre of Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Biochemistry Department, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia; 3) Henry Wellcome Building for Molecular Physiology, University of Oxford, Oxford, United Kingdom; 4) Department of Oncology, John Radcliffe Hospital, University of Oxford, Oxford, United Kingdom.

Hypoxia is an important physiological stress with a central role in many responses including adaptation to altitude, growth and development, as well as major pathophysiological processes such as ischaemic and cancer. Pan-genomic analyses of these responses have focussed on protein-coding genes. However, it is now recognised that many transcripts have functions that do not include coding for proteins (non-coding RNA). Here, we comprehensively profile this non-coding transcriptional output in hypoxia. We undertook an integrated pan-genomic analysis of normoxic and hypoxic MCF7 breast cancer cells, employing RNA-seq together with ChIP-seq for the major hypoxia-inducible transcription factor HIF and for chromosomal markers of active transcription (RNApol2 and histone H3K4 methylation). We further assessed these responses following knockdown of HIF transcription factors using RNAi. We establish a computational pipeline of strand specific ribo-depleted RNA-seq data to detect regulated non-coding transcripts including piwiRNA, miRNA, tRNA, sn/snoRNA, and lncRNA. Compared to other classes snRNAs and tRNAs are globally downregulated, whilst a significant number of lncRNAs are up-regulated. These up-regulated lncRNAs are associated both with chromosomal markers of transcriptional activation and with HIF binding, indicating direct transcriptional activation of non-coding transcripts by HIF. Dependence of lncRNAs on HIF was further confirmed by HIF RNAi. In addition we describe 105 "novel" previously unannotated transcripts bearing chromosomal marks of bon-fide genes. Four hypoxically induced lncRNAs were then analysed in 148 breast tumours and associated with clinicopathological features.Our findings extend knowledge of the hypoxic transcriptional response into the spectrum of non-coding transcripts. We demonstrate that HIF can transcriptionally activate lncRNAs in addition to coding transcripts and link these to clinicopathological features. These HIF-regulated non-coding transcripts have the potential to act as new biomarkers for breast cancer as well as potentially novel therapeutic targets.

266

Addressing the complexity of cancer: integrative genomic and transcriptomic analysis of 775 human cancer cell line reveals novel drivers and regulatory programs. *A.C. Villani[1], C. Ye[1], A. Regev[1,2,3], N. Hacohen[1,4,5].* 1) Genome Biology & Cell Circuits, Broad Institute of MIT and Harvard, Cambridge, MA; 2) Howard Hughes Medical Institute, Chevy Chase, MD; 3) Biology Dept, Massachusetts Institute of Technology, Cambridge, MA; 4) Harvard Medical School, Boston, MA; 5) Center for Immunology and Inflammatory Disease, Massachusetts General Hospital, Charlestown, MA.

Cancer is a genetically driven disease, and human cancer cell lines carrying such genetic drivers are a great model system for performing mechanistic studies of autonomous tumor-driven pathways in homogeneous controlled experimental settings. Any putative driver commonly altered in several cancer cell lines from different tissues of origin may point towards novel tumorigenesis mechanisms. We postulate that large-scale integrative genomics analysis of cancer cell lines repository could lead to the identification of novel master regulators of autonomous programs as well as key cell lines for follow-up functional studies. We performed an unbiased integrative analysis of chromosomal copy number aberrations (CNAs) and gene expression data (Affymetrix U133+2.0) of 775 human cancer cell lines, derived from 23 different tissues of origin. Using modified versions of GISTIC2.0 and CONEXIC, a Bayesian module network method, we sought to identify novel cancer driver genes that commonly alter transcriptional regulatory programs across multiple cell lines. To account for the inherent heterogeneity in expression data, we corrected for known (histology, sub-histology, ethnicity, gender, center of collection, experimental batches) and unknown confounding variables using Surrogate Variable Analysis (SVA). We identified a total of 60 amplified and 222 deleted somatic regions, of which 52% contained a candidate driver gene whose expression was associated with the expression of a target module. The driver genes consisted of putative novel and known cancer drivers, including deletions in CDKN2A-CDKN2B (q< 4.2E-176), WWOX (q< 5.3E-100), and amplifications in MYC (q< 2.1E-81) and CCND1 (q< 1.4E-44) genes. The target modules were enriched for known cancer dis-regulated processes, including cell cycle (q< 2.1E-91) and DNA damage/repair (q< 9.3E-50) associated programs. Interestingly, we also identified novel immune (q< 9.8E-26) and metabolic (q< 1.5E-17) cancer-autonomous programs common to several tissues of origin. These latter programs may be key to driving polarization of the tumor microenvironment. Several cell lines were nominated for ongoing follow-up functional studies to validate these predictions. Our results provide a powerful framework to identify putative novel drivers, nominate cell lines for follow-up functional studies, and highlight altered pathways common to several cancer models with biological, and possibly therapeutic, importance in cancer.

267

Diagnostic yield of clinical tumor exome sequencing for newly diagnosed pediatric solid tumor patients. *D.W. Parsons[1,2,3,4], A. Roy[1,5], F.A. Monzon[2,4,5], D.H. López-Terrada[1,4,5], M.M. Chintagumpala[1,4], S.L. Berg[1,4], S.G. Hilsenbeck[4,6], T. Wang[4], R.A. Kerstein[1], S. Scollon[1], K. Bergstrom[1], U. Ramamurthy[1,7], D.A. Wheeler[2,3,4], C.M. Eng[2], Y. Yang[2], J.G. Reid[2,3], D.A. Muzny[2,3], R.A. Gibbs[2,3], S.E. Plon[1,2,3,4].* 1) Pediatrics, Baylor College of Medicine, Houston, TX; 2) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 4) Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX; 5) Pathology and Immunology, Baylor College of Medicine, Houston, TX; 6) Medicine, Baylor College of Medicine, Houston, TX; 7) Institute for Clinical & Translational Research, Baylor College of Medicine, Houston, TX.

Although advances in sequencing technologies now allow provision of genome-scale data for cancer patients, experience with systematic application of clinical genomic sequencing remains sparse. The BASIC3 study follows pediatric patients with newly diagnosed solid tumors (targeted enrollment n=280) at Texas Children's Cancer Center for two years after performing CLIA-certified whole exome sequencing (WES) of blood and tumor. Since the study opened in August 2012, 70/82 (85%) eligible families have consented to participate, allowing an early assessment of the feasibility and utility of performing clinical tumor WES for these patients. Snap-frozen tumor and matched normal blood samples are submitted to the Baylor College of Medicine (BCM) Whole Genome and Cancer Genetics Laboratories for WES using a clinically-validated pipeline, with the resulting tumor and germline WES reports submitted into the electronic medical record. The first 70 patients comprise a diverse representation of diagnoses, including 23 (33%) with CNS tumors and 47 (67%) with non-CNS solid tumors. Despite limited diagnostic biopsies in many patients, snap-frozen tumor samples adequate for WES were available from 61/70 (87%). Tumor WES results have been reported for the first 32 patients, revealing a median of 9 (range 1-35) protein-altering mutations per tumor and alterations of known cancer genes such as ALK, KRAS, NRAS, MET, JAK2, FGFR3, ARID1A, CTNNB1, and TP53. Nine of 32 tumors (28%) contained mutations classified as having proven or potential clinical utility, including 9/24 (38%) non-CNS tumors and 0/8 CNS tumors (Fisher's exact test, p=0.07). Notably, only four of these mutations would have been detected on the clinically-available BCM Cancer Gene Mutation Panel (v.2). These results demonstrate the feasibility of routine tumor WES in the pediatric oncology clinic, indicate that potentially clinically-relevant mutations can be identified in a substantial minority of pediatric solid tumor patients, and suggest that the diagnostic yield of WES may be greater for non-CNS tumors as compared to CNS tumors. The diversity of mutations identified suggests that exome or genome-scale diagnostic approaches for pediatric solid tumors may be favored over targeted mutation panels which are often optimized for adult malignancies. Clinical utility of WES data as assessed by the pediatric oncologists will be measured longitudinally for each study patient. Supported by NHGRI/NCI 1U01HG006485.

268

Simultaneous estimation of population size changes and splits times from population level resequencing studies. *M. Forest, J. Marchini, S. Myers.* Department of Statistics, University of Oxford, Oxford, United Kingdom.

In the quest to understand human evolution, key questions might refer to the time of divergence of different human populations, and to the variation of the population sizes through time. In the recent past, genetic data have proven to offer a complementary insight to archeological discoveries in regard to population histories. Large genomic projects are now offering access to high quality data from a vast number of populations (ex. the 1,000 Genomes Project (1KGP)). Our aim is to study such population structure using large samples of genomic sequencing data from different populations. We have developed an approach that builds trees at thousands of loci, and uses these to infer demographic history allowing for arbitrary population splits, and size changes, over a series of epochs in the past, a previously unsolved problem. By jointly analysing hundreds of individuals, we show by simulation and real world application that we can accurately estimate population separation times and sizes from only a few thousand, to hundreds of thousands of years in the past. Our approach extends the Stephens and Donnelly importance sampler, to allow estimation of the divergence time and population sizes. The method is able to jointly utilise data from multiple regions that show very low levels of recombination (cold spots) in total covering hundreds of megabases. An iterative scheme allow us to: (1) obtain point estimates of the likelihood function using the coalescent process and the built genealogies, and (2) obtain maximum likelihood estimates of the effective population sizes using the previously built genealogies. The population sizes are modelled as piecewise constant and are allowed to vary freely in between different epochs. An optimization algorithm allows us to quickly find the maximum of the estimated likelihood function. We have applied the method to different populations from the 1KGP using more than 2,000 cold regions of the genome (average length of 30Kb). By analysing many pairs of populations using 1KGP sequencing data, we elucidate details of the relationships among multiple human groups, and changes in their effective population sizes, from a few thousand years ago. Our results unify and extend previous results on the split times between European groups, and among Europe, Africa and Asia, shared and non-shared bottlenecks in out-of-Africa groups and expansions following population separations, and the sizes of ancestral populations further back in time.

269

Inferring complex demographies from PSMC coalescent rate estimates: African substructure and the Out-of-Africa event. *S. Gopalakrishnan[1], P. Grabowski[2], M.C. Turchin[1], B. Henn[7], J. Kidd[5], G. Perry[3], A. Gebremedhin[6], C. Beall[4], C.D. Bustamante[7], A. Di Rienzo[1], Y. Gilad[1], A.A. Palmer[1], J.K. Pritchard[1].* 1) Human Genetics, University of Chicago, Chicago, IL; 2) Ecology & Evolution, University of Chicago, Chicago, IL; 3) Anthropology, Penn State University, University Park, PA; 4) Anthropology, Case Western Reserve University, Cleveland, OH; 5) Human Genetics, University of Michigan, Ann Arbor, MI; 6) Medicine, Addis Ababa University, Addis Ababa, Ethiopia; 7) Genetics, Stanford University, Stanford, CA.

Human population history is an intriguing and complex story with many events like population growth, bottlenecks, time-dependent/non-homogeneous migration, population splits and mixtures. Estimating complete demographies with population sizes, rates of gene flow and population split times has proven to be a challenging endeavor. We propose a framework for jointly estimating the demography parameters, especially gene-flow rates and split times, for a large number of populations. We use coalescent rate estimates obtained from Pairwise Sequentially Markovian Coalescent (PSMC) as the starting point for our analysis. Since PSMC works on only two chromosomes at a time, we apply PSMC to all pairs of individuals to obtain the pairwise coalescent rates for lineages from every pair of sampled populations. Using a mathematical model for calculating coalescent probabilites given population parameters, we estimate demography using the parameters that best fit the observed coalesecent rates.

In this study, we focus on two aspects of African population genetics, 1. the nature of population structure in Africa going back in time and 2. the timing of the Out-of-Africa event. To address these questions, we assembled a dataset with whole genome sequences from 162 individuals using both in-house sequencing and publicly available sources. These samples span 22 populations worldwide. These include eleven African populations which we use to dissect the population substructure in Africa. In addition, we also have 2 Middle Eastern, 5 European and 4 East/Central Asian populations which inform the population split time estimates for the Out-of-Africa event and the European-Asian split.

We find extensive population structure in Africa extending back to before the Out-of-Africa event. The Ethiopian populations, Amhara and Oromo, show evidence of mixing beyond 15 kya. The Maasai and Luhye merge with the Ethiopian populations to form a panmictic East African population ~40kya. We find evidence for extensive mixing between east and west African populations before 50kya. Among the pygmy populations, we see recent gene flow between the Batwa and Mbuti. All African populations except the San merge into a single population around 110 kya. The San exchange migrants with the other African populations beginning ~120 kya. We estimate the Out-of-Africa event to have occurred ~75kya and the European-Asian split to ~25kya.

270

Out of Africa, which way? *L. Pagani[1,2,3], T. Kivisild[2], S. Shiffels[1], A. Scally[4], Y. Chen[1], Y. Xue[1], P. Danecek[1], J. Maslen[1], M. Haber[5], R. Ekong[6], T. Oljira[7], E. Mekonnen[7], D. Luiselli[3], E. Bekele[7], P. Zalloua[5], C. Tyler-Smith[1].* 1) The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom; 2) Division of Biological Anthropology, University of Cambridge, CB2 1QH, UK; 3) Department of Biological, Geological and Environmental Sciences of the University of Bologna; Italy; 4) Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK; 5) The Lebanese American University, Chouran, Beirut, Lebanon; 6) Department of Genetics, Evolution & Environment, University College London, WC1E 6BT, UK; 7) University of Addis Ababa, and Center of Human Genetic Diversity, P.O. Box 1176, Ethiopia.

While the African origin of all modern human populations is well-established, the dynamics of the diaspora that led anatomically modern humans to colonize the lands outside Africa are still under debate. Understanding the demographic parameters as well as the route (or routes) followed by the ancestors of all non-Africans could help to refine our understanding of the selection processes that occurred subsequently, as well as shedding light on a landmark process in our evolutionary history. Of the three possible gateways out of Africa (via Morocco across the Gibraltar strait, via Egypt through the Suez isthmus or via the Horn of Africa across Bab el Mandeb strait) only the latter two are supported by paleoclimatic and archaeological evidence. Furthermore, recent studies (Pagani et al. 2012) showed that, although the modern Ethiopian populations might be good candidates for the descendants of the source population of such a migration, modern Egyptians could be an even better candidate. Unfortunately, however, only a few Egyptian samples have been genotyped and, as yet, none have been fully sequenced. Here, we have generated 125 Ethiopian and 100 Egyptian whole genome sequences (Illumina HiSeq, 8x average depth). The genomes were partitioned using PCAdmix (Brisbin et al. 2012) to account for the confounding effects of recent introgression from neighboring non-African populations. To explore the genetic legacy of migration routes out of Africa, and in particular to test whether the observed genetic data support one route over another, the African components of Egyptians and Ethiopians were then compared to a panel of available non-African populations from the 1000 Genomes Project (1000 Genomes Project Consortium, 2012). The high resolution provided by whole genome sequencing allows us to shed new light on the paths followed by our ancestors as they left Africa, as well as refining the current knowledge of the demographic history of the populations analyzed.

271

Insights into the genetic architecture of African genomes: the African Genome Variation Project. *I. Tachmazidou for the AGVP investigators.* The Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

Genome-wide association studies in populations from sub-Saharan Africa are eagerly anticipated, but there is a paucity of genetic data to inform powerful study design. Pronounced genetic diversity across ethnic groups within SSA, in conjunction with low levels of LD and differences in haplotype structure, give rise to statistical genetics challenges when designing and conducting genomic epidemiology studies. The African Genome Variation Project is a collaboration across the African Partnership for Chronic Disease Research, the Centre for Research on Genomics and Global Health and the Malaria Genomic Epidemiology Network. Our aim is to facilitate genome-wide association studies in diseases of relevance to African populations by providing first insights into the genetic variation landscape of different ethnic groups. To achieve this, we have genotyped 100 unrelated individuals from each of 18 ethnolinguistic groups from 7 SSA countries (Kenya, Nigeria, Uganda, Ethiopia, Ghana, the Gambia, South Africa) on the 2.5 million SNP Illumina platform. We are examining: 1) the allele frequency spectrum of variants on the chip; 2) patterns of LD; 3) the proportion of common variation captured by the array; 4) imputation-based approaches aiming to increase genetic association study power; and 5) analytical challenges and the need for new statistical genetics methods to address them. We find that between 1.10 and 1.36 million SNPs have MAF>5% and that between 240 and 490 thousand SNPs are monomorphic depending on the population examined. We also find that there are high levels of redundancy on the chip, as calculated based on pairwise correlation between variants in each ethnolinguistic group; for example, for 40-57% of common variants there is at least one more variant with r^2 over 0.8 on the chip, whereas 16-35% of common variation has a perfect proxy on the chip. Based on whole genome sequence data, we find an upper threshold of 70% of MAF>5% variants captured by the array (50% for MAF>1%) at an r^2 of 0.8. To explore the utility of SSA groups to serve as imputation reference panels for other SSA populations, we imputed Baganda, Ethiopia and Zulu samples on the 1000 Genomes low coverage sequence data. Correlation between the input genotypes and the expected genotypes varies between 60-75% for MAF<5% and between 70-88% for MAF>5% depending on the population examined.

272

Genetic evidence for multiple episodes of population mixture in southern and eastern African history. *J. Pickrell[1], N. Patterson[2], P. Loh[3], M. Lipson[3], B. Berger[3], M. Stoneking[4], B. Pakendorf[5], D. Reich[1,2].* 1) Department of Genetics, Harvard Medical School, Boston, MA; 2) Broad Institute, Cambridge, MA; 3) Department of Mathematics and Computer Science, MIT, Cambridge, MA; 4) Department of Evolutionary Genetics, Max Plank Institute, Leipzig, Germany; 5) Laboratoire Dynamique du Langage, CNRS and Université Lyon Lumière.

The history of southern Africa involved interactions between indigenous hunter-gatherers and a range of populations that moved temporarily or permanently into the region. The influence of these interactions on the genetic structure of current populations remains unclear. Here, using patterns of linkage disequilibrium inferred from genome-wide genetic data, we show that there are at least two admixture events in the history of Khoisan populations (southern African hunter-gatherers and pastoralists who speak non-Bantu languages with click consonants): one involving populations related to Niger-Congo-speaking African populations, and one that introduced ancestry most closely related to west Eurasian (European or Middle Eastern) populations. We date this latter admixture event to approximately 900-1,800 years ago, and show that it had the largest demographic impact in the subset of Khoisan populations who speak Khoe-Kwadi languages. A similar signal of west Eurasian ancestry is present throughout eastern Africa; in particular, we also find evidence for two admixture events in the history of several Kenyan, Tanzanian, and Ethiopian populations, the earlier of which involved populations related to west Eurasians and which we date to approximately 2,700 - 3,300 years ago. We thus suggest that west Eurasian ancestry entered southern Africa indirectly through eastern Africa. These results demonstrate how large-scale genomic datasets can inform complex models of population movements, and highlight the genomic impact of largely uncharacterized back-to-Africa migrations in human history.

273

Inferring the evolutionary history and the genetic basis of small stature in African Pygmies from whole-genome sequencing data. *M. Sikora[1], E. Patin[2], H. Costa[1], K. Siddle[2], B.M. Henn[1,6], J.M. Kidd[1,7], R. Kita[1], M.C. Yee[1], P. Verdu[3], L. Barreiro[4], J.M. Hombert[5], E. Heyer[3], C.D. Bustamante[1], L. Quintana-Murci[2].* 1) Department of Genetics, Stanford University, Stanford, CA; 2) Institut Pasteur, Unit of Human Evolutionary Genetics, CNRS URA3012, Paris, France; 3) CNRS-MNHN-Universite Paris 7, UMR 7206 Eco-anthropologie et Ethnobiologie, Paris, France; 4) CHU Sainte-Justine, Universite de Montreal, Montreal, Canada; 5) Laboratoire Dynamique Du Langage, CNRS UMR 5596 Universite Lumière Lyon 2, Lyon, France; 6) Present address: Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY, USA; 7) Present address: Departments of Human Genetics and Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, USA.

The Pygmy populations of the Central African rainforests are some of the last remaining hunter-gatherers among present-day human populations. They are characterized by distinct cultural and physical characteristics such as short stature, which are thought to be a consequence of the unique environment they inhabit. Here we present the results of a population genomic study investiagting the demographic and adaptive history of these populations using whole-genome sequencing data. We sequenced 47 individuals from two Pygmy populations (Western Baka, Eastern Mbuti), as well as one non-Pygmy farmer population (Nzebi) to a median sequencing depth of 5.5x. Multi-sample SNP calling followed by stringent QC filters resulted in a final dataset of >17 million SNPs. Analyses of population structure and pairwise identity-by-descent (IBD) sharing patterns revealed varying levels of farmer-related admixture in the Baka. Both Pygmy populations showed increased within-population IBD sharing and relatedness, consistent with recent genetic isolation. Demographic inference using the pairwise-sequential Markov Coalescent (psmc) showed that Pygmy populations diverged from the Nzebi around 50,000 years ago, with a subsequent split between the Mbuti and Baka starting 30,000 years ago. We also found evidence for a population expansion unique to the Baka after their divergence from the Mbuti. In order to investigate the genetic basis of small stature in the Pygmies, we scanned the genome for highly differentiated regions using the population-specific branch length method (PBS). Highly differentiated SNPs were annotated with a rich set of both coding and non-coding annotations from a variety of sources in order to identify variants of interest. We found that while GWAS SNPs previously associated with height do not show higher frequencies for the “small” allele in Pygmies, the chromosomal regions around those SNPs show significantly increased differentiation in Pygmies compared to random GWAS regions. Furthermore, we identified SNPs in genes of the growth hormone / IGF1 axis that are rare worldwide but common in Pygmies, and associated with regulatory features such as DNase hypersensitivity sites. These candidate SNPs were followed up in a larger replication panel of around 300 individuals from diverse ethnic groups, including phenotypic data on height. Our study provides an unprecedented view of the evolutionary history and the genetic basis of small stature in African Pygmies.

274

Reconstructing the Population Genetic History of the Caribbean. *A. Moreno Estrada[1], S. Gravel[1], F. Zakharia[1], J.L. McCauley[2], J.K. Byrnes[1], C.R. Gignoux[3], P. Ortiz Tello[1], K. Sandoval[1], P.J. Norman[4], P. Parham[4], J.C. Martinez Cruzado[5], E. Gonzalez Burchard[3], M.L. Cuccaro[2], E.R. Martin[2], C.D. Bustamante[1].* 1) Genetics, Stanford University, Stanford, CA; 2) Genetics, University of Miami, Miami, FL; 3) Genetics, University of California at San Francisco, San Francisco, CA; 4) Structural Biology, Stanford University, Stanford, CA; 5) Biology, University of Puerto Rico, Mayaguez, PR.

The Caribbean basin is home to some of the most complex interactions in recent history among previously diverged human populations. Here, by generating genome-wide SNP array data from 330 individuals, we characterize ancestral components of Caribbean populations on a sub-continental level and unveil fine-scale patterns of population structure distinguishing insular from mainland Caribbean populations as well as from other Hispanic/Latino groups. We combined these data with our unique database on genomic variation in over 3,000 individuals from diverse European, African, and Native American populations. We use local ancestry inference and a novel extended space Markov model of ancestry tract to test different demographic scenarios for the pre- and post-colonial history of the region. We provide genetic evidence for an inland South American origin of the Native American component in island populations and for extensive pre-Columbian gene flow across the Caribbean basin. The Caribbean-derived European component shows significant differentiation from parental Iberian populations, presumably as a result of founder effects during the colonization of the New World. Based on demographic models, we reconstruct the complex population history of the Caribbean since the onset of continental admixture. We find that insular populations are best modeled as mixtures absorbing two pulses of African migrants, coinciding with early and maximum activity stages of the transatlantic slave trade. These two pulses appear to have originated in different regions within West Africa, imprinting two distinguishable signatures in present day Afro-Caribbean genomes and shedding light on the genetic impact of the dynamics occurring during the slave trade in the Caribbean. These results demonstrate that dense population genomic data coupled with novel methods of analysis afford the possibility of recapitulating human population history from admixed genomes with far greater resolution than previously thought.

275

Selective interference driven by variable recombination impacts mutation load in humans. *J. Hussin[1,2,3], Y. Idaghdour[1,2,3], A. Hodgkinson[1,2], J-C. Grenier[2], J-P. Goulet[3], E. Gbeha[2], E. Hip-Ki[2], Y. Payette[3], C. Boileau[3], P. Awadalla[1,2,3].* 1) Faculty of Medecine, University of Montreal, Quebec, Canada; 2) CHU Ste-Justine Research Centre, Montreal, Quebec, Canada; 3) Cartagene project, CHU Ste-Justine, Montreal, Quebec, Canada.

A major prediction in population genetics is that linked functional mutations interfere with each other's rate of loss or accumulation, thus reducing the overall efficacy of natural selection in non-recombining systems. To date, it remains however unclear whether variation in crossover rate across recombining chromosomes translates in variation in the rate of adaptation along the human genome. To investigate the efficacy of selection across different recombination environments, we report genomic analyses of mutational load in 521 French-Canadian transcriptomes from the CARTaGENE project and 911 high-coverage exomes from the 1000 Genomes Project from European, Asian and African populations. In each population, we estimated the differential mutational load between high and low recombination regions for variants with different impact on fitness. Both at the population and individual levels, variants in low recombining regions are significantly enriched in highly constrained, low frequency missense mutations relative to variants in regions of high recombination. Using paired-end sequencing reads to unambiguously determine phased haplotypes, we observed that rare and weakly deleterious variants are preferentially linked with each other in low recombining regions, the signature of a Muller's ratchet process. We further observed that the mutational burden in regions of low recombination varies among human populations, with recently founded populations showing a larger differential mutational load at the individual level. Along with an excess of pseudogenes in low recombining regions, these results together indicate that weakly deleterious variants, accumulating on degenerating haplotypes, are less efficiently removed by natural selection in regions of low recombination rate. As low-recombining regions are enriched for genes with essential cellular function, this phenomenon likely impacts disease susceptibility at the individual level.

276

Human population assembly and error-correction of sequence reads. *Z. Iqbal[1], S. McCarthy[2], H. Zheng Bradley[3], C. Xiao[4], A. Marcketta[5], G. McVean[1,6].* 1) Wellcome Trust Centre for Huma, University of Oxford, Oxford, United Kingdom; 2) Wellcome Trust Sanger Institute, Hinxton, UK; 3) European Bioinformatics Institute, Hinxton, UK; 4) National Centre for Biotechnology Information, NIH, Bethesda, USA; 5) Albert Einstein School of Medicine; 6) Department of Statistics, University of Oxford.

As sequencing technologies improve and read-lengths increase, a major challenge is going to be the error-correction of reads, especially with low coverage. Current methods hinge either on the use of coverage as a proxy for truth, or a reference genome. For those interested in SNP and indel analysis, it is important not to throw out true polymorphisms in the process of removing errors. What we would like to be able to do is make use of prior knowledge, and have more confidence in reads that matches known sequence. A recent development in genome analysis, introduced in [1,2], has been the idea of using de novo assembly not just to study a single individual, but to learn about an entire species. This allows an unbiased access to all sequence - for example [1] found gene sequence that was highly differentiated between Europe, Africa and Asia, but which was missing from the reference genome.

Using the Cortex assembler, we have built assembly graphs of 1092 humans from 14 populations from Phase 1 of the 1000 Genomes Project. We show how the 1092-sample graph can be used as a repository of known sequence, allowing single-pass quality-aware error correction of reads, improving both power and concordance with genotype arrays. We demonstrate, both on a single high-coverage sample and on a cohort of low-coverage samples from a population absent from the graph. Although the majority of Illumina reads require zero or one base to be corrected, a non-negligible number have more than 10 bases corrected, including correction of Ns. These reads, which previously had no BLAST hits, now BLAST confidently to human sequence.

The method extends transparently, so it is possible to use trusted graphs of dbSNP, the 1000 Genomes SNP calls, or any assemblies. We discuss the value of using prior information in this manner. These approaches will be of great value in an era of low coverage long-reads.

[1] De novo assembly and genotyping of variants using colored de Bruijn graphs. Z Iqbal, M Caccamo, I Turner, P Flicek, G McVean, Nature Genetics (2012) [2] High-throughput microbial population genomics using the Cortex variation assembler. Z Iqbal, I Turner, G McVean, Bioinformatics (2012).

277

Fine mapping of the MHC in >60,000 samples by the International IBD Genetics Consortium: identification of multiple predisposing and protective variants that are mostly distinct between Crohn's disease and ulcerative colitis. *P. Goyette[1,2] on behalf of the International inflammatory bowel disease genetics consortium (IIBDGC).* 1) Research Center, Montreal Heart Institute, Montreal, Quebec, Canada; 2) Department of Medecine, Université de Montréal, Montreal, Quebec, Canada.

The major histocompatibility complex (MHC) is a well-established risk locus for both Crohn's disease (CD) and ulcerative colitis (UC). Previous studies of the MHC region have been hampered by insufficient density of variants tested in order to capture the extensive genetic variation of the region, in particular within the highly polymorphic classical HLA loci, as well as insufficient power to fine-map signals in this region with high linkage disequilibrium. In order to meet these challenges the IIBDGC has performed a large-scale high density genotyping project using the Immunochip custom genotyping array in over 67,000 samples. Following stringent QC, 7041 SNPs in the extended MHC (Chr6:25Mb-34Mb) were tested for association, using a logistic regression model controlling for ancestry, in 18,446 CD and 14,366 UC cases, as well as 33,933 healthy controls. In addition, alleles at the classical human leukocyte antigen (HLA) loci, and amino acids variation relevant to their antigen binding properties, were imputed from the SNP data. The primary association analyses in CD and UC confirm association to variants rs9264942 (1.16, P=1.52×10-25) and rs6927022 (1.56, P=2.7×10-169) for CD and UC, respectively, as previously reported by a recent IIBDGC study (Jostins et al., Nature, 2012). We used forward conditional logistic regression analysis to identify independent genome-wide significant associations. In CD we identified up to 10 independent signals, including 3 in the class II region and 4 near class I genes, mostly centered near HLA-B. In UC we identified up to 12 independent signals, including 6 in the class II region and 3 near classical class I genes. Interestingly, only three of the independent signals are shared between CD and UC, including a non-synonymous coding risk variant in CFB and a combination of rare protective alleles at HLA-B and HLA-C (HLA-B*5201 and HLA-C*1202). Both diseases show associations at HLA-A, but to SNPs tagging different HLA-A alleles; with CD associated to HLA-A*0301 while UC is associated the HLA-A*0201. While CD and UC share the majority of known risk alleles outside the MHC (110 of 163 known IBD risk loci), they only share a limited number of the independent alleles identified within the MHC. The current study demonstrates a strong but not exclusive role for classical class I and II loci in CD and UC, with only few MHC signals localizing outside the classical HLA loci.

278

Mapping the shared and distinct HLA alleles for seropositive and seronegative rheumatoid arthritis. *B. Han[1,2,3], S. Eyre[4,5], D. Diogo[1,3,6], J. Bowes[4,5], Y. Okada[1,3,6], L. Padyukov[7], R. Plenge[1,3,6], L. Klareskog[7], J. Worthington[4,5], P.K. Gregersen[8], P.I.W. de Bakker[1,3,9,10], S. Raychaudhuri[1,2,3,5,6].* 1) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 2) Partners HealthCare Center for Personalized Genetic Medicine, Boston, MA, USA; 3) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 4) Arthritis Research UK Epidemiology Unit, Musculoskeletal Research Group, University of Manchester, Manchester Academic Health Sciences Centre, UK; 5) NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester NHS Foundation Trust, Manchester Academic Health Sciences Centre, UK; 6) Division of Rheumatology, Immunology, and Allergy and Division of Genetics, Brigham and Women's, Hospital, Harvard Medical School, Boston, Massachusetts, 02115, USA; 7) Rheumatology Unit, Department of Medicine, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm, Sweden; 8) The Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, New York, USA; 9) Department of Epidemiology, University Medical Center Utrecht, Utrecht, The Netherlands; 10) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands.

Motivation: Investigators have long speculated that the two subtypes of rheumatoid arthritis (RA), anti-citrullinated protein autoantibody positive (ACPA+) and negative (ACPA-), have distinct underlying genetic factors. The MHC region is the strongest genetic risk factor to ACPA+ RA, but plays a much more modest role in ACPA- RA. To understand the similarities and differences between these two disease subtypes, we fine-mapped and compared MHC associations. **Results:** Using densely genotyped SNP data consisting of 7,222 ACPA+ RA cases, 3,339 ACPA- RA cases, and 15,870 controls from six different cohorts (**Eyre et al., Nat Gen, 2012**), we imputed and tested HLA alleles in the two RA subtypes separately. We mapped associations to ACPA+ RA using forward search conditional analysis and confirmed previously published associations at amino acid sites at positions 13 ($\mathbf{P < 10^{-705}}$), 71, and 74 in *HLA-DRB1*, position 9 in *HLA-B*, and position 9 in *HLA-DPB1*. In addition, we identified a novel association at position 77 in *HLA-A* ($\boldsymbol{P}$**=1.7×10^{-8}**) located in the peptide binding groove implicating antigen presentation as the major mechanism by which MHC variation confers risk. Then in parallel we mapped associations to ACPA- RA. We recognized that ACPA- RA associations to the MHC might be confounded due to the inclusion of misclassified samples that are actually ACPA+ RA (false negative testing) or ankylosing spondylitis. We developed a novel statistical approach that estimates the proportion of misclassified samples and regresses out their effects. Using this approach we observed that each cohort consistently contained 3-9% of cases that likely had ankylosing spondylitis, and a variable number of cases that likely had ACPA+ RA. Controlling for misclassification effects, we identified the amino acid residues at position 13 in *HLA-DRB1* as strongly associated with risk (**Omnibus test $\boldsymbol{P}$= 1.2×10^{-16}**). Serine conferred the highest risk (**OR=1.28**, $\boldsymbol{P}$**=5.7×10^{-13}**); in stark contrast serine conferred protection to ACPA+ disease (**OR=0.4**). We also observed a shared association to the presence of an aspartate in position 9 in *HLA-B* (**$\boldsymbol{P}$=1.3×10^{-15}**, **OR=1.38**) with a more modest effect size than for ACPA+ disease (**OR=2.1**). **Conclusions:** Our analysis is the first to define specific amino acid sites for ACPA- RA, and demonstrates a distinct genetic basis for ACPA+ and ACPA- RA in the MHC region. Our analysis also underscores the importance of phenotypic classification for accurate fine-mapping.

279

Common genetic variants of autoimmunity confer susceptibility to candididemia. *V. Magadi Gopalaiah[1], S. Cheng[2], M.D. Johnson[3,4], S.S. Smeekens[2], L.A.B. Joosten[2], J.R. Perfect[3,4], B. Kullberg[2], C. Wijmenga[1], M.G. Netea[2].* 1) Genetics, University Medical Center Groningen, Groningen, The Netherlands; 2) Department of Internal Medicine and Nijmegen Institute for infection, Inflammation, and Immunity, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands; 3) Duke University Medical Center, Durham, North Carolina, USA; 4) Department of Clinical Research, Campbell University School of Pharmacy, Buies Creek North Carolina USA.

Candidemia is the fourth most common cause of sepsis, with a high mortality rate of up to 40% despite the availability of potent antifungal drugs. Identification of host genetic factors that confer susceptibility to candidemia may aid in designing adjunctive preventive and therapeutic strategies. Ten monogenetic diseases are associated with increased susceptibility to chronic mucocutaneous candidiasis. Intriguingly, single nucleotide polymorphisms (SNPs) in eight out of the ten of these monogenetic disease genes are also associated with susceptibility to different immune-mediated diseases. Thus, we hypothesized that genes involved in the immune system could be strong susceptibility candidates to candidemia. To test our hypothesis, we analyzed around 200,000 SNPs in 186 loci that are associated with 12 different immune-mediated diseases using Immunochip SNP array in the largest candidemia cohort to-date of 316 cases, and in 11,995 controls. The significant associations were replicated in two independent candidemia cohorts. We found SNPs in the CD58 ($P = 1.97 \times 10^{-11}$; OR = 4.68), LCE4A-C1orf68 ($P = 1.98 \times 10^{-10}$; OR = 4.25) and TAGAP ($P = 1.84 \times 10^{-8}$; OR = 2.96) regions to be significantly associated with candidemia. The cumulative effect of these three SNPs indicated that individuals with 2 or more risk alleles have 17.6-fold higher risk of candidemia compared with those with one or no risk allele, implying that these SNPs could be of clinical use for risk prediction. All three SNPs are located within non-coding region and thus may be regulatory in function. Transcriptomics and pathway analysis on macrophages in which CD58 was blocked by siRNA suggested defects in phagocytosis and IL-6 pathway. Blocking CD58 by siRNA induced the incapacity of human macrophages to inhibit yeast-fungal transformation and fungal growth. The physical co-localization of CD58 and Candida during the phagocytosis suggested that CD58 might play a role in recognition and phagocytosis of Candida in macrophages. In contrast, TAGAP was involved in Candida-induced cytokine production. The role of TAGAP was tested in an in vivo disseminated candidiasis model. The fungal loads in liver and kidney were significantly increased in TAGAP -/- mice, demonstrating that TAGAP is important for host defense against Candida infection. In summary, our study not only identified novel genetic risk factors to candidemia but also validated these pathways to be important for antifungal host defense.

280

PXK and Lupus: Defining novel immunobiology for an SLE risk gene. *S.E. Vaughn[1,2,3], I.T.W. Harley[3], C. Foley[4], L.C. Kottyan[1,2], K.M. Kaufman[1,2,5], J.B. Harley[1,2,3,5], SLEGEN.* 1) Immunobiology Graduate Program, Cincinnati Children's Hospital and Medical Center, Cincinnati, OH; 2) Center for Autoimmune Genomics and Etiology, Division of Rheumatology, Cincinnati Children's Hospital and Medical Center, Cincinnati, OH; 3) Medical Scientist Training Program, University of Cincinnati, Cincinnati, OH; 4) Spellman University, Atlanta, GA; 5) US Department of Veterans Affairs Medical Center, Cincinnati, OH.

Systemic Lupus Erythematous (SLE) is a systemic autoimmune disease with a strong genetic component. Over 50 risk genes have been associated with SLE, many with no immediate biological connection to disease. We previously identified one such gene, PXK, as being a candidate gene associated with SLE in women of European descent. These findings have since been replicated. PXK has additionally been identified as a risk gene for RA as well, suggesting that PXK may have a broad role in the pathobiology of autoimmune disease. In this work we undertake the fine mapping of the PXK genetic locus in an effort to refine the association signal. We identify one independent effect in the region occurring strictly in individuals of European ancestry. In an attempt to capture more of the potential variation in the region, we performed imputation, which confirmed our findings of a single associated region found only in Europeans. In tandem with refinement of the genetic signal, we also attempt to identify the SLE relevant biological import of PXK by examining the role it plays in B cells. Public datasets reveal that PXK is highly expressed in the blood, specifically in B cell subsets. PXK has been shown to participate in receptor internalization, and we find using ImageStream technology that PXK colocalizes with the B cell receptor (BCR) upon BCR internalization. Finally, we show that the BCR is more rapidly internalized in cells carrying the PXK SLE-associated allele. These results suggest that PXK may play an important role in the regulation of BCR signaling and B cell differentiation and survival. As B cell regulation is crucial to SLE pathogenesis, understanding the specific changes induced by SLE-associated variants in PXK will provide important insight into SLE pathogenesis.

281

Allelic heterogeneity of and interactions between polymorphic RET enhancers affecting Hirschsprung disease risk. *S. Chatterjee, A. Kapoor, A. Chakravarti.* Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.

Hirschsprung disease (HSCR, congenital aganglionosis) is the most common form of functional intestinal obstruction in neonates (1 in 5,000 live births). We, and others, have identified >15 genes with rare, deleterious mutations in HSCR, the most common being those at the receptor tyrosine kinase gene RET. We have previously identified a polymorphism (rs2435357 (C/T); risk allele T) within a intron 1 RET enhancer with marked effect on HSCR risk and associated with the more common male, short segment, isolated probands with HSCR. By genome-wide association studies, we have now identified a second non-coding polymorphism (P=2.7E-16; rs2506030 (A/G); risk allele G) located ~125 kb upstream of the RET transcription start site. This second variant also disrupts an enhancer since: (a) a 1kb human DNA fragment containing either the G or A allele shows significant loss-of-expression from the risk allele in Neuro2A cells; (b) transient transgenic zebrafish enhancer assays show that the 1 kb element can drive expression in vivo, overlapping ret, in the migratory neural crest cells at 48hpf and dorsal root ganglion at 3dpf in the developing embryo; (c) deletion of a 10nt fragment containing rs2506030 leads to loss of activity in in vitro and in vivo enhancer assays. Furthermore, enhancer assays demonstrate synergistic effects in Neuro2A cells when the 1 kb element contains both rs2435357 and rs2506030. For establishing genetic-phenotypic interactions, we studied the joint genotypes of 337 HSCR patients and 379 controls of European ancestry (1000 Genomes). We find the expected enrichment of both risk alleles in cases, as expected (rs2435357 T: 57%/27% in cases/controls, P=5.46E-31; rs2506030 G: 56%/40% in cases/controls, P=1.24E-9) but a very wide variation in risk by genotype (odds ratio: 0.31 - 9.24, P<0.0015). Moreover, the haplotype distribution of both risk alleles shows greater association in cases (r^2=0.25, P=9.8E-18) than in controls (r^2=0.10, P=9.2E-8), a difference that is statistically significant (P=0.0023) and signifies interaction of their effects. Analysis of the joint effects at both loci suggests that the primary RET susceptibility to HSCR is 'recessive' and from the enhancer at rs2435357 with the new enhancer at rs2506030 modulating that effect. In summary, this is one of the first descriptions of the molecular basis and genetic effects of allelic heterogeneity and interactions of enhancer polymorphisms in human disease.

282

Identification of Human Craniofacial, Thyroid and Heart Enhancers at the FOXE1 Locus. *A.C. Lidral[1], S.A. Bullard[2], R.A. Cornell[3], G. Bonde[3], A. Visel[5], L.M. Moreno[1], J. Machida[4], B. Amendt[3], M.L. Marazita[6].* 1) Dept Orthodontics, Univ Iowa, Iowa City, IA; 2) Department of Endocrinology, University of Iowa, Iowa City, USA; 3) Department of Anatomy, University of Iowa, Iowa City, USA; 4) Department of Oral and Maxillofacial Surgery, Toyota Memorial Hospital, Toyota City, Aichi, Japan; 5) Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. Department of Energy Joint Genome Institute, Walnut Creek, CA, USA; 6) Department of Oral Biology, School of Dental Medicine; Department of Human Genetics, Graduate School of Public Health; Department of Psychiatry, School of Medicine - University of Pittsburgh, Pittsburgh, PA, USA.

Mutations in FOXE1 cause the congenital Bamforth-Lazarus syndrome comprised of agenesis or dysgenesis of the thyroid, cleft palate and hair anomalies. Three common diseases, isolated cleft lip and cleft palate, hypothyroidism and thyroid cancer have all been mapped to the FOXE1 locus at 9q22.33. However, very few coding mutations have been found, suggesting that the common risk alleles reside in nearby regulatory elements that have yet to be identified. Using a combination of zebrafish and mouse transgenesis, we screened 15 conserved non-coding sequences for enhancer activity, identifying 3 that regulate expression in a tissue specific pattern consistent with endogenous foxe1 expression. These 3, located -82.4 and -67.7 upstream of FOXE1 and another 22.6 kb downstream, are all active in the developing jaws and branchial arches. Two of these, -67.7 and +22.6 are also active in the heart. The -67.7 enhancer also directs expression in the developing thyroid and contains the SNP rs7850258 that is the most significantly associated marker at this locus with both hypothyroidism and thyroid cancer based on genome-wide association studies. Studies to evaluate the functional effects of rs7850258 are ongoing. Our previous studies of isolated cleft lip and cleft palate have indicated the presence of three different FOXE1 risk haplotypes, two of which correspond with the -67.7 and +22.6 enhancers. Given that thyroid diseases and orofacial clefting map to the same locus and that biological function for the affected tissues converge on one enhancer, there may a shared risk for both diseases within families or a population. To this end we evaluated the frequency of thyroid diseases amongst pedigrees segregating for CLP. Personal and family history of thyroid cancer and other thyroid conditions was obtained from case and control families from the Pittsburgh Orofacial Cleft Study (MLM). Interestingly, significantly more case families (27.4%; 248/905) reported history of any thyroid condition than control families (19.66%; 163/829; p-value=0.0002). Focusing on thyroid cancer alone, while there was a greater frequency in cleft families (2.89%) than control (2.43%), the difference was not significant (p=0.64). In conclusion, we have demonstrated there are genetic, biological and clinical links between orofacial clefting and thyroid disease at the FOXE1 locus.

283

Irf6 homeostasis is required for neurulation through a direct interaction with Tfap2a. *Y.A. Kousa[1], H. Zhu[2], A. Kinoshita[3], W.D. Fakhouri[4], M. Dunnwald[5], R.R. Roushangar[1], T.J. Williams[6], B.A. Amendt[7], Y. Chai[8], R.H. Finnell[2], B.C. Schutte[9].* 1) Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI; 2) Dell Pediatric Research Institute, Department of Nutritional Sciences, University of Texas at Austin, 78723 Austin, Texas, USA; 3) Department of Human Genetics, Nagasaki University, Nagasaki, Japan; 4) Department of Diagnostic & Biomedical Sciences, School of Dentistry, University of Texas at Houston, 77054 Houston, Texas, USA; 5) Department of Pediatrics and Interdisciplinary Program in Genetics and Molecular and Cellular Biology, University of Iowa, 52242 Iowa City, Iowa, USA; 6) Department of Craniofacial Biology, University of Colorado Denver at Anschutz Medical Campus, 80045 Aurora, Colorado, USA; 7) Department of Anatomy and Cell Biology, University of Iowa, 52242 Iowa City, Iowa, USA; 8) 8Center for Craniofacial Molecular Biology, Ostrow School of Dentistry, University of Southern California, 90033, Los Angeles, California, USA Irvine, California, USA; 9) Department of Microbiology and Molecular Genetics, Michigan State University, 48824 East Lansing, Michigan, USA.

Mutations in Interferon Regulatory Factor 6 (IRF6) cause Van der Woude syndrome (VWS), an autosomal dominant form of cleft lip and palate. A DNA variant (rs642691) in the IRF6 enhancer MCS9.7 is found in 30% of the world's population and contributes 12% of the risk for nonsyndromic orofacial clefting. In vitro, rs642691 abrogates one of four TFAP2 binding sites within MCS9.7 and increases risk for orofacial clefting. Mutations in TFAP2A cause branchio-oculo-facial syndrome, which typically includes an orofacial cleft. In the mouse, Tfap2a is also required for neurulation. In this work, we dissected the nature of Irf6-Tfap2a interaction. We show that AP-2a is necessary for complete MCS9.7 enhancer activity in the mouse, but that AP-2a was ectopically expressed in Irf6-/- embryos. To determine the mechanism of this interaction, we altered the dosage of these two genes in mutant mice. We showed that 10% of Tfap2a+/- embryos have exencephaly. However, consistent with a negative feedback loop, the Irf6 null allele completely rescues Tfap2a+/- embryos. Furthermore, we observed that 19% of transgenic embryos over-expressing Irf6 have neural tube defects (NTD) that phenocopy Tfap2a heterozygous and knockout embryos. At the molecular level, we found that Irf6 and AP-2a co-localized in non-neural ectoderm of wild type embryos. However, in embryos that over-express Irf6, there was a loss of AP-2a expression in early delaminating neural crest cells. These data suggest that Tfap2a and Irf6 interact through a negative feedback loop. In addition, we found that compound heterozygous embryos for the Irf6 null and hypomorphic alleles have a curly tail, a NTD considered to be analogous to spina bifida in humans. Thus, both over-expression and under-expression of Irf6 can lead to NTD in the mouse. Finally, to assess the role of IRF6 in human neurulation, we sequenced 192 spina bifida patients. We found a nonsynonymous substitution at a highly conserved amino acid in exon 9 of IRF6 (D427Y) that is predicted to be damaging by PolyPhen/SIFT and is not found in control databases. Interestingly, D427Y was previously identified in a patient with VWS. While IRF6 haploinsufficiency has not been associated with spina bifida, our murine studies support the possibility that IRF6 mutations can lead to human NTDs. These results establish a negative feedback loop between Irf6 and Tfap2a and identify a role for Irf6 in neurulation.

284

Deletion of a distant-acting enhancer on Chr16p13.3 causes recessive Intractable Diarrhea of Infancy Syndrome. *D. Oz-Levi[1], I. Bar-Joseph[3,4], T. Olender[1], EK. Ruzzo[5], D. Yagel[3], H. Reznik-Wolf[3,4], A. Alkelai[1], R. Milgrom[1], C. Hartman[6], R. Shamir[6], R. Kleta[7,8], L. Pennacchio[9], DB. Goldstein[5], E. Pras[3,4], Y. Anikster[2,4], D. Lancet[1].* 1) molecular genetics, weizmann institute of science, Rehovot, Rehovot, Israel; 2) Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Ramat Gan, Israel; 3) The Danek Gertner Institute of Human Genetics, Sheba Medical Center, Ramat Gan, Israel; 4) The Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel; 5) Center for Human Genome Variation, Duke University School of Medicine, Durham, North Carolina 27708, USA; 6) Schneider children's Medical Center, PetachTikva, Israel; 7) Centre for Nephrology, University College London, London, UK; 8) Institute of Child Health, University College London, London, UK; 9) Genomics Division, MS 84-171, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

We report on ten patients from eight seemingly unrelated families of Jewish Iraqi origin who presented with a recessive form of Intractable Diarrhea of Infancy Syndrome (IDIS). This is an early onset non-infectious malabsorptive diarrhea starting within the first few weeks of life, and the patients receive prolonged intravenous nutrition. Exome sequencing together with homozygosity mapping identified two independent deletion alleles of noncoding DNA (deltaS, deltaL) on 16p13.3, overlapping with falsely predicted exons and present homozygously or as compound heterozygote. The 1500bp overlap region of deltaS, deltaL was tested in a transgenic embryonic mouse assay and found to act as a robust enhancer in the stomach, duodenum and pancreas at days e11.5 to e13.5. Candidate target genes in the vicinity of this enhancer, C16ORF91, the CCSMST1 protein precursor; UNKL, a RING finger transcription factor and CCDC154, a scantly annotated coiled-coil protein, are being scrutinized. One hypothesized mechanism of action being considered is that embryonic modified expression of a transcription factor could interfere with a developmental process, leading to adult abnormal gastrointestinal function. Currently, isolation of enhancer positive intestinal cells at e13.5 followed by RNA-seq is in progress to uncover gene expression changes and help identify the target gene. Beyond, mouse homozygous knockout of the enhancer would reveal its role in gastrointestinal dysfunction. Our findings constitute a rare case of noncoding deletion underlying human disease etiology, a finding likely to become more widespread with the increasing use of whole genome sequencing.

285

The role of SIX6 in primary open-angle glaucoma. *M. Ulmer[1], Y. Liu[2], E. Oh[2], Y. Liu[1], L. Pasquale[3], J. Wiggs[3], A. Ashley-Koch[1], R. Allingham[4], M. Hauser[1,4], NEIGHBORHOOD Consortium.* 1) Center for Human Genetics, Duke University, Durham, NC; 2) Center for Human Disease Modeling, Duke University, Durham, North Carolina; 3) Department of Ophthalmology, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts; 4) Department of Ophthalmology, Duke University School of Medicine, Durham, North Carolina.

Purpose: To follow up a genome-wide association by identifying sequence variants within and near the SIX6 gene, to assess their association with primary open angle glaucoma (POAG), and to perform functional testing. Methods: Meta-analyses of the NEIGHBOR and GLAUGEN datasets (>6500 samples) found significant association between POAG disease status and the SIX6 locus on chromosome 14q23 (top SNP rs10483727, p-value = 3.9 × 10-11, OR= 1.32). SIX6 is a homeobox gene that plays a role in ocular development; it has been associated with quantitative optic nerve endophenotypes and haploinsufficiency in humans causes bilateral anophthalmia. To identify potential causal variants, we sequenced a well described enhancer element, the promoter, and the exons of the SIX6 gene in 262 POAG cases and 279 controls. Variants identified are currently being functionally tested using morpholino antisense oligonucleotide technology in developing zebrafish and dual luciferase assays. Results: The common, nonsynonymous variant, rs33912345 (His141Asn), was identified by sequencing and is significantly associated with POAG (p-value = 4.2 × 10-10, OR = 1.27). This variant is in strong LD with rs10483727 (r2 = 0.95) first identified in the NEIGHBOR/GLAUGEN meta-analysis. We identified five additional nonsynonymous variants within the SIX6 gene, (Glu93Gln, Glu129Lys, Leu205Arg, Thr212Met, and Ser242Ile). Morpholino knockdown of SIX6 in the zebrafish results in reduced eye size. This phenotype can be rescued by the wild-type human SIX6 cDNA, but at least 3 variants identified by sequencing prevent full functional rescue (His141Asn, Thr212Met, Ser242Ile; all p < 0.0001). Within the SIX6 enhancer, 5 variants were identified, including 4 novel variants. These are currently being tested to determine their effects on SIX6 expression levels using dual luciferase assays. Conclusions: SIX6 plays a pivotal role in retinal development including direct regulation of retinal ganglion cells, the cell type whose apoptotic death leads to blindness in POAG patients. Through direct in vivo testing, we have demonstrated that rs33912345 (His141Asn) has reduced functional activity and may well be the primary driver of the genome-wide association signal in this locus. We have also identified rare protein coding variants that affect gene function in vivo, as well as variants in the regulatory regions of the SIX6 gene that may be functional through altering gene dosage.

286

A dominant-negative *GFI1B* mutation causes autosomal dominant gray platelet syndrome. *L. Van Laer[1], D. Monteferrario[2], N.A. Bolar[1], K. Hebeda[3], S. Bergevoet[2], H. Veenstra[2], B. Laros[2], M. MacKenzie[2], E. Fransen[1,5], G. Van Camp[1], F. Preijers[2], S. Salemink[4], W. van Heerde[2], G. Huls[2], J.H. Jansen[2], M. Kempers[4], B.A. van der Reijden[2], B.L. Loeys[1].* 1) Center of Medical Genetics, Antwerp University Hospital/University of Antwerp, Antwerp, Belgium; 2) Department of Laboratory Medicine, Laboratory of Hematology Nijmegen Centre of Molecular Life sciences, Radboud University Medical Centre, Nijmegen, The Netherlands; 3) Department of Pathology, Radboud University Medical Centre, Nijmegen, The Netherlands; 4) Department of Clinical Genetics, St. Radboud University Medical Centre, Nijmegen, The Netherlands; 5) StatUa Center for Statistics, University of Antwerp, Antwerp, Belgium.

Gray platelet syndrome (GPS) is a hereditary, usually autosomal recessive bleeding disorder caused by defective production of α-granules in platelets. It has been proposed that the α-granule deficiency in GPS can be attributed to the failure of megakaryocytes to efficiently route endogenous proteins into α-granules, thereby hampering their maturation. Here we describe a large family with an autosomal dominant type of GPS characterized by mild to severe bleeding complications. In addition to large gray platelets lacking α-granules, other GPS-associated phenomena like thrombocytopenia, emperipolesis, myelofibrosis and lower expression of platelet factor 4 were observed in affected individuals. To determine the disease causing mutation we first performed linkage analysis and identified a candidate locus on chromosome 9q34 with a LOD score of 3.9. We considered *GFI1B* (Growth factor independence 1B), located within this region, an excellent candidate gene because of its function as a transcriptional repressor essential in erythroid and megakaryocyte lineage development. Sequence analysis identified a nonsense mutation in exon 6 (c.859C>T, p.Gln287*) that co-segregated with the GPS in this family. The mutated transcript was not targeted for nonsense-mediated mRNA decay, resulting in a protein truncated within its 5th zinc finger, a domain essential for DNA binding. Using luciferase reporter assays we demonstrated that the truncated GFI1B protein was unable to repress gene expression and that it inhibited wild type GFI1B in a dominant-negative manner. Subsequently, we performed an immunophenotypical analysis of the peripheral blood of affected family members. Myeloid and erythroid lineages were unaffected, but we observed aberrant platelets, dysplastic abnormalities and disturbed lineage marker expression in GFI1B-mutated megakaryocytes. These studies define GFI1B to be key to proper megakaryocyte development, platelet production and α-granule biogenesis.

287

Identification of disease causing mutations in a new congenital neutrophil defect syndrome. *T. Vilboux[1], A. Lev[2], M.C. Malicdan[1], S. Amos[2], R. Sood[3], Y. Anikster[4], C. Klein[5], W.A. Gahl[1], R. Somech[6].* 1) Medical Genetics Branch, NHGRI, NIH, Bethesda, MD; 2) The Claim Sheba Medical Center, Sheba Cancer Research Center, Pediatric Immunology, JMF center for PID; 3) Zebrafish Core Genetics and Molecular Biology Branch; 4) Edmond and Lily Safra Children's Hospital Sheba Medical Center, Sackler School of Medicine, Tel Aviv University, Metabolic Disease Unit; 5) Ludwig Maximilians University Munich, Pediatrics; 6) Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Tel Hashomer. Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv Israel, Pediatrics B.

Neutrophils are the predominant phagocytes that protect the host against bacterial and fungal infections. The important role of neutrophils in innate immunity is underscored by the multiple congenital defects that lead to severe infections. These disorders are characterized by genetic abnormalities that alter the neutrophil number and/or their function. To identify the causative gene, we combined the data of single-nucleotide polymorphism arrays of multiple families and a single exome. Functional studies were performed on patient neutrophils and fibroblasts, and on a zebrafish model. We describe seven patients from five families, who displayed a unique constellation of severe neutrophil dysfunction, bone marrow fibrosis and nephromegaly. From our analysis, this new immunodeficiency is associated with biallelic mutations in the *VPS45* gene, encoding a protein that regulates membrane traffic through the endosomal system. Two distinct homozygous *VPS45* mutations were identified in patients from different ethnic origins. Functional studies showed that *VPS45* mutations affected protein recycling, altered cell migration and neutrophil maturation, and induced cell apoptosis. Both the migration defect and the enhanced apoptosis were rescued by expressing the wild-type *VPS45* in patient cells. Neutrophil defect was recapitulated in a zebrafish model. Mutations in *VPS45* underlie a new immunodeficiency syndrome with bone marrow fibrosis and severe neutropenia. This can pave the way into understanding and finding more immunodeficiencies that are associated with endocytosis or phagosome defect.

288

A functional variant in the *CFI* gene confers a high risk of age-related macular degeneration. *A. den Hollander[1,14], J.P.H. van de Ven[1], S.C. Nilsson[2], P.L. Tan[3,4,5], G.H.S. Buitendijk[6,7], T. Ristau[8], F.C. Mohlin[2], S.B. Nabuurs[9], F.E. Schoenmaker-Koller[1], D. Smailhodzic[1], P.A. Campochiaro[10,11], D.J. Zack[10,11], M.R. Duvvari[1], B. Bakker[1], C.C. Paun[1], C.J.F. Boon[1], A.G. Uitterlinden[7,12], S. Liakopoulos[8], B.J. Klevering[1], S. Fauser[8], M.R. Daha[13], N. Katsanis[3,4,5], C.C.W. Klaver[6,7], A.M. Blom[2], C.B. Hoyng[1].* 1) Dept Ophthalmology, Radboud Univ Nijmegen MC, Nijmegen, Netherlands; 2) Section of Medical Protein Chemistry, Dept Laboratory Medicine Malmö, Lund University, Malmö, Sweden; 3) Center for Human Disease Modeling, Duke University, Durham, NC; 4) Dept Cell Biology, Duke University, Durham, NC; 5) Dept Pediatrics, Duke University, Durham, NC; 6) Dept Ophthalmology, Erasmus MC, Rotterdam, the Netherlands; 7) Dept Epidemiology, Erasmus MC, Rotterdam, the Netherlands; 8) Center for Ophthalmology, University of Cologne, Cologne, Germany; 9) Center for Molecular and Biomolecular Informatics, Radboud Univ Nijmegen MC, Nijmegen, the Netherlands; 10) Dept Ophthalmology, Johns Hopkins School of Medicine, Baltimore, MD; 11) Dept Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD; 12) Dept Internal Medicine, Erasmus MC, Rotterdam, the Netherlands; 13) Dept Nephrology, Leiden University MC, Leiden, the Netherlands; 14) Dept Human Genetics, Radboud University Nijmegen MC, Nijmegen, the Netherlands.

Age-related macular degeneration (AMD) is the most common cause of blindness among the elderly. Genome-wide association studies have identified common variants in or near genes within the complement cascade, high-density lipoprotein cholesterol pathway, and extracellular matrix that are associated with AMD. Combined, these variants explain up to half of the heritability of AMD. To explore the role of rare, highly penetrant variants in AMD we focused on *CFI*, a gene that encodes Factor I, a serine protease that circulates in its inactive form and can inactivate all complement pathways by cleaving the α' chain of activated complement factors C3b and C4b. We identified a rare missense mutation, p.Gly119Arg, in 20 of 3,567 cases versus only one of 3,937 controls, consistent with p.Gly119Arg conferring a high risk for developing AMD ($P = 3.79 \times 10^{-6}$, odds ratio = 22.20; 95% confidence interval 2.98-164.49). The p.Gly119Arg mutation affects a highly conserved glycine residue in the CD5 domain of Factor I and is predicted to alter the packing and stability of this domain. The cases carrying the p.Gly119Arg variant had lower plasma Factor I levels compared to both controls and cases without a *CFI* variant. Plasma and serum samples of cases carrying the p.Gly119Arg substitution mediated the degradation of C3b both in the fluid phase and on the cell surface at a lower level compared to controls. Recombinant protein studies revealed that the p.Gly119Arg mutant protein is both expressed and secreted at lower levels than wildtype Factor I protein. Consistent with these findings, 119Arg-encoding human mRNA had reduced activity compared to 119Gly in regulating vessel thickness and branching in the zebrafish retina. This study provides the first direct evidence that Factor I deficiency confers a high risk of developing AMD. Taken together, these findings demonstrate that rare, highly penetrant mutations contribute to the genetic variance of AMD, with important implications for predictive testing and personalized treatment.

289

Recurrent genomic mutation 507delT in three Lipoid Proteinosis (Urbach-Wiethe) pedigrees from central Iran. *L. Youssefian[1], H. Vahidnezhad[1,2], M. Daneshpazhooh[3], S. Abdollahzadeh[4], H. Talari[5], C. Chams-Davatchi[3], S. Akhondzadeh[6], R. Mobasher[3], M. Tabrizi[1].* 1) Department of Medical Genetics, Tehran University of Medical Sciences, Tehran, Tehran, Iran; 2) Molecular Medicine Division, Biotechnology Research center, Pasteur institute of Iran, Tehran, Iran; 3) Department of Dermatology, Bullous Diseases Research Center, Tehran University of Medical Sciences, Tehran, Iran; 4) Department of Neurosurgery, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran; 5) Radiology Department and Anatomic Science Research Center, Kashan University of Medical Sciences,Kashan, Iran; 6) Psychiatric Research Center, Roozbeh Hospital, Tehran University of Medical Sciences, Tehran, Iran.

Lipoid proteinosis reports from Iran are rare. Recently, a mutation has been reported from Iran. We aimed to collect and analyze a large number of individuals in affected families and conduct genetic testing especially in isolated areas of Iran where consanguineous marriages are very common and thus genetic conditions pose a great burden on the society. Genetic counseling and follow-up of such isolated populations has great potential and can drastically improve living conditions. Genetic testing can be offered to such populations and for those who continue to engage in consanguineous marriages can have the option of PND. Diagnosis of Lipoid Proteinosis had been established by dermatologtists previously and patients with their families were referred for genetic counseling. After obtaining written informed consent from families, extensive family history was documented on-site in three separate villages within each other's 20-30 km vicinity. Patients were coordinated by the local health service center. DNA was extracted from patient blood samples. PCR reaction was set up for ten exons and nearby intronic regions of the ECM1 gene. PCR products from twelve patients were sequenced and analyzed. All patients harbored the 507delT mutation in exon 6. This mutation generates a new SmaI restriction site at the mutation site. To see if PCR-RFLP can distinguish affected individuals and heterozygotes from healthy homozygotes, thirty-eight individuals were screened by PCR-RFLP. Results documented ten affected individuals (324 bp and 339 bp fragments), nine healthy homozygotes (one 663 bp fragment), and nineteen heterozygotes (663 bp, 324 bp, and 339 bp fragments). Five affected individuals from various age groups underwent MRI and CT scans. Five out of Five demonstrated signal void area in bilateral amygdalae which was confirmed as calcification on CT. Three of the patients with calcification also reported periodic episodes of seizure. Although all our patients harbored the 507delT mutation, all possible clinical presentations were observed; therefore, this unique population negates the possibility of establishing a genotype/phenotype correlation for Lipoid proteinosis in this region of Iran. In conclusion, we can provide genetic testing to affected individuals and propose the possibility of using PCR-RFLP as a cheap and quick screen of the heterozygotes in the population. Genetic counseling,follow-up and education could reduce burden of this disease in Iran.

290

The neuronal endopeptidase *ECEL1* is a frequent cause of autosomal recessive distal arthrogryposis associated with limited knee flexion, ptosis, and limb muscle and tongue atrophy. *K. Dieterich[1,2,3,4], S. Quijano-Roy[5], N. Monnier[1,13], J. Zhou[3], J. Faure[1,13], D. Avila-Smirnow[5], RY. Carlier[6], C. Laroche[7], P. Marcorelles[8], S. Mercier[9], A. Megabarne[10], S. Odent[9], N. Romero[11], D. Sternberg[12], I. Marty[1], B. Estournet[5], PS. Jouk[2,4], J. Melki[3], J. Lunardi[1,13].* 1) INSERM U836 Grenoble Institut des Neurosciences, Grenoble, France; 2) Département de Génétique et Procréation, Hôpital Couple Enfant, CHU Grenoble, Grenoble, France; 3) INSERM UMR-788, University of Paris 11, Biomedical Institute of Bicêtre, Le Kremlin-Bicêtre, France; 4) Centre de Référence des Anomalies du Développement, Hôpital Couple Enfant, CHU Grenoble, Grenoble, France; 5) AP-HP, Service de Pédiatrie, Hôpital Raymond Poincaré, Garches; Hôpitaux Universitaires Paris-Ile-de-France Ouest, Pôle pédiatrique; Centre de Référence Maladies Neuromusculaires GNMH; CIC-IT - Faculté 6) AP-HP, Service d'Imagerie Médicale, Hôpital R. Poincaré, Garches; Hôpitaux Universitaires Paris-Ile-de-France Ouest, Pôle neuro-locomoteur; GNMH; UVSQ, France; 7) Service de Pédiatrie, CHU Limoges, Limoges, France; 8) Laboratoire d'Anatomo-Pathologie, CHU Brest, Brest, France; 9) Service de Génétique Clinique, CHU Rennes, Université Rennes 1, UMR62920, Rennes, France; 10) Unité de Génétique Médicale et Laboratoire Associé INSERM UMR_S910, Pôle Technologie Santé, Université Saint Joseph, Beyrut, Lebanon; 11) Service d'Anatomopathologie, Pavillon Riessler, Institut de Myologie, Paris, France; 12) Laboratoire de Cardiogénétique et myogénétique moléculaire et cellulaire, GH Pitié-Salpétrière, AP-HP, Paris; 13) Laboratoire de Biochimie Génétique et Moléculaire, Département de Biochimie, Pharmacologie, Toxicologie, CHU de Grenoble, Grenoble, France.

Distal arthrogryposis (DA) constitutes a frequent but heterogeneous subgroup among multiple congenital contractures (MCC). Despite its frequency only a limited number of genes have been associated with rare but well characterized types of DA, implicating almost exclusively genes of the contractile apparatus. Our aim was to identify new genes associated with DA. We therefore performed a SNP array based homozygosity mapping approach in two consanguineous African families presenting with an unclassified DA phenotype. We further screened potential candidate genes in 18 familial or sporadic cases with DA that did not show mutations in known genes associated with DA. Combined multipoint linkage analysis of the two families revealed an overlapping locus at 2q37 of 5.7 Mb with a LOD score Zmax = 5.1 at ϑ = 0.0 and harboring 77 annotated genes. We excluded pathogenic mutations in the genes encoding the gamma (*CHRNG*) and delta (*CHRND*) subunits of the acetylcholine receptor at the neuromuscular junction known to cause MCC. We then searched for candidate genes of the neuromuscular apparatus. The Endothelin Converting Enzyme Like 1 (*ECEL1*) gene is predominantly expressed during fetal life and encodes a neuronal endopeptidase crucial for intramuscular motoneuron branching during embryogenesis and was therefore highly candidate. Screening for *ECEL1* mutations in the total of 20 familial and sporadic cases identified homozygous or compound heterozygous mutations (four non-sense, two splicing and one missense mutation) in six families including the two consanguineous African families. Detailed clinical evaluation showed a recognizable phenotype with limited knee flexion, pes talus or talus valgus, severe muscle atrophy predominant on the lower limbs and the tongue, and ptosis. Muscle MRI evidenced muscle atrophy and frequent fatty replacement, especially in biceps femoris and vastus lateralis. Electrophysiological analyses excluded a neuromuscular transmission defect after birth. Interestingly, *ECEL1*-deficient mice show acetylcholine receptor deficiency of prenatal onset due to deficient axonal branching of motor neurons. The prenatal expression of *ECEL1* and the non-progressive disease state in our patients suggests a developmental defect of the lower motoneuron or neuromuscular junction in patients with *ECEL1* mutations. Further work is needed to better understand the role of *ECEL1* in human, but our findings already highlight a new mechanism for DA.

291

Expanding molecular basis for rasopathies - a new player? *M. Ludwig[1,2], C.Y. Hung[1], J.L. McCauley[1], J. Dallman[3], E.J. Back[3], I. Mihalek[4], G.X. Shi[5], D.A. Andres[5], O. Bodamer[1]*. 1) Department of Human Genetics, University of Miami - Miller School of Medicine, Miami, Florida, USA; 2) Department of Otorhinolaryngology, Medical University of Vienna, Vienna, Austria; 3) Department of Biology, University of Miami, Miami, Florida, USA; 4) Bioinformatics Institute A*STAR Singapore, Singapore; 5) Department of Molecular and Cellular Biochemistry, University of Kentucky - College of Medicine, Lexington, Kentucky, USA.

Background: Gain-of-function mutations in 13 different genes of the ras-MAPK-pathway cause a heterogeneous group of genetic disorders known as rasopathies. The clinical phenotypes of these disorders overlap due to the shared pathway, including Costello (OMIM #218040), CFC (OMIM #115150), LEOPARD (OMIM #151100), Noonan syndrome (OMIM #163950) and others. Patient: We report a 24 month-old boy born full term via C-section following a pregnancy complicated by pre-eclampsia and polyhydraminos. Postnatally he presented with prominent overgrowth: birth weight: 5.14 kg (>97th percentile), birth length: 53.5 cm (90th percentile) and head circumference: 38.5 cm (>90th percentile). His parents are of Cuban-Hispanic descent, healthy and non-consanguineous. Physical examination at presentation revealed a phenotype reminiscent of Costello syndrome, including facial dysmorphism, severe developmental delay, muscular hypotonia, broad fingers with deep palmar and plantar creases, congenital heart defects and non-progressive hypertrophic cardiomyopathy. Methods and Results: Sanger sequencing revealed no pathogenic mutation in any of the known rasopathy genes including BRAF, CBL, HRAS, KRAS, MAP2K1, MAP2K2, NF1, NRAS, PTPN11, RAF1, SHOC2, SOS1 and SPRED1. The high density SNP microarray (Illumina HumanOmni1-Quad BeadChip, >1 million markers) showed no copy number variation. Whole exome sequencing (Illumina HiSeq2000; 82 million pass filter paired-end reads; average depth of 0.74 at 20X) revealed a de-novo missense variant in a novel gene predicted to be within the ras-MAPK-pathway. The variant was confirmed by Sanger sequencing, absent in 654 Hispanic control alleles and not reported in the SNP database (NCBI). The gene is expressed in all tissues and conserved throughout all species. The de-novo variant was introduced into PC6 cells, leading to increased MEK-ERK signaling in comparison to wild-type. This gain-of-function is characteristic for rasopathies. Conclusion: We identified a gain-of-function variant in a gene within the ras-MAPK-pathway associated with a characteristic clinical phenotype. Additional testing is ongoing to further delineate the functional consequences including a zebrafish model. Additional patients with a similar phenotype without an underlying molecular cause are currently screened for variants in this gene. Note: The name of the gene will be disclosed at the meeting.

292

MC3R modifies CF lung disease by increasing the level of CFTR. *J. Park[1], J. Pilewski[2], D. Belchis[3], S. Blackman[4], G.R. Cutting[1]*. 1) Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 2) Departments of Medicine and of Cell Biology and Physiology, University of Pittsburgh, Pittsburgh, PA; 3) Pathology, Johns Hopkins School of Medicine, Baltimore, MD; 4) Pediatric Endocrinology, Johns Hopkins University, Baltimore, MD.

Lung disease accounts for 90% of the mortality in cystic fibrosis (CF), an autosomal recessive disorder caused by mutations in the *CFTR* gene affecting 70,000 individual worldwide. Heritability of lung disease severity estimated from affected twin analysis (H^2=0.54) indicates a prominent role for genetic modifiers. Linkage analysis of affected siblings identified a region on 20q13.2 (LOD=5.03) harboring a modifier of lung function. The 1.3 Mb region of maximum linkage (1 LOD drop) contains only 5 genes, one of which is the melanocortin-3-receptor (*MC3R*). *MC3R* is a compelling candidate for modifying CF lung disease as an empiric screen for new CF therapies using siRNAs revealed that reduction of MC3R resulted in restoration of chloride channel activity to cells expressing CFTR bearing the common CF mutation F508del. To assess whether *MC3R* is a modifier of CF lung disease, we searched for mutations, analyzed its tissue expression and evaluated its effect on CFTR. Sequencing of *MC3R* in 150 sibpairs with CF identified 4 common and 4 rare variants. Three rare coding variants showed evidence of association with lung disease severity (as measured by survival-adjusted CF-specific FEV_1 Z-score) [g.54824368A>C (ΔZ-score=+1.6, $p<0.05$), rs41274722 (ΔZ-score=-1.2, $p<0.03$), rs61735259 (ΔZ-score=+1.5, $p<0.01$)]. *MC3R* mRNA was detected by RT-PCR in normal human bronchial epithelial (HBE) cells and CF bronchial epithelial (CFBE) cells homozygous for F508del-CFTR. MC3R was detected in the lysates of primary airway epithelial cells by Western blot (n=5). IHC staining with 2 anti-MC3R antibodies (Abcam & LS Bio) revealed a signal in surface airway epithelial cells of normal and CF lung. A reduction in MC3R expression by 40-79% in K562 cells using siRNAs increased the quantity of heterologously expressed GFP-WT-CFTR (298%, $p<0.0001$), GFP-F508del-CFTR (479%, $p<0.03$) and F508del-CFTR (336%, $p<0.04$). Similarly, transfection of CFBE cells with *MC3R* siRNAs resulted in a dramatic increase in the amount of endogenously expressed F508del-CFTR. Together, these results suggest that MC3R modifies CF lung disease by increasing the level of mutant forms of CFTR, such as F508del, that retain residual function, which enables recovery of CFTR-mediated chloride transport to airway epithelia. Multiple synthetic ligands have been developed that selectively target melanocortin receptors suggesting a novel therapeutic approach to CF lung disease based on targeting of a genetic modifier.

293

Combined exome and whole-genome sequencing identifies mutations in *ARMC4* as a cause of Primary Ciliary Dyskinesia. *A. Onoufriadis[1], A. Shoemark[2], M. Munye[3], C. James[4], EM. Rosser[5], C. Bacchelli[4], SL. Hart[3], M. Schmidts[1], JE. Danke-Roelse[6], G. Pals[7], C. Hogg[2], EMK. Chung[8], . UK10K[9], HM. Mitchison[1].* 1) Molecular Medicine Unit and Birth Defects Research Centre, Institute of Child Health, University College London, London, UK; 2) Department of Paediatric Respiratory Medicine, Royal Brompton and Harefield NHS Trust, London, UK; 3) Molecular Immunology Unit, Institute of Child Health, University College London, London, UK; 4) Centre for Translational Genomics - GOSgene, Institute of Child Health, University College London, London, UK; 5) Clinical Genetics Unit, Great Ormond Street Hospital, London, UK; 6) Department of Pediatrics, Atrium Medical Center, Heerlen, the Netherlands; 7) Department of Clinical Genetics, VU University Medical Center, Amsterdam, the Netherlands; 8) General and Adolescent Paediatric Unit, Institute of Child Health, University College London, London, UK; 9) uk10k.org.uk.

Primary ciliary dyskinesia (PCD) is a recessively inherited, genetically heterogeneous disorder arising from abnormal function of motile cilia and sperm flagella. It affects one per 15,000-30,000 births. Abnormal motility of cilia and flagella leads to symptoms including neonatal respiratory distress, chronic respiratory infections and obstructive lung disease, otitis media, subfertility and situs abnormalities. So far, mutations in 20 genes have been identified to cause PCD including *DNAH5, DNAH11, DNAI1, DNAI2* and *DNAL1* which encode subunits of the axonemal outer dynein arm (ODA) and *CCDC114* which encodes an ODA docking complex component. Exome-sequencing was performed in one PCD patient at the Wellcome Trust Sanger Institute (Cambridge, UK) as part of the UK10K project. This affected individual is from a consanguineous marriage so we focused on homozygous non-synonymous or splice-site substitutions or indels, that were novel or present in the 1000 Genomes Project with a frequency <0.01. To prioritize candidate genes we used our internal allele count data, removing variants detected more than ten times across the UK10K cohort, as PCD causing mutations would not likely appear in multiple well-phenotyped non-PCD patients. This filtering strategy revealed a homozygous protein-truncating nonsense variant in *ARMC4* (c.2675C>A; pSer892*), a gene previously shown to be involved in ciliogenesis. In parallel, whole genome sequencing was performed in the two parents of a consanguineous marriage since material from their affected offspring was not available. We filtered per chromosome, to identify protein altering heterozygous variants shared by both parents with a 1000 Genomes MAF<0.01. This strategy revealed a second *ARMC4* protein-truncating nonsense variant (c.1972G>T; p.Glu658*). Segregation analysis in all available members of the pedigrees confirmed recessive inheritance of both c.2675C>A and c.1972G>T variants. Respiratory cilia from affected individuals were immotile and ultrastructural analysis by transmission electron microscopy showed loss of the ciliary outer dynein arms. *ARMC4* transcript levels were investigated by means of qPCR in respiratory epithelial cells during ciliogenesis, showing a highly significant increase of *ARMC4* expression in ciliated cells when compared to non-ciliated basal cells. Together these data suggest that *ARMC4* mutations cause PCD due to loss of the outer dynein arms.

294

Identification of Novel Molecular Defects in Chronic Intestinal Pseudo-Obstruction (CIPO). *G. Romeo[1], E. Bonora[1], F. Bianco[1], L. Cordeddu[2], M. D'Amato[2], G. Lindberg[2], M. Bamshad[3], D. Nickerson[3], Z. Mungan[4], K. Cefle[4], S. Palanduz[4], S. Ozturk[4], T. Ozcelik[5], A. Gedikbasi[6], V. Stanghellini[1], R. Cogliandro[1], E. Boschetti[1], C. Graziano[1], M. Seri[1], R. De Giorgio[1], University of Washington Center for Mendelian Genomics (UW CMG).* 1) Medical Gen Unit, Univ Bologna, Bologna, Italy; 2) Karolinska Institutet, Stockholm, Sweden; 3) University of Washington, Seattle, U.S.A; 4) University of Istanbul, Istanbul, Turkey; 5) Bilkent University, Ankara, Turkey; 6) Bakirkoy Sadi Konuk Hospital, Istanbul, Turkey.

INTRODUCTION: CIPO is characterized by severe impairment of gut motility mimicking a mechanical sub-occlusion in the absence of occluding causes. Although most patients are sporadic, some CIPO cases show familial clustering with a likely genetic origin. We wanted to identify the gene involved in a previously studied large Turkish consanguineous pedigree with recurrence of CIPO together with megaduodenum, Barret esophagus and cardiac anomalies (MIM # 611376) and test at the same time whether the genetic alterations found in that pedigree may have any implication in sporadic CIPO. RESULTS: WES analysis revealed a novel homozygous mutation, c.1864 G>A, p.622 Ala>Thr in RAD21 in the consanguineous Turkish pedigree where we had previously mapped CIPO on 8q23-q24 using homozygosity mapping 1. RAD21 belongs to the cohesin complex regulating cell replication and gene expression. The reported change cosegregated with the disease and was not detected in 1000 Turkish control chromosomes. PolyPhen-2 predicted it as probably damaging (HumDiv score: 0.999), since it resides in a conserved position. RAD21 screening in 12 Swedish and 24 Italian well-characterized sporadic CIPO patients did not show any coding mutation. However, we established a relationship between RAD21 and sporadic CIPO starting from our finding related to the expression of APOB 2. APOB mRNA and protein levels were found to be increased in CIPO patients vs. controls. In silico analysis of the human APOB promoter identified three RAD21-putative binding sites. EMSA with anti-RAD21 antibodies showed that RAD21 binds to these regions. A significant increase in APOB and RAD21 immunoreactive cells was also detected in the lamina propria of gut biopsies of CIPO patients vs. controls. CONCLUSION: We identified a novel mutation of RAD21 in a familial syndromic form of CIPO and an altered expression of RAD21 and APOB in sporadic CIPO. These previously uncharacterized pathways contribute to clarify the pathogenesis of CIPO and bear clinical implications for the diagnosis and prognosis of this disabling condition. ACKNOWLWDGEMENT: Sequencing was provided by the University of Washington Center for Mendelian Genomics (UW CMG) funded by NIH grant 1U54HG006493 to Drs. Debbie Nickerson, Jay Shendure and Michael Bamshad. REFERENCES: 1. Degl'Incerti et al Eur J Hum Genet 15, 889-897, 2007; 2. Evangelisti et al Neurogastroenterology & Motility, 24, 497-508, 2012.

295

Rare highly penetrant variants of late onset Alzheimers disease. *J. Rehker[1], R. Levy[1], R. Nesbitt[1], Q. Yi[2], B. Martin[2], D. Nickerson[2], W. Raskind[2], J. Shendure[2], Z. Brkanac[1].* 1) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA; 2) Department of Genome Sciences, University of Washington, Seattle, WA.

Alzheimer's disease (AD) is the most common cause of dementia in US, affecting approximately 5% of individuals older than age 70. Highly penetrant mutations in three genes (APP, PSEN1 and PSEN2) are responsible for majority of cases of early-onset disease (<60-65 years). Late onset disease (age >60-65 years) represents more then 95% of cases and it is familial in 15% to 25%. The main risk factor for late-onset disease is the ApoE gene that contributes to some of family clustering. The whole genome association studies have identified approximately dozen low-risk genes associated with the disease. Recently rare variants in TREM2 gene were identified as significant AD risk factor with odds ratio comparable to ApoE. This indicates that additional rare highly penetrant variants remain to be found. To detect additional highly penetrant AD genes we have identified 19 families with four or more affected cases in NIA and NIMH AD collections. In our sample average age of onset of the disease was 68.97 ± 10.55 years. Under assumption that functional private variants are causal, exome sequencing allows for gene identification in small families. For each person the exome on average has less then 200 private functional variants. We define private functional variants as not seen in 1000 genomes, ESP6500 and dbSNP135 data and having potential for function altering like missense, frameshift, gain of stop codons or changes in splicing sites. Exomes were sequenced at University of Washington Northwest Genomic Center. Sequencing, read alignment and variant calling were performed using stringent parameters for discovery of private variants. We have sequenced 37 exomes. Our variant call and filtering criteria identified 1-23 functional private variants shared between exome sequenced cases in each family. Identified variants were confirmed with capillary sequencing and present in remaining affected cases in each family. This analysis has resulted in the identification of 1-2 private functional variants that are present in all, or most affected cases in each family. To identify genes associated with AD for 23 candidate genes identified in our families we are performing a large case control study involving ~ 1000 cases of familial AD and 500 non-demented elderly controls. For case-control study we are employing gene-based approach and sequencing of candidate genes using molecular inversion probes and next generation sequencing.

296

The identification of high-penetrance variants in late-onset Alzheimer disease by whole-exome sequencing in extended families. *M.A. Kohli[1], B.W. Kunkle[1], A.C. Naj[2], L.-S. Wang[2], K.L. Hamilton[1], R.M. Carney[1], W.R. Perry[1], P.L. Whitehead[1], J.R. Gilbert[1], E.R. Martin[1], G.W. Beecham[1], R.P. Mayeux[3], J.L. Haines[4], L.A. Farrer[5], G.D. Schellenberg[2], S. Zuchner[1], M.A. Pericak-Vance[1], Alzheimer's Disease Genetics Consortium (ADGC).* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA; 2) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 3) Taub Institute of Research on Alzheimer's Disease, Columbia University, New York, NY, USA; 4) Center for Human Genetics Research, Vanderbilt University, Nashville, TN, USA; 5) School of Medicine, Boston University, Boston, MA, USA.

We chose 6 multi-generational, late-onset Alzheimer disease (LOAD) families to identify rare LOAD risk variants. The pedigrees are consistent with dominant inheritance, have an average of 6 AD-affected subjects, and are free of known AD mutations. We performed whole-exome sequencing (WES) on 4-9 affected subjects per pedigree, comprising 2-6 cousin pairs. For validation, we used the Alzheimer's Disease Genetics Consortium's (ADGC) case-control dataset composed of 13,748 individuals (7,652 cases) genotyped on the Illumina Exome SNP Array (195,039 SNPs with Minor Allele Frequencies (MAF) <2%).

We filtered WES-identified variants for potential high-penetrance AD variants by applying the following criteria: (1) segregation with LOAD allowing for one phenocopy per pedigree, (2) non-synonymous variants at evolutionarily conserved sites predicted to be deleterious, (3) MAF<1%, and (4) allelic case-control association odds ratio (OR) >2, if available, in the ADGC dataset. Nineteen variants passed these criteria, among which only 2 had ADGC data (*MYO3B*, *PPP1R12A*). We further genotyped 4 of the remaining variants in the ADGC dataset. Of these six variants, 3 were exclusively found in cases (in *DYM*, *LONRF3* and *TBC1D8B*). The fourth reached an OR of 1.54 in an analysis excluding controls with age <80 (*RTTN*). Known functions of mentioned genes are actin-binding, regulation of actin-myosin interaction, brain development, and Rab GTPase activation. Of the remaining 13 genes, the androgen receptor (*AR*) has most often been related to AD. In a second approach, we did not require an OR>2 and lowered the segregation criterion to two phenocopies per pedigree. We obtained a total of 60 genes, of which five (*CAPZA3*, *LONRF3*, *LRP1B*, *MYO3B*, *TBC1D17*) showed nominally significant associations in gene-based tests in the ADGC data. Interestingly, *TBC1D17* is another Rab GTPase-activating protein implicated in endocytosis, and *LRP1B* is a low-density lipoprotein receptor that allows endocytosis of a complex of ApoE-containing lipoprotein particles bound to β-amyloid.

These variants are extremely rare and may well qualify as high-penetrance AD variants. Most promising are completely segregating variants in genes that also obtained gene-based nominal significance (*LONRF3*, *MYO3B*), or have been functionally linked to AD (*LRP1B*).

297

Functional rare genetic variation in Alzheimer's disease: an exome-wide association study in the CHARGE consortium. *J. Jakobsdottir[1], S.J. van der Lee[2], J.C. Bis[3], V. Chouraki[4,5], A.V. Smith[1,6], A.L. DeStefano[7,5], J. Brody[3], N. Amin[2], L.J. Launer[8], C.A. Ibrahim-Verbaas[2,9], S. Choi[7,5], A. Beiser[7,5], R. Au[4,5], P.A. Wolf[4,5], O.L. Lopez[10], M.A. Ikram[2], A. Hofman[2], A.G. Uitterlinden[2,11], D. Levy[5,12,4], C.J. O'Donnell[5,12], M.L. Grove[13], E. Boerwinkle[13,14], A.L. Fitzpatrick[15], B. Psaty[16,15], S. Seshadri[4,5], V. Gudnason[1,6], C.M. van Duijn[2].* 1) Icelandic Heart Association, Kopavogur, Iceland; 2) Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 3) Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA; 4) Boston University School of Medicine, Boston, MA, USA; 5) Framingham Heart Study, Framingham, MA, USA; 6) Faculty of Medicine, University of Iceland, Reykjavik, Iceland; 7) Boston University School of Public Health, Boston, MA, USA; 8) Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Bethesda, MD, USA; 9) Department of Neurology, Erasmus Medical Center, Rotterdam, The Netherlands; 10) Department of Neurology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA; 11) Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands; 12) Heart, Lung, and Blood Institute, Framingham, MA, USA; 13) School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA; 14) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; 15) Department of Epidemiology, University of Washington, Seattle, WA, USA; 16) Department of General Internal Medicine, University of Washington, Seattle, WA, USA.

Alzheimer's disease (AD) is a complex disease with a heritability of ~60%. Relatively common SNPs with modest effects on AD risk have been identified as well as high-risk variants that lead to early onset forms of AD. In CHARGE, a large consortium of population-based studies, we tested the hypothesis that functional coding and splice-site variants modulate the risk of AD. Four cohorts (AGES-Reykjavik Study, Cardiovascular Health Study, Framingham Heart Study, and Rotterdam Study) with 1327 AD cases and 8912 controls of European ancestry were genotyped for the ExomeChip (EC). The EC is a comprehensive genotyping array of functional, mostly rare, variation in the human exome. Of ~240,000 variants on the EC 90% are predicted to alter the protein structure and 45% are predicted to be damaging. Over 62,000 samples were jointly called within CHARGE, which allowed for high accuracy in the rare variant genotype calls. In each study we tested for associations of AD to single and multiple variants, grouped by genes, using score tests and then meta-analyzed the scores. The strongest associations to single SNPs were found to be at common SNPs near the *APOE* locus, previously established to affect AD risk (rs769449 in *APOE*, MAF 12%, $p=4\times10^{-34}$). The Sequence-Kernel Association Test (SKAT) was used for the gene-based analysis. We tested ~14,000 genes, containing non-synonymous or splice-site variants of MAF≤5%. 14,000 tests correspond to $p=4\times10^{-6}$ for a Bonferroni multiple testing correction. The SKAT analysis identified the *ERVFRD-1* gene at a suggestive level of exome-wide significance ($p=5\times10^{-5}$, cumulative MAF 2.5%). The most significantly associated SNP in *ERVFRD-1* was rs55714642 (MAF 2%, $p=7\times10^{-6}$), which we observed with the same direction of protective effect in all four studies. Associations were also found with the *SKAP2* and *TM2D3* genes. Due to low cumulative MAF, the association signal in the *SKAP2* gene (cumulative MAF 0.06%, $p=3\times10^{-8}$) was driven by three studies while the signal at *TM2D3* (cumulative MAF 0.23%, $p=5\times10^{-6}$) was driven by one study enriched for the putative high-risk SNP. β-amyloid plaques in the brain are an important characteristic of AD and dementia and all three genes are directly connected to β-amyloid precursor or binding proteins, supporting the role of the β-amyloid pathway in AD. Further studies, such as replication, sequencing, and functional studies are warranted to firmly establish these newly found associations.

298

Whole-exome sequencing in early-onset Alzheimer disease families identifies rare variants in multiple Alzheimer-related genes and processes. *B.W. Kunkle[1], M.A. Kohli[1], B.N. Vardarajan[2], C. Reitz[2], A.C. Naj[3], P.L. Whitehead[1], W.R. Perry[1], E.R. Martin[1], G.W. Beecham[1], J.R. Gilbert[1], L.A. Farrer[3], J.L Haines[4], G.D. Schellenberg[5], R.P. Mayeux[2], M.A. Pericak-Vance[1], Alzheimer's Disease Genetics Consortium.* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA; 2) Taub Institute for Research on Alzheimer's Disease, Columbia University, New York, NY, USA; 3) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 4) School of Medicine, Boston University, Boston, MA, USA; 5) Center for Human Genetics Research, Vanderbilt University, Nashville, TN, USA.

Mutations in APP, PSEN1 and PSEN2 lead to familial, early-onset Alzheimer disease (EOAD). These mutations account for only 60-70% of familial EOAD and ~11% of EOAD overall, leaving the majority of genetic risk for the most severe form of Alzheimer disease (AD) unexplained. We performed Whole-Exome Sequencing in Caribbean Hispanic and Caucasian EOAD families previously screened negative for APP, PSEN1, and PSEN2 to search for rare variants contributing to risk for EOAD. 60 individuals in 21 families were sequenced using the Agilent 50Mb kit on an Illumina HiSeq2000. Variant filtering for segregating, conserved and functional rare variants (MAF<0.1%) was performed on the families assuming both autosomal-dominant and X-linked dominant models. Filtered loci were examined for implication as AD candidate genes from GWAS or in biologically relevant KEGG Pathways. Variants were also followed up for association with AD in 13,748 individuals (7,652 affected) from the Alzheimer's Disease Genetics Consortium (ADGC) genotyped on the exome chip, which included 195,039 variants with MAF<2%. Enrichment analysis of the variant list was conducted using DAVID. 984 variants in 886 genes passed our stringent filtering criteria, including 63 genes with rare segregating, conserved and functional variants in two or more families. A frameshift mutation in ABCA7 and a missense variant in ZCWPW1 are present in one of the 23 GWAS-confirmed AD candidate genes. Seven variants are in AD KEGG Pathway genes (BID, CYC1, ITPR1, ITPR2, LRP1, ATP2A1), including two variants in LRP1, a gene involved in AD through its roles in cholesterol transport and β-amyloid modulation. Follow up in ADGC exome chip association results comparing EOAD vs. late-onset AD identified 13 of our filtered genes with suggestive associations (P<10-3), including ITM2C (P=1.22×10-4), a gene known to inhibit the processing of APP by blocking access to alpha- and beta-secretase. Enrichment analysis of the list of rare conserved, functional variants showed significant, Benjamini FDR-adjusted enrichment for several AD-related processes including the 'ECM-receptor interaction' and 'ABC transporters' KEGG pathways; GO terms including 'homophilic cell adhesion' and 'microtubule-based movement'; and multiple INTERPRO 'cadherin' classes. Exome sequencing of EOAD pedigrees identified multiple rare segregating variants with potential roles in AD pathogenesis, several of which were shared in two or more families.

299

Integrated whole-transcriptome and DNA methylation analysis identifies new gene network in Alzheimer disease. *C.E. Humphries[1], M.A. Kohli[1], P.W. Whitehead[1,2], G. Beecham[1,2], D.C. Mash[3], M.A. Pericak-Vance[1,2], J. Gilbert[1,2].* 1) Hussman Inst Human Genomics, Univ Miami, Miami, FL; 2) Dr. John T. Macdonald Department of Human Genetics, Univ Miami, Miami, FL; 3) Department of Neurology, Univ Miami, Miami, FL.

To understand the processes contributing to Late-onset Alzheimer Disease (LOAD), we investigated differences in transcription and DNA methylation in human post-mortem brain tissue. Transcription was examined using total RNA-seq which permits the identification and quantitation of both known genes and non-coding RNAs. DNA methylation was examined using Illumina's Infinium HumanMethylation450 BeadChip Kit. Neuropathological specimens were sampled from age, sex and race-matched temporal poles (BA 38) from ten cases each of LOAD, Diffuse Lewy-Body (DLB) disease and pathologically and clinically normal controls. RNA was extracted using Qiagen's miRNeasy kit and the library was prepared with Epicentre's ScriptSeq protocol. Samples were run at a density of two per lane on Illumina's HiSeq2000, generating 40-65 million reads per library. The program DESeq was used to examine transcriptional differences between Gencode genes, a union of genomic sequences encoding a coherent set of potentially overlapping functional products. Transcriptional analysis between LOAD and Controls revealed 2,700 genes out of a total of 37,000 genes (Gencode) to be differentially expressed ($p<0.05$). Of these 2700 genes, 60% were protein coding and 40% were non-coding RNAs. Seventeen of these genes survived correction for multiple testing (FDR) of which 6 were non-coding RNAs. Next, the 2,700 genes that were differentially expressed in LOAD were compared between LOAD and DLB. Three hundred of these genes were differentially expressed ($p<0.05$) with 140 being protein coding. StringDB, a network program, found 49 of the protein coding genes to be functionally connected in a network. The hub of the network was VCAM-1, a gene known to be upregulated in LOAD. This hub had 3 branches consisting of genes involved in: 1) angiogenesis, 2) immune responses, and 3) axonal growth and cell adhesion. DNA methylation within these 49-networked genes was examined for differences between LOAD and CON+DLB brains. 35 of the 49 genes had altered methylation in LOAD. Of the 35 genes, changes of DNA methylation in the promoter region were negatively correlated(r-0.76) with transcription in 24 genes. The study of this network of genes with altered expression and methylation specific to LOAD may offer a fruitful approach to advancing our understanding of the etiology of late-onset Alzheimer's disease and the role that methylation changes may play in LOAD gene transcription.

300

Novel mutations uncovered from exome sequencing of Norwegian families with Parkinson's disease. *M. Lin[1], J. Aasly[2], D. Evans[1], C. Vilarino-Guell[1], B. Shah[1], C. Szu Tu[1], H. Han[1], H. Sherman[1], C. Thompson[1], M. Toft[3], K. Wirdefeldt[4], A.C. Belin[5], M.S. Petersen[6], J. Trinh[1], V. Silva[1], F. Pishotta[1], M. Farrer[1], GEO-PD Consortium.* 1) Centre for Applied Neurogenetics, University of British Columbia, Vancouver, BC, Canada; 2) Department of Neuroscience, Norwegian University of Science and Technology, Trondheim, Norway; 3) Department of Neurology, Oslo University Hospital, Oslo, Norway; 4) Department of Medical Epidemiology and Biostatistics and Department of Clinical Neuroscience, Karolinska Instiutet, Stockholm, Sweden; 5) Department of Neuroscience, Karolinska Instiutet, Stockholm, Sweden; 6) Department of Occupational Medicine and Public Health, The Faroese Hospital System, Torshavn, Faroe Islands.

Approximately 14% of patients with Parkinson's disease (PD) have a family history of parkinsonism. In larger, multi-incident families, pathogenic mutations/genes have now been identified using traditional linkage and contemporary exome sequencing methods. We report a clinical and comparative genetic study of Norwegian families. Six Norwegian families (pedigree structure of 3-5 generations, 111 individuals, of whom 28 have parkinsonism) were invited; 25 affected (mean age 76.3 ± 9.9 SD, mean age of onset 61.1 ± 7.0 SD, range = 45-75 years) and 16 (mean age 70.8 ± 16.1 SD) unaffected members participated in the study. Late-onset, asymmetric, levodopa-responsive PD appears to segregate in an autosomal dominant fashion. Two to three affected members (cousins) from each pedigree were screened negative for known genetic mutations using a proprietary diagnostic panel, then further exome-sequenced on an Illumina HiSeq platform. Genome alignment, annotation and pair-wise bioinformatic comparisons of affected family members were performed. Coding variants observed at <1% frequency were validated by Sanger sequencing in all family members and matched control subjects. Mutations segregating with PD but not observed in controls were genotyped in 3112 subjects (1613 patients) of Scandinavian origin. Subsequently, mutations were genotyped in multi-ethnic GEO-PD samples using Sequenom and TaqMan technologies. Additional exonic sequencing was performed in candidate genes. Our preliminary data showed one mutation in each family, for four families, which segregates with PD in four genes (*NOVA2*, *PABPC1L*, *RPE65* and *OR56B4*). Each mutation was highly conserved through evolution, and those substitutions were predicted to have deleterious consequences on protein function. Mutations in *PABPC1L*, *RPE65* and *OR56B4* were observed in additional unrelated patients but not control subjects. The functions of the encoded proteins are not well characterized but are involved in RNA splicing in developing neurons, translation, retinal regeneration and olfaction. In conclusion, pair-wise exome-sequencing is an efficient method to identify novel gene mutations in familial PD that enhances traditional linkage efforts in disease gene mapping.

301

Gene-environment interaction reveals hidden heritability: plasma vitamin D concentration and its interaction with vitamin D receptor gene polymorphisms in Parkinson disease. *L. Wang[1,2], M.L. Evatt[3], L. Maldonado[1], W.R. Perry[1], J.C. Ritchie[4], G.W. Beecham[1,2], E.R. Martin[1,2], J.L. Haines[5], M.A. Pericak-Vance[1,2], J.M. Vance[1,2], W.K. Scott[1,2].* 1) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL; 3) Department of Neurology, Emory University School of Medicine, Atlanta, GA; 4) Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA; 5) Center for Human Genetics Research, Vanderbilt University School of Medicine, Nashville, TN.

Parkinson disease (PD) is a complex disease with both genetic and environmental risk factors. Low plasma vitamin D concentrations (vit D) have been associated with increased risk of PD in several studies. Single nucleotide polymorphisms (SNPs) in the vitamin D receptor (VDR) gene, which encodes the major mediator of vit D's biological actions, are also associated with PD in multiple candidate gene studies. However, similar associations were not observed in other candidate gene and genome-wide association studies (GWAS). We hypothesize that these genetic effects are modified by vit D, and that joint analysis of genotype and vit D concentration may reconcile the disagreement across studies. Thus we jointly analyzed the association of vit D and VDR SNPs with PD in a GWAS sample of 484 PD cases and 409 controls (Edwards et al., 2010). LC-MS/MS was used to measure vit D concentration in stored plasma samples. Association between vit D and PD was evaluated by logistic regression. A 4 degree of freedom joint test of SNP genotype and interaction effects was conducted as suggested by Kraft et al. (2007). For each SNP, a full model containing the SNP (coded as genotypes), vit D deficiency (<20 ng/mL vs >=20 ng/mL), interaction terms, and three covariates (age, sex and sampling season) was compared to a restricted model containing only vit D deficiency and the covariates. Vit D deficiency was strongly associated with PD (Odds Ratio (OR)=2.7, 95% Confidence Interval (CI): (1.9, 3.7), P<0.0001). Joint tests were significant (likelihood ratio test P<0.05) for four SNPs: rs12721364, rs886441, rs2189480, and rs11574026. Analysis stratified by vit D deficiency found that the effect of VDR SNP genotypes on PD depended on vit D concentration. A significant dose-effect of the rare allele at rs886441 on PD is observed in the vit D non-deficient stratum (OR_{het}=1.6, CI=(1.1, 2.3); OR_{hom}=3.7, CI=(1.6, 8.8)) but not in the vit D deficient stratum (OR_{het}=1.1, CI=(0.6, 2.0); OR_{hom}=1.0, CI=(0.3, 4.2)). Similar patterns were observed at the other three SNPs. Our data not only strongly support vit D deficiency as a risk factor for PD but also, for the first time, demonstrate that the effect of VDR polymorphisms on PD depends on vit D concentration. This observation provides a potential explanation of the inconsistency across studies of VDR SNPs in PD where vit D concentration was not considered.

302

Genetic variants in longevity gene *KLOTHO* are associated with increased brain volumes in aging. *J.S. Yokoyama[1], V.E. Sturm[1], L.W. Bonham[1], E. Klein[2], K. Arfanakis[3], L. Yu[3], G. Coppola[2], J.H. Kramer[1], D.A. Bennett[3], L. Mucke[1,4], B.L. Miller[1], D.B. Dubal[1].* 1) Dept of Neurology, University of California San Francisco, San Francisco, CA; 2) Dept of Neurology and Semel Institute for Neuroscience and Human Behavior, The David Geffen School of Medicine at University of California Los Angeles, Los Angeles, CA; 3) Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL; 4) Gladstone Institute of Neurological Disease, San Francisco, CA.

Identification of genetic variants associated with human brain structures in aging may elucidate new biologic mechanisms underlying resilience to age-dependent cognitive decline and disease - and identify brain regions critical to healthy aging. Two variants in the gene *KLOTHO* (KL), rs9536314 (F352V) and rs9527025 (C370S) segregate together and form a haplotype (KL-VS) associated with longevity and protection from cardiovascular disease in heterozygous carriers. Since brain size decreases with age, we sought to determine whether carrying one copy of the protective KL-VS allele is associated with larger brain volumes in healthy aging individuals. Using voxel-based morphometry, we blindly and broadly analyzed brain regions by magnetic resonance imaging (MRI) in a cohort of 217 healthy older adults (mean±SE 70.2±0.5 years). Using linear regression models adjusted for total intracranial volume, age, sex, education, and APOE ε4 status, we found the KL-VS haplotype to be associated with increased volumes of frontal cortical brain regions. After adjusting for multiple testing, one of the strongest associations was a greater grey matter volume of the right dorsolateral prefrontal cortex (DLPFC, MaxT=4.55, P_{FWE}=0.03), a finding that replicated in an independent cohort of 224 healthy older adults (age: 81.2±0.5 years, beta=0.49±0.24, P=0.04). Because the right DLPFC is a critical component of neural networks engaged in executive cognitive abilities, we analyzed working memory and processing speed in individuals of both cohorts. By meta-analysis, greater executive function was associated with carrying one copy of the KL-VS haplotype (beta=0.39±0.11, P=0.0003), and the volume of the right DLPFC correlated with increased executive function in both cohorts (r=0.11, P=0.03). These results identify genetic variation in *KLOTHO* as a potential determinant of DLPFC volume, confirm the involvement of the right DLPFC in executive function, and implicate klotho in mechanisms of healthy cognitive aging.

303

ENIGMA2: Genome-wide scans of subcortical brain volumes in 16,125 subjects from 28 cohorts. *S. Medland, Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) Consortium.* Quantitative Genetics, Queensland Inst Med Res, Brisbane, Australia.

The ENIGMA consortium (http://enigma.loni.ucla.edu) was founded in 2010 and brings together researchers in imaging genomics, to understand brain structure and function, based on MRI, DTI, fMRI and genome-wide association scan (GWAS) data. The consortium model is particularly important in neuroimaging, where the cost of phenotype collection almost three times the cost of genome-wide genotyping and contrary to the expectations of many, the distribution of effect sizes in neuroimaging genetics closely follows that seen in other complex traits such as height. The ENIGMA Consortium is comprised of 28 groups that span 5 continents, including 16,125 subjects. ENIGMA follows a meta-analysis framework, where analyses are conducted at local sites and group-level, de-identified statistics are contributed for meta-analysis. Image analysis is conducted using fully-automated and validated neuroimaging segmentation algorithms (FSL FIRST, FreeSurfer, or validated DTI processing pipelines). In the initial flagship project (Stein et al, 2012) we undertook GWAS meta-analysis of hippocampal volume, intracranial volume, and total brain volume. Here we report the findings of the second ENIGMA meta-analyses, which extended the initial project to examine genetic influences on the volumes of subcortical structures using data imputed to the 1,000 Genome Project references. In this second round of analysis, we have already identified numerous novel genome wide significant associations including two regions influencing amygdala volume (8p23.1 and 9q22.1), two regions influencing the putamen (14q22.3 and 18q21.2), and one region influencing the thalamus (20p12.1). In addition, we have replicated our earlier findings for hippocampal volume (located at 12q24.22 and 12q14.3). Of particular note is the chromosome 18 region influencing putamen located in DCC which is involved in axonal guidance of dopaminergic neurons and has previously been associated with schizophrenia. These findings have important implications for the identification and use of neuro-imaging endophenotypes for psychiatric and neurological disorders. The full list of contributing authors is available at http://enigma.loni.ucla.edu/ongoing/gwasma-of-subcortical-structures/ASHG2013/.

304

Epigenome-wide association studies in the era of meta-epigenomics. *J.M. Greally[1], N.A. Wijetunga[1], F. Delahaye[2], Y.M. Zhao[2], A. Golden[1], J.C. Mar[3], F.H. Einstein[2].* 1) Department of Genetics, Albert Einstein College of Medicine, Bronx, NY; 2) Department of Obstetrics & Gynecology and Women's Health, Albert Einstein College of Medicine, Bronx, NY; 3) Department of Computational and Systems Biology, Albert Einstein College of Medicine, Bronx, NY.

The epigenome-wide association study (EWAS) is increasingly employed to assess the possible role of epigenetic dysregulation in the determination of phenotypes and disease. As part of ongoing EWAS, we explored the extent of variability of the epigenome between normal individuals. We found substantial variation of DNA methylation to occur in human CD34+ hematopoietic stem and progenitor cells (HSPCs) from different neonates without apparent disease, born at term with appropriate weight for gestational age. Empirical annotation of the genome of this cell type using *Roadmap in Epigenomics* reference data was performed, using Segway and Self-Organising Map (SOM) analyses. Promoter, enhancer, transcribed and repressed regions were identified and annotated, revealing among other findings that so-called CpG island shores encode poised enhancers in CD34+ HSPCs.

We then tested where DNA methylation variability was occurring relative to these empirically-derived annotations. We found the epigenetic variability to be targeted to functional sequences, candidate promoters and enhancer sequences, with a discrimination of enhancer variability proximal to genes expressed at lower levels. The enrichment of variability at loci with intermediate DNA methylation values, occurring at apparently “poised” enhancers (encoded by H3K4me1 and H3K27me3), suggests cell subpopulation heterogeneity between individuals as the basis for the epigenomic variability observed.

We find that even in purified cell types the population of cells has a mixture of epigenomes present, which could be described as a **meta-epigenome**. We developed new meta-epigenomic approaches to study the CD34+ HSPCs. The variability of DNA methylation at candidate enhancers was low for housekeeping genes but high for genes associated with leukocyte differentiation. We also adapted an approach used for transcriptional studies to infer the number of cell subpopulations in the CD34+ HSPC pool.

Recognition of the meta-epigenomic structure of cell pools studied by EWAS approaches establishes a foundation for more informed design and interpretation of EWAS, more rational estimates of statistical power and potentially new analytical approaches that take into account the unusual variability structure of the data, as we will describe.

305

Genome-wide DNA methylation analysis of uniparental disomy cases reveals many novel imprinted loci in the human genome. *R.S. Joshi, P. Garg, C. Borel, F. Cheung, A. Guilmatre, A.J. Sharp.* Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY.

Genomic imprinting is a mechanism in which the expression of genes varies depending on whether they are maternally or paternally inherited. Imprinting occurs through an epigenetic mechanism involving differential DNA methylation and histone modifications on the two parental alleles, with most imprinted genes marked by CpG-rich differentially methylated regions (DMRs). Approximately 70 human imprinted genes have been identified to date, and imprinted loci have been associated both with rare syndromes, and with common diseases including diabetes and cancer. Previously we showed that DNA methylation profiling in cases of uniparental disomy (UPD) provides a unique system that allows the isolated study of DNA derived from a single parent (PMID: 20631049). We have assembled an unprecedented collection of DNA from 113 patients with UPD for 18 different chromosomes, allowing the efficient detection of DMRs associated with imprinted genes for 84% of the human genome. We performed DNA methylation profiling in our UPD cohort using Illumina Infinium 450K Methylation BeadArrays that yield quantitative data on ~482,000 CpGs at single-nucleotide resolution. Imprinted DMRs were defined by sites at which the maternal and paternal methylation levels for multiple probes diverged significantly from the biparental average. Using stringent thresholds, we were able to detect 19 out of 20 (95%) previously described DMRs, validating our methodology for the detection of imprinted loci. Using these same criteria we identified a total of 46 DMRs spanning all 16 chromosomes analysed, 27 of which were not previously described. This included novel DMRs on chromosomes 3, 5, 21 and 22 which have previously been considered to be devoid of imprinting, highlighting potential parent-of-origin effects in chromosomal aneuploidies such as Down syndrome. The quantitative nature of our data also allowed us to detect subtle but consistent reciprocal methylation differences between the two parental alleles that encompass hundreds of CpG sites extending across large (>1Mb) chromosomal regions at known imprinted regions such as 15q11-q13. We confirmed novel DMRs by bisulfite sequencing of informative trios to determine parent-of-origin methylation marks. Our data provide the first comprehensive genome-wide map of imprinted sites in the human genome, and provide novel insights into potential parent of origin effects in a variety of human disorders.

306

Correlation Between CpG DNA Methylation Levels in peripheral CD4+ T cells and brain in aging individuals. *C.M. McCabe[1,2,3], L.L. Rosenkrantz[1,2,3], G. Srivastava[1,2,3,4], A. Kaliszewska[1,2,3,4], S. Imboywa[1,2,3], J. Schneider[5], D.A. Bennett[5], P.L. De Jager[1,2,3,4].* 1) Program in Medical and Population Genetics, The Broad Institute, 7 Cambridge Center, Cambridge, MA; 2) Center for Neurologic Disease, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 3) Program in Translational Neuro-Pshychiatric Genomics, Institute for the Neurosciences, Departments of Neurology and Psychiatry, Brigham and Women's Hospital, 77 Avenue Louis Pasteur, NRB168, Boston MA; 4) Harvard Medical School, Boston, MA; 5) Rush Alzheimer's Disease Center, Rush University Medical Center, 600 S Paulina St., Chicago IL.

Background: DNA methylation has been reported to be strongly affected by aging. Here, we explore the extent of DNA methylation changes found in subjects from two prospective cohort studies of aging: the religious order study (ROS) and the memory and aging project (MAP). We have previously characterized a profile of DNA methylation from the dorsolateral prefrontal cortex of 748 subjects from these studies to identify loci whose level of methylation in cortex is altered relative to aging and Alzheimer’s disease. Here, we have evaluated the utility of profiling a peripheral blood cell population, CD4+ T cells, in the same subjects for studies of cognitive aging and dementia. **Materials:** Forty-eight deceased subjects with archived frozen PBMC were sampled at two time points, one obtained at study enrollment, up to 12 years prior to death (t1) and one within a year of death (t2), and 216 subjects were sampled at one time point (t2). CD4+ T cells were purified from frozen PBMCs using a Miltenyi Biotec autoMACS machine. The Qiagen All-Prep DNA kit was used for DNA extraction. Epigenomic data was obtained on 470,913 CpGs using the Illumina HumanMethylation450 beadset. **Results:** Profiles from the 216 subjects sampled only once are being generated currently and will be analyzed in relation to the subjects’ trajectory of decline. In looking at the subset of subjects with two time points, there is a large excess of positive correlation highlighting the existence of stable methylation signals in an individual’s CD4+ T cells over several years’ time. Further, when comparing brain methylation to methylation in one of the CD4+ time points, the mean of the distribution of total correlations is near zero, but we see a skewed heavy tail to these distributions, i.e., there is an excess of CpGs with high levels of positive correlation in methylation levels between blood and brain. The proportion of CD4+ T cell CpGs methylation is more highly correlated (r>0.5) with brain methylation in the blood samples proximal to death (t2) (11.5% of the total CpGs) than in those obtained at enrollment into the study (t1) (4.3% of the total). **Conclusion:** These initial results demonstrate that for a subset of CpGs, methylation levels in CD4+ T cells are correlated with brain methylation levels. This intriguing result requires validation and suggests that, in this subset of loci, further investigation into the role of CD4+ T cell methylation as a surrogate for brain is warranted.

307

Genome-wide analysis of Mecp2 dependent DNA methylation and hydroxymethylation at base-resolution in neurons. *KE. Szulwach[1], M. Yu[2], X. Li[1], CR. Street[1], C. He[2], P. Jin[1].* 1) Dept Human Gen, Emory Univ, Atlanta, GA; 2) Dept of Chem and Inst for Biophysical Dynamics, The University of Chicago, Chicago, IL.

5-methylcytosine (5mC) imparts epigenetic function to genomic elements and has been implicated in an array of human neurological disorders. In neurons, 5mC exhibits stimulus dependent dynamics and is enzymatically oxidized to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) by TET-family proteins. When coupled with base-excision repair, TET-mediated oxidation of 5mC can result in active DNA demethylation. We have established whole-genome DNA methylation states at base-resolution using complementary methods to distinguish 5mC and 5hmC (WGBS and TAB-Seq, respectively) in mature neurons isolated from adult mouse hippocampal dentate gyrus as well as from a mouse model of the human neurological disease, Rett Syndrome (RTT). RTT is primarily caused by mutations in MECP2, which encodes a 5mCpG specific binding protein. In brain, Mecp2 expression is inversely correlated with 5hmC levels, and Mecp2 has affinity for 5hmC, implicating it in the dynamic regulation of 5mC. Our base-resolution analysis of 5mC and 5hmC indicate that, although Mecp2 is extremely abundant in mature neurons, DNA methylation states are largely preserved in Mecp2-/y mice (Pearson >0.99). Yet, stringent identification of differentially methylated regions (DMRs) consistent among biological replicates identified 5,126 DMRs within a small fraction of the genome (~2.5MB in total). These DMRs exhibited three key features; 1.) DMRs were distal to promoters 2.) DMRs were initially depleted of 5mC relative to the genome-wide average in wildtype neurons 3.) DMRs were preferentially hypomethylated in Mecp2-/y neurons. These features indicate Mecp2 dependent regulation of DNA methylation at sites likely undergoing dynamic regulation under normal conditions, which is supported by the enrichment of 5hmC at DMRs. Lastly, DMRs that become hypomethylated in Mecp2-/y neurons are enriched for a distinct set of transcription factor binding motifs ($P \leq 1e\text{-}6$), including Creb1 and CTCF, among others. These DMRs, therefore, represent sites of Mecp2 dependent dynamic DNA methylation, providing links to the underlying gene regulatory networks dysregulated in Mecp2-/y mice. Overall, our data demonstrate a novel approach whereby genome-wide analyses of 5mC and 5hmC can be utilized to define DMRs reflective of the underlying alterations in the transcriptional regulatory circuitry, supporting a role for Mecp2 in localized DNA methylation dynamics at key gene regulatory regions.

308

Random Replication of the Inactive X Chromosome. *A. Koren[1,2], S.A. McCarroll[1,2].* 1) Genetics, Harvard Medical School, Boston, MA; 2) Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA.

In eukaryotic cells, genomic DNA replicates in a defined temporal order. The inactive X chromosome (Xi), the most extensive instance of facultative heterochromatin in mammals, replicates later than the active X chromosome (Xa), but the replication dynamics of inactive chromatin are not known. By profiling human DNA replication in an allele-specific, chromosomally phased manner, we determined for the first time the replication timing along the active and inactive chromosomes (Xa and Xi) separately. Replication of the Xi was different than that of the Xa, varied among individuals, varied across experimental replicates, and resembled a random, unstructured process. The Xi replicated rapidly and at a time largely separable from that of the euchromatic genome. Late-replicating, transcriptionally inactive regions on the autosomes also replicated in an unstructured manner, similar to the Xi. We conclude that DNA replication follows two strategies: slow, ordered replication associated with transcriptional activity, and rapid, random replication of silent chromatin. The two strategies co-exist in the same cell yet are segregated in space and time.

309

The epigenetic profile of the SOX9 regulatory region appears Y chromosome dependent. *G. Houge, H. Lybæk.* Center for medical genetics, Bergen, Norway.

Recently it was shown that duplications of the RevSex element, located 0.5 Mb upstream of SOX9, cause XX-DSD and deletions XY-DSD (DSD; disorder of sex development). To explore how a 148 kb RevSex duplication could have turned on gonadal SOX9 expression in the absence of SRY in an XX-male, we examined the chromatin landscape of the SOX9 regulatory region by custom ChIP-on-chip experiments in fibroblast lysates from the index, his RevSex duplication carrier father and six controls. In addition, regional CpG methylation was examined. The RevSex duplication was associated with chromatin changes that could have facilitated the establishment of an auto-regulatory SOX9 self-maintaining loop through better accessibility of the TESCO enhancer 14-15 kb upstream of SOX9. Here a dip in the H3K9me3 signal was found together with a strong H3K4me3 profile towards the SOX9 transcription start site. A similar type of effect was found when three control males were compared to three control females. Like females, the XX-DSD case had an enhanced H3K27me3 profile in the100 kb region upstream of SOX9. The RevSex duplication carrier father had the most open chromatin profile of all. Our findings suggest that RevSex copy number-induced sex reversal could be due to long-range changes in chromatin conformation. Furthermore, the differences in chromatin state maps between males and females suggest that the Y chromosome may induce large-scale conformational changes of genomic regions. Our findings are preliminary, but give reason to further explore whether SRY regulates gene expression through regional epigenetic modification.

310

Global reduction of 5-hydroxymethylcytosine in a FMR1 premutation mice model. *B. Yao[1], L. Lin[1], C. Street[1], Z. Zalewski[2], J. Galloway[2], D. Nelson[2], P. Jin[1].* 1) Human Genetics, Emory University, Atlanta, GA; 2) Human and Molecular Genetics, Baylor College of Medicine, Houston, TX.

Fragile X-associated tremor/ataxia syndrome (FXTAS) is a late onset neurodegenerative disorder, with patients carrying permutation alleles of 55-200 CGG repeats on the FMR1 gene. The cerebella of FXTAS patients usually develop reduced and ectopic Purkinje cells. It has been proposed that the CGG expansion disrupts the balance and availability of RNA-binding proteins, therefore affecting their cellular functions. How this alteration of transcriptional states and epigenetic regulation modulates FXTAS has yet to be explored. DNA 5-hydroxymethylcytosine (5hmC) converted from 5-methylcytosine (5mC) by ten-eleven translocation (TET proteins), has been recently reported. This novel epigenetic mark provides a completely new perspective on the plasticity of 5mC-dependent epigenetic regulation. Here we applied genome-wide profiling of cerebellar 5hmC in a FXTAS mice model (CGG mice) with their Purkinje neurons ectopically expressing 90 CGG repeats. 16 weeks CGG mice showed overall reduced 5hmC levels genome-wide compared to their wildtype littermates. The reduced 5hmC can be readily detected on gene bodies and CpG islands. However, gain-of-5hmC regions have also been observed in repetitive elements such as SINE, LTR and simple repeats. Importantly, in cerebellum, tissue-specific enhancers but not general enhancers showed higher 5hmC levels in CGG mice, implying that 5hmC regulates cerebellum-specific gene expression via manipulating enhancer cytosine modification. Separately mapping wild-type and CGG specific DhMRs (differential 5-hydroxymethylated regions) revealed their highly correlation with neuronal developmental gene ontology terms and functional pathways. Genomic annotation separately identified 6026 wildtype- and 2969 CGG-specific 5hmC-associated genes. Among those, 148 wildtype- or 78 CGG- DhMR associated genes are overlapping with 498 genes that are differentially associated with ribosomes in CGG mice identified by bacTRAP ribosomal profiling when they are in 4 weeks of age. Taken together, our data strongly indicates the functional importance of 5hmC in the etiology of FXTAS, possibly through the regulation of transcription, and shed light on the potential therapeutic interventions.

311

Coordination of engineered factors with TET1/2 promotes early stage epigenetic modification during somatic cell reprogramming. *Y. Li[1], G. Zhu[2], F. Zhu[2], T. Wang[1], W. Jin[2], W. Mu[2], W. Lin[2], W. Tan[2], W. Li[2], Y. Feng[3], S. Warren[1], Q. Sun[2], D. Chen[2], P. Jin[1], State Key Laboratory of Reproductive Biology and State Key Laboratory of Biomembrane and Membrane.* 1) Dept Human Gen, Emory Univ, Atlanta, GA; 2) State Key Laboratory of Reproductive Biology State Key Laboratory of Biomembrane and Membrane Biotechnology, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang, Beijing, P.R. China 100101; 3) Dept Pharmacology, Emory Univer, Atlanta, GA.

Somatic cell reprogramming towards induced pluripotent stem cells (iPSCs) holds great promise in future regenerative medicine, however, the reprogramming process mediated by the traditional defined factors (OSNK) is slow and extremely inefficient. Here we develop a combination of modified reprogramming factors (OySyNyK), in which the transactivation domain of the Yes-associated protein is fused to defined factors, and establish a highly efficient and rapid reprogramming system. We show that the efficiency of OySyNyK-induced iPSCs was up to 100-fold higher than the OSNK, and the reprogramming by OySyNyK is very rapid and is initiated around 24 h. Notably, we found that OySyNyK-factors significantly increased the TET1 expression at the early stage, which interact with defined factors, and co-occupy the pluripotency loci. Our studies not only establish a rapid and highly efficient iPSC reprogramming system, but also uncover a novel mechanism by which engineered factors coordinate with TETs to regulate 5hmC-mediated epigenetic control.

312

SMCHD1 mutations in facioscapulohumeral muscular dystrophy type 2. *R.J.L.F. Lemmers[1], M.P. Nieuwenhuizen[2], P.J. van der Vliet[1], M. Vos-Versteeg[2], J. Balog[1], J.J. Goeman[3], D.G. Miller[4], S.J. Tapscott[5], S. Saconni[6], R. Tawil[7], B. Bakker[2], S.M. van der Maarel[1].* 1) Human Genetics, Leiden University Medical Center, Leiden, Netherlands; 2) Clinical Genetics, Leiden University Medical Center, Leiden, Netherlands; 3) Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, Netherlands; 4) Department of Pediatrics, University of Washington, Seattle, Washington, USA. Seattle Children's Hospital, Seattle, WA, USA; 5) Divisions of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; 6) Neuromuscular Diseases Specialized Center, Nice University Hospital, France; 7) University of Rochester Medical Center, Department of Neurology, Rochester, NY, USA.

Facioscapulohumeral muscular dystrophy (FSHD) is one of the commonest muscular dystrophies. FSHD is characterized by a high clinical variability in onset, progression and severity. This variability is observed both between families and within families. FSHD is associated with an opening of the chromatin structure in the D4Z4 macrosatellite repeat array localized on chromosome 4 and transcriptional derepression of the D4Z4-encoded DUX4 gene in skeletal muscle. Two variants of chromosome 4 have been described of which only one is permissive to DUX4 expression because of the presence of a polymorphic DUX4 polyadenylation site. In most patients, chromatin relaxation-associated DUX4 expression is caused by contraction of the D4Z4 repeat array (autosomal dominant FSHD1) on a permissive allele. However, in some patients D4Z4 chromatin decondensation occurs independent of D4Z4 repeat array size and is caused by mutations in the chromatin modifier structural maintenance of chromosomes flexible hinge domain containing 1 (SMCHD1) on chromosome 18. The chromatin modifier SMCHD1 binds directly to the D4Z4 repeat where it is involved in establishment and/or maintenance of CpG methylation. In FSHD2 patients there is reduced binding of SMCHD1 to D4Z4, resulting in D4Z4 CpG hypomethylation and causing disease when combined with a permissive allele. In support, knock down of SMCHD1 in normal myoblasts containing a permissive haplotype leads to DUX4 expression. We performed a SMCHD1 mutation screen in 62 independent FSHD2 patients with the objectives to define the mutation spectrum of SMCHD1. In 51 families (82%) we identified SMCHD1 variations that were inferred to be disease causing. The mutation spectrum is unique and strongly suggests a selection bias that predicts a balanced interplay between SMCHD1 activity and D4Z4 repeat size in the somatic repression of DUX4. We anticipate that further genotype-phenotype analyses, combined with functional studies will facilitate our understanding of repeat mediated epigenetic silencing mechanisms in mammalian cells.

313

A Genome-wide meta-analysis of the response to inhaled bronchodilators among subjects with chronic obstructive pulmonary disease. *M. Hardin[1,2], M.H. Cho[1,2], M. McDonald[1], E. Wan[1,2], D.A. Lomas[3], H.O. Coxson[4], L.D. Edwards[5], W. MacNee[6], J. Vestbo[7], J.C. Yates[5], A. Agusti[8], P. Calverley[9], B. Celli[2], C. Crim[5], S. Rennard[10], E. Wouters[11], P. Bakke[12], E.A. Regan[13], B. Make[13], A. Litonjua[1,2], J.E. Hokanson[14], J.D. Crapo[13], T.H. Beaty[15], E.K. Silberman[1,2], C.P. Hersh[1,2], the ECLIPSE and COPDGene Investigators.* 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA; 2) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA; 3) Department of Medicine, University COllege London, London, UK; 4) UBC Department of Radiology, Vancouver General Hospital, Vancouver, Canada; 5) GLaxoSmithKline, Research Triangle Park, NC, USA; 6) Department of Respiratory and Environmental Medcine, University of Edinburgh, Edinburgh, Scotland; 7) Department of Respiratory Medicine, Manchester Academic Health Sciences Centre, University Hospital of South Manchester, Manchester, UK; 8) Thoracic Institute, Hospital Clinic, Barcelona, SP; 9) Department of Pulmonaryr and Rehabilitation Medicine, University of Liverpool, Liverpool, UK; 10) Department of Medicine, Nebraska Medcial Center, Omaha, Nebraska, USA; 11) Center for Chronic Diseases, University Hospital Maastricht, Maastricht, The Netherlands; 12) Department of Clinical Science, University of Bergen, Bergen, Norway; 13) DIvision of Pulmonary and Critical Care Medicine, National Jewish Health, Denver, CO, USA; 14) Department of Epidemiology, Colorado School of PUblic Health, University of Colorado Denver, Denver, Colorado; 15) Johns Hopkins School of Public Health, Baltimore, MD, USA.

Rationale: Recent genome-wide association studies have identified genes associated with lung function, both in population-based and chronic obstructive pulmonary disease (COPD) case/control cohorts. The response to inhaled bronchodilators is a measure of lung function that correlates with FEV1 decline and survival in COPD and has been found to be heritable. Identifying single nucleotide polymorphisms (SNPs) associated with bronchodilator responsiveness (BDR) may reveal novel genetic pathways associated with the pathogenesis of COPD. Methods: We performed a genome-wide association study of BDR in four cohorts of Caucasian subjects with GOLD stage II or higher COPD. These previously described cohorts included ECLIPSE (1764 cases with > 10 pack-years smoking), COPDGene (2797 subjects with > 10 pack-years smoking), NETT (364 subjects > 5 pack-years smoking), and GenKOLs (864 cases with > 2.5 pack-years smoking). We performed an additional analysis in over 700 African Americans from the COPDGene study. Pre-and post-bronchodilator spirometry was performed on all subjects. BDR was calculated in response to administered albuterol as absolute change in FEV1 (BDRABS), change as percentage of predicted FEV1 (BDRPRED), and change as percentage of baseline FEV1 (BDRBASE). Genotyping was performed using Illumina SNP arrays. Additional genotypes were imputed using the 1000 Genomes reference panel. Linear regression was performed in all cohorts, for all three outcomes, adjusting for age and pack-years smoking. We then performed a meta-analysis of results from the four cohorts using a random-effects model to investigate the top SNPs from each outcome. SNPs were filtered for minor allele frequency > 0.01. Results Over 6.3 million unique SNPs from 5789 COPD subjects (GOLD stage II or higher) were investigated. In a meta-analysis, SNPs in HS6ST3 (P=1.7×10-8) were associated with BDRBLINE, SNPs in XKR4 were associated with BDRPRED (P=5.8×10-8) and BDRBLINE (P=1.3×10-7), and SNPs in the CUBN gene were associated with BDRABS (P=2.6×10-7) and BDRPRED (P=7.0×10-7). Among African American subjects, SNPs in CDH13 were significantly associated with BDRABS (P=1×10-11). Conclusions: Bronchodilator responsiveness in COPD is an important measure of lung function that is likely to have multiple genetic determinants. In a meta-analysis including four case-control cohorts of COPD, we identified several novel variants associated with bronchodilator responsiveness.

314

Genome-wide association study of opioid-induced vomiting in the 23andMe cohort. *J.L. Mountain[1], N. Eriksson[1], J.Y. Tung[1], A.S. Shmygelska[1], H.L. McLoed[2], U. Francke[1,3], A.K. Kiefer[1], D.A. Hinds[1].* 1) 23andMe, Inc, Mountain View, CA; 2) Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC; 3) Department of Genetics, Stanford University, Stanford, CA.

To date the vast majority of pharmacogenetic studies have been conducted in vitro or within a relatively small cohort. Here we present results of a large genome-wide association study of side effects of codeine. Codeine is an opiate medication used to control mild to moderate pain. In the body, codeine is activated to morphine which is then thought to bind to receptors in the brain and spinal cord that play a role in transmitting the sensation of pain throughout the body. Opioids also bind to receptors in the gastrointestinal tract. Among the more common side effects of codeine are vomiting, itching, euphoria, and constipation. Participants in this study were unrelated customers of 23andMe, a personal genetics company, with European ancestry. They consented to take part in research and reported their experiences with commonly used medications via a web-based survey. 2,479 cases reported vomiting in response to codeine use; 10,789 controls reported none of the four side effects named above. Participants were genotyped at up to 1 million positions; genotypes for 20 million positions were imputed. We performed logistic regression assuming an additive model for allelic effects, controlling for age, sex, and genetically-derived ancestry. We identified two genome-wide significant regions. The top hit, rs9620007 (p=1.4e-09; OR= 0.79), is located within the *WBP2NL* gene on chromosome 22. Located about 0.1 Mb from the *WBP2NL* gene is the *CYP2D6* gene. Previous studies have yielded strong evidence for a link between *CYP2D6* genotype and codeine metabolism. However, these studies have focused primarily on drug metabolism (e.g. excretion of morphine after codeine use) rather than on specific side effects. The second hit, rs10732842 (p=2.9e-08;OR=1.29), is located near the *LRRIQ3* gene on chromosome 1. We identified a third association, nearly significant at the genome-wide level (p=2.9e-07; OR= 0.77), between vomiting in response to codeine and a variant (rs1799971, g.A118G) in the opioid receptor gene *OPRM1*. Previously this SNP was found to be linked to susceptibility to opioid dependence. This study is novel in demonstrating associations between the side effect of vomiting in response to codeine and two genes linked to opioid pathways. The significance likely reflects several factors including the relatively high frequency of use of codeine, the relatively high frequency of this particular side effect, and the participants' ability to recall a more dramatic side effect.

315

Transcriptome Profiling of Human Airway Smooth Muscle Cells Stimulated with Dexamethasone Identifies *CRISPLD2* as a Regulator of Steroid and Immune Response. *B.E. Himes[1,2,3], X. Jiang[4], P. Wagner[4], R. Hu[4], B. Klanderman[2], Q. Duan[1], J. Lasky-Su[1], C. Nikolos[5], W. Jester[5], M. Johnson[5], R.A. Panettieri Jr.[5], K.G. Tantisira[1], S.T. Weiss[1,2], Q. Lu[4].* 1) Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2) Partners Center for Personalized Genetic Medicine, Boston, MA; 3) Children's Hospital Informatics Program, Boston, MA; 4) Program in Molecular and Integrative Physiological Sciences, Department of Environmental Health, Harvard School of Public Health, Boston, MA; 5) Pulmonary, Allergy and Critical Care Division, University of Pennsylvania, Philadelphia, PA.

Asthma is a chronic inflammatory airway disease that affects over 300 million people around the world. Glucocorticoids are common medications used to manage inflammatory diseases, and one of the primary tissues that glucocorticoids target in the treatment of asthma is the airway smooth muscle. RNA-Seq is a high-throughput sequencing method that provides comprehensive expression analysis, including discovery of novel genes, non-coding transcripts, and splice variants. We used RNA-Seq to characterize changes in the human airway smooth muscle (HASM) transcriptome in response to treatment with a glucocorticoid (i.e. dexamethasone, 1 μM for 18h) using primary HASM cells from four white male donors. The Illumina TruSeq method was used to prepare RNA-seq libraries that were sequenced with an Illumina HiSeq 2000 instrument and aligned to the hg19 reference genome using TopHat. Differential expression analysis, carried out using Cufflinks and CummeRbund, revealed that 316 genes were significantly differentially expressed between control and dex-treated cells. Such genes included ones that have been previously related to steroid responsiveness and inflammation (i.e., *DUSP1, KLF15, PER1,* and *TSC22D3* (a.k.a. *GILZ*), as well as novel candidates that are not known to be associated with glucocorticoid response. Previously conducted GWAS of two glucocorticoid- and HASM-related phenotypes, inhaled corticosteroid resistance and bronchodilator response, were used to screen differentially expressed genes. One of the top dex-induced genes, *CRISPLD2*, had SNPs with nominal associations (P-value $<10^{-03}$) with both of these asthma phenotypes and was selected for further functional studies. After verifying that dexamethasone significantly increases both *CRISPLD2* mRNA and protein levels in HASM cells, we performed knock-down experiments in which *CRISPLD2* was found to (1) reverse the effect of dexamethasone on two known steroid-response genes (i.e. *ENaC* and *TSC22D3*) and (2) increase the IL1β responsiveness of a inflammatory gene (i.e. *IL6*). Our results suggest that *CRISPLD2* regulates both steroid and NFκB-dependent immune responses and may play a role in asthma pharmacogenetics.

316

Potential of integrating human genetics and electronic medical records for drug discovery: the example of *TYK2* and rheumatoid arthritis. *D. Diogo[1,2,3], K.P. Liao[2], R.S. Fulton[4], R.R. Graham[5], J. Cui[2], J.C Denny[8], T. Behrens[5], M.F. Seldin[6], P.K. Gregersen[7], E. Mardis[4], R.M. Plenge[1,2,3], The RACI, i2b2-Rheumatoid Arthritis, CORRONA.* 1) Division of Rheumatology, Immunology & Allergy, Brigham and Women's, Hospital, Boston, Massachusetts; 2) Division of of Genetics, Brigham and Women's, Hospital, Harvard Medical School, Boston, Massachusetts; 3) Medical and Population Genetics Program, Broad Institute, Cambridge, Massachusetts; 4) The Genome Institute, Washington University School of Medicine, St. Louis, Missouri; 5) ITGR Human Genetics Group, Genentech Inc, San Francisco, California; 6) Department of Biochemistry and Molecular Medicine, University of California, Davis, California; 7) The Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, New York; 8) Department of Biomedical Informatics and department of Medicine, Vanderbilt University, Nashville, Tennessee.

Human genetics has the potential to uncover disease-associated alleles that result in gain-of-function (GOF) or loss-of-function (LOF), thereby linking target perturbation with relevant human physiology. In particular, genes with LOF alleles that protect from disease represent promising targets for pharmacological manipulation. Here, we integrated deep sequencing with large-scale genotyping to search for protein-coding variants that influence risk of rheumatoid arthritis (RA). We selected 845 genes, based on findings from genome-wide association studies (GWAS) in RA and other autoimmune diseases, as well as candidate genes based on the pathophysiology of RA, and targeted the protein-coding exons for sequencing in 1,118 RA cases and 1,118 matched controls of European ancestry. We observed an excess of rare alleles (MAF<1%) predicted to be damaging in controls in *TYK2*, a gene coding for a member of the Janus kinases (JAK) (P=0.003). The signal of association was driven by 2 variants with MAF>0.5%. In addition, we observed an excess of true rare variants in controls in the kinase 1 domain-coding region of *TYK2* (P=0.006). We further analyzed the role of low-frequency and common *TYK2* protein-coding variants in two large collections of case-control samples genotyped on the Exomechip or the Immunochip. We demonstrate that three protein-coding variants predicted to be damaging independently protect against RA ($P=1.6\times10^{-27}$). Importantly, recent studies have shown that two of these variants are LOF mutations that affect TYK2 kinase activity. To assess for pleiotropic effects of the three alleles - and especially phenotypes that may be considered adverse events following pharmacological inhibition of TYK2 - we linked *TYK2* genetic data with clinical data from electronic medical records (EMR) for 3,102 individuals (European ancestry). In a permutation-based analysis, we observed an enrichment of association at clinical diagnoses related to infection, suggesting that individuals carrying *TYK2* RA-protective haplotypes are at higher risk of infection. Together, our results highlight the role of *TYK2* in RA pathogenesis and suggest that pharmacological manipulation of TYK2 would be effective at treating the symptoms of RA with an adverse drug event profile that is predictable from human clinical data. These findings highlight the potential of integrating information from electronic medical records with human genetics for drug discovery in complex traits such as RA.

317

Rare variants contribute to bronchodilator drug response in Latino children with asthma. *D.G. Torgerson[1], K.A. Drake[1], C.R. Gignoux[1], J.M. Galanter[1], L.A. Roth[1], S. Huntsman[1], D. Hu[1], C. Eng[1], S.W. Yee[2], L. Lin[2], C.D. Campbell[3], E.E. Eichler[3,4], R.D. Hernandez[2], K.M. Giacomini[2], E.G. Burchard[1,2], the GALA II investigators.* 1) Department of Medicine, University of California San Francisco, San Francisco, CA; 2) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA; 3) Department of Genome Sciences, University of Washington, Seattle, WA; 4) Howard Hughes Medical Institute, Seattle, WA.

There is substantial variation in how well an individual responds to albuterol, the primary rescue medication used to treat asthma symptoms. Although these differences may be due to environmental factors, there is mounting evidence for a genetic contribution to variability in bronchodilator drug response (BDR). We performed a genome-wide association study (GWAS) and admixture mapping for BDR in 1,782 Latino children with asthma genotyped on the Axiom LAT1 array (World Array 4, Affymetrix), followed by exon sequencing of 3 genes in 787 Puerto Ricans. We identified six variants associated with BDR at a genome-wide significant threshold in our GWAS ($p<5\times10^{-8}$), all of which had frequencies below 5%. Carriers of these variants were validated using Sanger sequencing. Furthermore, we observed an excess of small p-values driven by variants at frequencies below 5%, and by variants near solute carrier genes (SLC genes), which include membrane transport proteins involved in the transport of endogenous compounds and xenobiotics. Admixture mapping identified five significant peaks, one of which contained population-specific associations with rare variants in *SLC22A15*: two variants were locus-wide associated with higher BDR in Mexicans (rs1281743 and rs1281748, $p = 8.8\times10^{-5}$), and one variant identified through sequencing was gene-wide associated with lower BDR in Puerto Ricans (rs146673261, $p=8.4\times10^{-4}$). RtPCR and immunohistochemistry confirmed that *SLC22A15* is highly expressed in lung epithelial cells. Exon sequencing of 3 SLC genes in 787 Puerto Ricans revealed significant associations with multiple rare variants and BDR in those identified through GWAS: *SLC24A2* (SKAT test, p=0.037), and *SLC24A4* (p=0.018). Overall our results suggest that rare variation contributes to individual differences in response to albuterol in Latinos, notably in solute carrier genes. Resequencing in larger, multi-ethnic population samples and additional functional studies are required to further understand the role of rare variation in BDR.

318

Characterization of Statin Dose-response within Electronic Medical Records. *W.Q. Wei[1], Q.P. Feng[2], L. Jiang[3], M.S. Waitara[2], O.F. Iwuchukwu[2], D.M. Roden[2,4,5,6], M. Jiang[7], H. Xu[7], R.M. Krauss[8], J.I. Rotter[9], D.A. Nickerson[10], R.L. Davis[11], R.L. Berg[12], P.L. Peissig[12], C.A. McCarty[13], R.A. Wilke[14], J.C. Denny[1].* 1) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 2) Department of Medicine, Vanderbilt University, Nashville, TN; 3) Department of Molecular Physiology and Biophysics, Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN; 4) Department of Pharmacology, Vanderbilt University School of Medicine, Nashville, TN; 5) Oates Institute for Experimental Therapeutics, Vanderbilt University School of Medicine, Nashville, TN; 6) Office of Personalized Medicine, Vanderbilt University School of Medicine, Nashville, TN; 7) Department of Biomedical Informatics, University of Texas, Houston, TX; 8) Children's Hospital Oakland Research Institute, Oakland, CA; 9) Medical Genetics Institute, Cedars-Sinai Medical Center, West Los Angeles, CA; 10) Department of Genome Sciences, University of Washington, Seattle, WA; 11) Kaiser Permanente Georgia, Center for Health Research Southeast, Atlanta, GA; 12) Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield, WI; 13) Essentia Institute of Rural Health, Duluth, MN; 14) Department of Internal Medicine, Sanford Healthcare, Fargo, ND.

Background: Statin-mediated reduction in cardiovascular events represents one of the greatest advances of modern medicine. However, there is wide variability in the degree to which these drugs reduce low density lipoprotein (LDL) cholesterol within an individual. Efforts to define the genetic architecture underlying this variability have met with limited success, in part because genetic studies examining statin response have been limited to a single dose. By facilitating the reconstruction of full dose-response curves, electronic medical records (EMRs) offer a potential solution. In this study, we extract dose-response curves for simvastatin and atorvastatin, the two most commonly prescribed drugs in this class. We then leveraged these phenotypes to identify genetic predictors of statin potency and lipid lowering efficacy in the context of routine care. Methods: Two EMR-linked biobanks were utilized to construct dose-response curves for 2,026 subjects exposed to multiple doses of simvastatin and 2,252 subjects exposed to multiple doses of atorvastatin. We then fitted these data to a non-linear mixed effects model and extracted parameters representing potency (ED50), and maximal lipid lowering efficacy (Emax). We tested these parameters for association with 144 gene variants pre-selected based on (A) prior association with baseline lipids, (B) prior association with statin response, or (C) proven impact on statin pharmacokinetics. Findings: Atorvastatin was more efficacious, more potent, and demonstrated less inter-individual variability than simvastatin. A pharmacodynamic variant emerging from randomized trials (PRDM16) is associated with maximal efficacy (Emax) for both simvastatin ($p = 0.04$) and atorvastatin ($p = 0.008$). The impact of this variant on effect size was striking for atorvastatin (Emax = 51.7 mg/dl in subjects homozygous for the minor allele versus Emax = 75.0 mg/dl in subjects homozygous for the major allele). We also identified several loci associated with atorvastatin ED50, including SORT1, and several loci that were associated with simvastatin ED50, including SLCO1B1. Conclusion: Biobanks linked to EMRs improve our understanding of genetic factors contributing to drug response. The extraction of rigorously defined traits for pharmacogenetic association studies represents another promising approach to the meaningful use of EMRs.

319

Genome-wide analysis of creatine kinase levels in statin-users. *M.P. Dubé[1-3], R. Zetler[1-3], Y. Feroz Zada[1,3], V. Normand[1,3], I. Mongrain[1,3], N. Laplante[1], A. Barhdadi[1,3], G. Asselin[1,3], S. Provost[1,3], J.D. Rioux[1,2], S. deDenus[1-3], E. Kritikou[1,2], J. Turgeon[2,4], M.S. Phillips[1-3], J.C. Tardif[1-3].* 1) Montreal Heart Institute, Montreal, QC, Canada; 2) Université de Montréal, Montreal, QC, Canada; 3) Beaulieu Saucier Pharmacogenomics Centre, Montreal, QC, Canada; 4) Centre de recherche du CHUM Hôtel-Dieu, Montreal, QC, Canada.

Statins (HMG-CoA reductase inhibitors) are the most prescribed class of lipid lowering drugs used in the treatment and prevention of cardiovascular disease. Despite their benefits, statins are associated with muscle-related adverse effects. Approximately 10% of statin-users report symptoms of myopathy including myalgia without abnormal creatine kinase (CK) elevation and myositis with CK elevation. In order to better understand the predisposing genetic factors to statin-induced myalgia, we conducted a case-control association study of 4600 statins-users, half present myalgia and half without. We found that patients with ongoing muscle pain presented with slightly higher plasma CK levels ($P=9.9\times10^{-5}$), despite the fact that the majority of CK values were within the "normal" clinical range. To follow on this observation, we did a genome-wide study of on-statin CK levels as secondary phenotype analysis. We included 3412 patients on statin at the time of recruitment into the case-control myalgia study. Recruitment proceeded from 9 clinical care centres throughout the province in Quebec, Canada. Genotyping was performed by using the Illumina Human610-Quad BeadChip and an iSelect panel enriched for genes of lipid homeostasis, statin pharmacokinetics and pharmacodynamics, and general drug metabolism. We found a strong association signal between plasma levels of CK and the muscle CK gene (CKM) ($P=5.0\times10^{-16}$) and with the LILRB5 gene ($P=2.6\times10^{-11}$) which is located 9Mb downstream of CKM. Genetic variants at those two genes were shown to be independently associated to CK levels in statin users. Results were successfully replicated in a cohort including 4000 statin users and 1500 non-users from the Montreal Heart Institute hospital cohort and genotyped with the ExomeChip (CKM: $P=7.6\times10^{-16}$; LILRB5: $P=3.5\times10^{-13}$). The genetic factors alone were shown to explain 1.5% and 2% of the variability in CK values in statin users and non-statin users respectively. A predictive model including genetic factors plus known clinical factors contributing to CK variability could explain over 10% of the variability in CK values. This is the first genome-wide study to report on the underlying genetic determinants of CK, a widely used biomarker of statin-induced myotoxicity.

320

The Return of Pharmacogenomic Variants in the MedSeq Project: Reporting Approach and Physician Response. *J.B. Krier[1,4], H.M. McLaughlin[2,6], W.J. Lane[2,5], D. Metterville[6], I. Leshchiner[3,4,6], J.L. Vassy[3,4,7], C. MacRae[3,5], M.S. Lebo[2,5,6], D. Lautenbach[4], R.C. Green[3,4], H.L. Rehm[2,5,6].* 1) Department of Pediatrics, Harvard Medical School, Boston, MA; 2) Department of Pathology, Harvard Medical School, Boston, MA; 3) Department of Medicine, Harvard Medical School, Boston, MA; 4) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA; 5) Department of Pathology, Brigham and Women's Hospital, Boston, MA; 6) Laboratory for Molecular Medicine, Partners Center for Personalized Genetic Medicine, Cambridge, MA; 7) VA Boston Healthcare System, Boston, MA.

Background. The refinement of drug selection and dosage poses an exciting opportunity for the integration of genomic sequencing data into clinical medicine. However, few examples of pharmacogenomic (PGx) testing have been adopted clinically and no standards or best practices exist in utilizing whole genome sequencing (WGS) for PGx variant reporting. The MedSeq Project is a randomized clinical trial that seeks to develop tested approaches for the evaluation of individual genomes and to assess the impact of integrating WGS, including PGx variants, into medical practice in primary care and cardiology clinics. **Methods.** MedSeq participants randomized to WGS receive a General Genome Report (GGR). Consistent with the goal of the GGR to provide a concise summary of genomic data with potential clinical relevance, a set of PGx variants were selected for inclusion in the GGR based on the level of evidence supporting their associations with specific drug responses. Among the Class I and Class II PGx variants in the PharmGKB "Clinical Annotation Levels of Evidence" classification system, a short, final list of variants and associations were selected based on their relevance in primary care and cardiology: warfarin dosing, clopidogrel efficacy, metformin efficacy, digoxin dosing, and risk of simvastatin-associated myopathy. Additional PharmGKB Class I PGx variants are also available for validation and reporting if clinically indicated by concurrent medical therapy for any MedSeq participant. **Results.** The report of PGx variants includes brief summary statements accompanied by detailed interpretations that describe specific genotype, population frequencies, clinical implication and, where relevant, links to genotype-based dosing guidelines. Initial anecdotal data suggests that physicians have significant enthusiasm for receiving PGx variants, and we describe qualitative data from MedSeq primary care physicians and cardiologists on the perceived utility of the PGx variants. Additionally, we will present data from the disclosure of the PGx results to patients and early experiences on clinical decision-making impacted by PGx variant status. **Conclusion.** PGx variants are a key component of clinical WGS reports and of great interest to physicians, and PGx reporting must recognize clinicians' desire for concise and consistent reporting formats while also providing sufficient supporting evidence and contextualization.

321

Genetic evidence improves chances of drug discovery success. *M.R. Nelson[1], H. Tipney[2], J.L. Painter[1], J. Shen[1], P. Nicoletti[3], Y. Shen[3,4], A. Floratos[3,4], P.C. Sham[5], M.J. Li[6], J. Wang[6], P. Agarwal[7], J.C. Whittaker[2], P. Sanseau[2].* 1) Quantitative Sciences, GlaxoSmithKline, Research Triangle Park, NC, USA; 2) Quantitative Sciences, GlaxoSmithKline, Stevenage, UK; 3) Initiative in Systems Biology, Columbia University, New York, NY, USA; 4) Department of Biomedical Informatics, Columbia University, New York, NY, USA; 5) Centre for Genomics Sciences, University of Hong Kong, Hong Kong SAR, China; 6) Department of Biochemistry, University of Hong Kong, Hong Kong SAR, China; 7) Quantitative Sciences, GlaxoSmithKline, Upper Merion, PA, USA.

Attrition is a major challenge in drug discovery and development with >90% of projects failing before clinical trials and >50% of the remainder failing in clinical development due to lack of efficacy. Therefore, selecting and validating the best targets is the key decision in developing medicines. Often the human evidence supporting the chosen target in a disease context is limited. However, rapid progress in deciphering the genetic basis of disease, and associated pathways, offers an opportunity to transform this process and leads to a key question: what weight should be given to a genetic association in selecting targets and indications? To address this question we merged GWASdb, a database with over 100000 genetic associations corresponding to 1228 unique traits mapped to 603 Medical Subject Heading (MeSH) terms, with Informa Pipeline, a commercial database of >23000 drugs with known human targets (including >2400 marketed) with 915 unique indications mapped to 708 MeSH terms. We drew on linkage disequilibrium, eQTL data, ENCODE-related data, and location to map genetic variants to one or more genes. We mapped 14614 genome-wide significant variants ($p < 1e-8$) to 3781 protein-coding genes (19%). We found that among 1855 unique drug targets in Pipeline, 522 (28%) had one or more genome-wide significant associations. Interestingly, targets for marketed drugs had a significantly higher percentage of genes connected to genome-wide significant associations (128 of 379 genes, 34%). Furthermore, targets for marketed drugs were 8 times more likely to be in OMIM or have a genetic association ($N = 221$; 58%) compared to all coding genes in the genome, demonstrating that genetic evidence at the target increases the chance that a drug will reach the market. We used the MeSH hierarchy to estimate relative similarities between terms to investigate the similarity between drug indication and genetically associated trait: as term similarity increases, so does the proportion of marketed drugs. Conversely, the many marketed drugs that have different indications than the genetically associated traits may suggest exciting opportunities for new indications. This work also identifies many drug indications for which there are no genetic associations, highlighting opportunities for future genetic studies. We conclude that genetic associations should play an important role in making decisions about target selection and indications to be investigated in drug development.

322

A novel mitochondrial SLC25A gene causes CMT and optic atrophy. *A.J. Abrams[1], M.A. Gonzalez[1], A.P. Rebelo[1], I.J. Campeanu[1], F.G. Speziani[1], A. Nemeth[2], J. Dallman[1], S. Züchner[1].* 1) Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Churchill Hospital, Universtiy of Oxford, Oxford, UK.

Charcot-Marie-Tooth neuropathy (CMT) is one of the most common inherited diseases in neurology. Mutations in MFN2 cause an axonal form of CMT that leads to axon degeneration in the lower limbs and is sometimes accompanied by optic atrophy. Vice versa, patients with dominant optic atrophy, caused by OPA1, can also harbor symptoms of a peripheral neuropathy. MFN2 and OPA1 are both essential factors for mitochondrial fusion suggesting a conserved mechanism involved in the overlapping phenotypes. By exome sequencing families with optic atrophy and CMT, we identified a novel solute carrier gene. Sanger sequencing confirmed that the compound heterozygous changes completely segregate in the pedigree. The gene is conserved in zebrafish and we decided to use the morpholino technology to knock down the SLC25A gene. In great congruence to our patients, morphant fish showed spinal deformities, swimming defects, abnormal motor neuron growth, and appear to be blind. Members of the SLC25 family of mitochondrial solute carrier proteins are predicted to localize to the inner mitochondria membrane and transport a large variety of molecules across the membrane. At least 11 other genetic diseases are caused by mutations in SLC25 genes and all are recessively inherited. This is the first reported case of CMT and optic atrophy to be caused by a mitochondrial carrier gene. Moreover, given the functions of MFN2 and OPA1 and the phenotypic and subcellular overlap, we are currently testing whether SLC25A is a novel, not yet characterized mitochondrial fusion protein. Indeed SLC25A shows significant conservational orthology to ugo1, a yeast mitochondrial fusion gene with no known human orthologue as of today. In summary, with the identification of SLC25A we are expanding the family of mitochondrial membrane proteins involved in optic atrophy and CMT neuropathy and open new opportunities to specifically study these pathways.

323

Molecular defects in the motor adaptor BICD2 cause proximal spinal muscular atrophy with autosomal-dominant inheritance. *K. Peeters[1,2], I. Litvinenko[3], B. Asselbergh[1,2], L. Almeida-Souza[1,2], T. Chamova[4], T. Geuens[1,2], E. Ydens[1,2], M. Zimoń[1,2], J. Irobi[1,2], E. De Vriendt[1,2], V. De Winter[1,2], T. Ooms[1,2], V. Timmerman[1,2], I. Tournev[4,5], A. Jordanova[1,2,6].* 1) Department of Molecular Genetics, VIB, Antwerp, Belgium; 2) Neurogenetics Laboratory, Institute Born-Bunge, University of Antwerp, Antwerp, Belgium; 3) Clinic of Child Neurology, Department of Pediatrics, Medical University-Sofia, Sofia, Bulgaria; 4) Department of Neurology, Medical University-Sofia, Sofia, Bulgaria; 5) Department of Cognitive Science and Psychology, New Bulgarian University, Sofia, Bulgaria; 6) Department of Medical Chemistry and Biochemistry, Molecular Medicine Center, Medical University-Sofia, Sofia, Bulgaria.

The most common form of spinal muscular atrophy (SMA) is a recessive disorder caused by deleterious *SMN1* mutations in 5q13, whereas the genetic etiologies of non-5q SMA are very heterogeneous and largely remain to be elucidated. In a Bulgarian family affected by autosomal-dominant proximal SMA, we performed genome-wide linkage analysis and whole-exome sequencing and found a heterozygous de novo c.320C>T (p.Ser107-Leu) mutation in bicaudal D homolog 2 (Drosophila) (*BICD2*). Further analysis of *BICD2* in a cohort of 119 individuals with non-5q SMA identified a second de novo *BICD2* mutation, c.2321A>G (p.Glu774Gly), in a simplex case. Detailed clinical and electrophysiological investigations revealed that both families are affected by a very similar disease course, characterized by early childhood onset, predominant involvement of lower extremities, and very slow disease progression. The amino acid substitutions are located in two interaction domains of BICD2, an adaptor protein linking the dynein molecular motor with its cargo. Our immunoprecipitation and localization experiments in HeLa and SH-SY5Y cells and affected individuals' lymphoblasts demonstrated that p.Ser107Leu causes increased dynein binding and thus leads to accumulation of BICD2 at the microtubule-organizing complex and Golgi fragmentation. In addition, the altered protein had a reduced colocalization with RAB6A, a regulator of vesicle trafficking between the Golgi and the endoplasmic reticulum. The interaction between p.Glu744Gly altered BICD2 and RAB6A was impaired, which also led to their reduced colocalization. Our study identifies *BICD2* mutations as a cause of non-5q linked SMA and highlights the importance of dynein-mediated motility in motor neuron function in humans.

324

***AMPD2* regulates *de novo* GTP synthesis and is mutated in a new form of pontocerebellar hypoplasia.** *V. Cantagrel[1, 2], N. Akizu[2], J. Schroth[2], J. Van Vleet[3], N. Cai[2], K. Vaux[2], A. Crawford[2], J.S. Silhavy[2], F.M. Sonmez[4], F. Celep[5], A. Oraby[6], M. Zaki[7], R. Al-Baradie[8], E. Faqeih[9], E. Nickerson[10], S. Gabriel[10], T. Morisaki[11], E.W. Holmes[12], J.G. Gleeson[2].* 1) INSERM U781-Institut IMAGINE, Hôpital Necker-Enfants Malades, Paris, France; 2) Neurogenetics Laboratory, Institute for Genomic Medicine, Rady Children's Hospital, Howard Hughes Medical Institute, University of California, San Diego, CA 92093, USA; 3) Glycotechnology Core Resource, University of California, San Diego, CA 92093, USA; 4) Department of Pediatric Neurology, Turgut Ozal University, Ankara, Turkey; 5) Medical Biology Department, Karadeniz Technical University, Trabzon, 61080, Turkey; 6) Pediatric Neurology Department, Cairo University Children's Hospital, Cairo 406, Egypt; 7) Clinical Genetics Department, Human Genetics and Genome Research Division, National Research Centre, Cairo 12311, Egypt; 8) Division of Pediatric Neurology, King Fahd University Hospital, Dammam, 31444, Kingdom of Saudi Arabia; 9) Division of Medical Genetics, Department of Pediatrics, King Fahad Medical City, Children's Hospital, Riyadh 11525, Kingdom of Saudi Arabia; 10) The Broad Institute of MIT and Harvard, Cambridge, MA 02141, USA; 11) National Cerebral and Cardiovascular Center Research Institute, Department of Molecular Pathophysiology, Osaka University Graduate School of Pharmaceutical Sciences, Osaka, 565-0871 Japan; 12) Sanford Consortium for Regenerative Medicine, La Jolla, CA 92037, USA.

AMP deaminase (AMPD) enzymes play a role in the maintenance of purine nucleotide homeostasis by converting adenosine monophosphate (AMP) into inosine monophosphate (IMP), the common precursor for both adenine and guanine nucleotides. Using a combination of whole-exome sequencing and linkage analysis, we identified a new distinct early-onset neurodegenerative condition with characteristic brain imaging features of pontocerebellar hypoplasia (PCH), which we propose to be called PCH type 8. We identified recessive mutations in the *AMPD2* gene in five unrelated families. All affected individuals present atrophy of the cerebellum and the pons and corpus callosum hypoplasia. Among the three human *AMPD* genes, *AMPD2* is the only one highly expressed in the cerebellum. Functional conservation of the human AMPD2 enzyme was demonstrated in yeast, however none of the patients' mutations were able to rescue the yeast phenotype. The study of *Ampd2/Ampd3* double knockout mice revealed a neurodegenerative phenotype in the hippocampus and the cerebellum. This observation supports a role for AMP deaminase activity in neuroprotection through adenosine nucleotide metabolism and/or maintenance of the guanine nucleotide pool. Using yeast and patient-derived cell lines, we quantified nucleotides levels and investigated the underlying mechanism. In the presence of precursor of adenosine, patients' cells are not able to control the feedback inhibition caused by the accumulation of adenine nucleotide which cannot get converted into guanine nucleotides. This exacerbated regulatory loop strongly inhibits the *de novo* purine biosynthesis pathway and leads to a severe depletion of guanine nucleotide resulting in a defective GTP-dependent initiation of protein translation. Administration of a purine precursor allowed us to rescue the cellular defects associated with this nucleotide imbalance.

325

Truncating mutations of *MAGEL2* cause autism and Prader-Willi Syndrome-like phenotypes. *C.P. Schaaf[1, 2], M.L. Gonzalez-Garay[3], F. Xia[1], L. Potocki[1], K.W. Gripp[4], B. Zhang[1], B.A. Peters[5], M.A. McElwain[5], R. Drmanac[5], A.L. Beaudet[1], C.T. Caskey[1], Y. Yang[1].* 1) Department of Molecular & Human Genetics, Baylor College Med, Houston, TX; 2) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX; 3) The Brown Foundation Institute of Molecular Medicine for the Prevention of Human Diseases, The University of Texas, Health Science Center at Houston, Houston, TX; 4) Division of Medical Genetics, Alfred I. duPont Hospital for Children, Wilmington, DE; 5) Complete Genomics, Inc., Mountain View, CA.

Prader-Willi Syndrome is caused by the absence of paternally expressed, maternally silenced genes at 15q11-q13. We report four individuals with truncating mutations on the paternal allele of *MAGEL2*, a gene within the Prader-Willi syndrome (PWS) domain. The first case was ascertained by whole genome trio analysis for PWS features. Three additional patients were identified in a cohort of 400 cases submitted for clinical whole exome sequencing to a clinical laboratory. The phenotypes of the four probands ranged from meeting criteria for PWS to some features of PWS, but autism spectrum disorders (ASDs) were present in all four probands. The reported *MAGEL2* mutations are de novo in three cases, and not inherited from the mother in one case (father unavailable). Using two different methodologies, we show that the mutations are on the paternal allele of the *MAGEL2* gene in all four cases, and therefore probably pathogenic. First, we performed long fragment analysis in conjunction with parental SNP genotypes in proximity to the mutated locus. Second, we digested genomic DNA with the methylation-sensitive restriction enzyme SmaI, followed by long-range PCR and Sanger sequencing. As the unmethylated (paternal) allele is digested by SmaI, only the maternal allele is amplified in the subsequent PCR, and all sequencing reads reflect the sequence of the maternal *MAGEL2* allele. In summary, our findings suggest *MAGEL2* is a novel gene causing complex ASDs, and lack of a functional *MAGEL2* copy can contribute to several aspects of the PWS phenotype.

326

Defective ubiquitination underlies oligogenic cerebellar degeneration and reproductive endocrine axis defects. *M. Kousi[1], D. Margolin[2], Y.M. Chan[3,4], V. Muto[5], S. Servidei[6], E.T. Lim[3], J.D. Schmahmann[2], M. Hadjivassiliou[7], J.E. Hall[3], I. Adam[8], A. Dwyer[3], L. Plummer[3], S.V. Aldrin[3], J. O'Rourke[3], A. Kirby[9], K. Lage[9,10,11,12], A. Milunsky[13], J.M. Milunsky[13], J. Chan[14], E.T. Hedley-Whyte[15], M.J. Daly[9], M. Tartaglia[5], S.B. Seminara[3], N. Katsanis[1,16].* 1) Center for Human Disease Modeling, Department of Cell Biology, Duke University Medical Center, Durham, NC; 2) Department of Neurology, Massachusetts General Hospital, Boston, MA; 3) Harvard Reproductive Sciences Center and Reproductive Endocrine Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA; 4) Division of Endocrinology, Department of Medicine, Boston Children's Hospital, Boston, MA; 5) Dept. Hematology, Oncology and Molecular Medicine, Section of "Physiopathology of Genetic Diseases", Istituto Superiore di Sanità, Rome, Italy; 6) Dept. Neurosciences, Institute of Neurology, Università Cattolica del Sacro Cuore, Rome Italy; 7) Department of Neurology, The Royal Hallamshire Hospital, Glossop Road, Sheffield, S10 2JF, UK; 8) Specialty Hospital, Amman, Jordan; 9) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA; 10) Pediatric Surgical Research Laboratories, Massachusetts General Hospital, Boston, MA; 11) Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark; 12) Center for Protein Research, University of Copenhagen, Copenhagen, Denmark; 13) Center for Human Genetics, Inc., Boston, MA; 14) Department of Pathology, Brigham and Women's Hospital, Boston, MA; 15) Department of Neuropathology, Massachusetts General Hospital, Boston, MA; 16) Department of Pediatrics, Duke University Medical Center, Durham, NC.

We have studied a cohort of ataxia patients who manifest a unique constellation of cerebellar degeneration and defects of the reproductive endocrine axis, hypothesizing that understanding of the genetic drivers in these individuals might inform biological pathways relevant to multiple neuronal populations. Subsequent to whole-exome sequencing in a large consanguineous pedigree and systematic annotation of all rare alleles that segregate with the phenotype, we identified candidate pathogenic mutations in two genes, RNF216 and OTUD4 both of which encode components of the ubiquitin-proteasome pathway. Further analyses in unrelated patients showed that in some families, truncating mutations in RNF216 are likely sufficient to cause the disorder. By contrast, patients with functionally interrogated hypomorphic alleles require additional trans mutations in OTUD4. In vivo testing of not only each allele found in our patients but also of the potential for genetic interaction in zebrafish embryos showed that RNF216 and OTUD4 interact genetically to affect both neurogenesis and cerebellar architecture during development. Encouraged by these findings, we analyzed additional families and identified mutations in additional members of the ubiquitin pathway, highlighting a central role of this fundamental cellular process in cerebellar development and maintenance. Further, these observations have suggested that small molecules known to augment proteasomal function and correct autophagy might be of broad benefit to these phenotypes; preliminary in vivo testing is supportive of this hypothesis. Taken together, our data implicate dysfunction of a core biological process as a driver for neuronal degeneration, highlight the importance of mutational load on biological pathways as a determinant of disease onset and suggest that chemical amelioration of the pathway might be of therapeutic benefit.

327

Periventricular heterotopia in 6q terminal deletion syndrome: role of the C6orf70 gene. *V. Conti[1], A. Carabalona[2,3], E. Pallesi-Pocachard[2,3], E. Parrini[1], R. Leventer[4,5,6], E. Buhler[7], G. McGillivray[8], F. Michel[2,3], P. Striano[9], D. Mei[1], F. Watrin[2,3], S. Lise[10], A. Pagnamenta[10], J. Taylor[10], U. Kini[11], J. Clayton-Smith[12], F. Novara[13], O. Zuffardi[13], W. Dobyns[14], I. Scheffer[15,16], S. Robertson[17], S. Berkovic[15], A. Represa[2,3], D. Keays[18], C. Cardoso[2,3], R. Guerrini[1,19]*. 1) Pediatric Neurology and Neurogenetics Unit and Laboratories, A. Meyer Children's Hospital - Department of Neuroscience, Pharmacology and Child Health, University of Florence, Florence, Italy; 2) INMED INSERM U901, Marseille, France; 3) Aix-Marseille University, Marseille, France; 4) Department of Neurology, Royal Children's Hospital, Parkville Victoria, Australia; 5) Murdoch Children's Research Institute, Parkville Victoria, Australia; 6) Department of Paediatrics, University of Melbourne, Victoria, Australia; 7) Plateforme postgenomique INMED-INSERM, Marseille, France; 8) Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Victoria, Australia; 9) Pediatric Neurology and Muscular Diseases Unit, Department of Neurosciences, Rehabilitation, Ophtalmology, Genetics, Maternal and Child Health, University of Genoa, "G. Gaslini" Institute, Genova, Italy; 10) NIHR Biomedical Research Centre, Wellcome Trust Centre for Human Genetics, Oxford, UK; 11) Department of Clinical Genetics, Oxford Radcliffe NHS Trust, Oxford, UK; 12) Genetic Medicine, St Mary's Hospital, University of Manchester, UK; 13) Department of Pediatric Science and Human and Hereditary Pathology, General Biology and Medical Genetics Section, University of Pavia, Pavia, Italy; 14) Center for Integrative Brain Research Seattle Children's Research Institute Seattle WA; 15) Epilepsy Research Centre, Department of Medicine, University of Melbourne, Austin Health, Victoria, Australia; 16) Florey Institute for Neuroscience and Mental Health, and Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Parkville Victoria, Australia; 17) Department of Paediatrics and Child Health, Dunedin School of Medicine University of Otago, Dunedin, New Zealand; 18) Research Institute of Molecular Pathology, Vienna, Austria; 19) IRCCS Stella Maris Foundation, Pisa, Italy.

Periventricular nodular heterotopia (PNH) is caused by defective neuronal migration that results in heterotopic neuronal nodules lining the lateral ventricles. Mutations in Filamin-A (FLNA) or ADP-ribosylation factor guanine nucleotide-exchange factor-2 (ARFGEF2) cause PNH, but most patients with this malformation do not have an assigned aetiology. Using comparative genomic hybridization (array-CGH), we identified 12 patients with developmental brain abnormalities, variably combining PNH, corpus callosum dysgenesis, colpocephaly, cerebellar hypoplasia and polymicrogyria, harboring a common 1.2 Mb minimal critical deletion in 6q27. These anatomic features were mainly associated with epilepsy, ataxia and cognitive impairment. Using whole exome sequencing in 14 patients with isolated PNH but no copy number variants, we identified one patient with PNH, developmental delay and epilepsy and a de novo missense mutation in the chromosome 6 open reading frame 70 (C6orf70), mapping in the minimal critical deleted region. Using immunohistochemistry and western blot, we demonstrated that in human cell lines, C6orf70 shows primarily a cytoplasmic vesicular puncta-like distribution and that the mutation affects its stability and subcellular distribution. We also performed in utero silencing of C6orf70 and of Phf10 and Dll1, the two additional genes mapping in the 6q27 minimal critical deleted region that are expressed in human and rodent brain. Silencing of C6orf70 in the developing rat neocortex produced a PNH phenotype that was rescued by concomitant expression of wild-type human C6orf70 protein. Silencing of the contiguous Phf10 or Dll1 genes only produced slightly delayed migration but not PNH. The complex brain phenotype observed in the 6q terminal deletion syndrome likely results from the combined haploinsufficiency of contiguous genes mapping to a small 1.2 Mb region. Our data suggest that, of the genes within this minimal critical region, C6orf70 plays a major role in the control of neuronal migration and its haploinsufficiency or mutation causes PNH.

328

A GENE IMPLICATED IN THE NEUROBEHAVIOURAL ABNORMALITIES OF WILLIAMS-BEUREN SYNDROME, GTF2IRD1, ENCODES A NOVEL EPIGENETIC REGULATOR. *P. Carmona-Mora[1], J. Widagdo[1], F. Tomasetig[1], K.M Taylor[1], Y. Cha[1], R.T-W Pang[1], N.A. Twine[2], M.R. Wilkins[2], P.W. Gunning[3], E.C. Hardeman[1], S.J. Palmer[1]*. 1) School of Medical Sciences, Neuromuscular and Regenerative Medicine Unit, University of New South Wales, Sydney, NSW 2052, Australia; 2) School of Biotechnology and Biomolecular Sciences, The New South Wales Systems Biology Initiative, University of New South Wales, NSW 2052, Australia; 3) School of Medical Sciences, Oncology Research Unit, University of New South Wales, Sydney, NSW 2052, Australia.

Williams-Beuren syndrome (WBS) is an autosomal dominant disorder resulting from a hemizygous microdeletion within chromosome 7q11.23. The clinical presentation of this multisystem disorder includes craniofacial abnormalities and a distinctive neurocognitive profile. WBS neurobehavioural abnormalities involve intellectual disability, reduced social anxiety, high incidence of psychopathologies such as ADHD and phobias and a visuospatial construction deficit. Genotype/phenotype correlations in patients with atypical deletions implicate a gene discovered in our laboratory, *GTF2IRD1*, as responsible for the distinctive neurocognitive profile of WBS. However, the molecular and cellular consequences of *GTF2IRD1* haploinsufficiency remain unknown. To reveal the normal function of GTF2IRD1 we have combined protein interaction studies, gene expression and transcript profile analyses in *Gtf2ird1* knockout mice. The first approach comprised a yeast two-hybrid screening system to discover the molecular interactions of GTF2IRD1. We identified a panel of interacting partners that fall into functional groups, such as DNA binding proteins, post-translational modification machinery (SUMO and ubiquitin ligation proteins) and chromatin modifying factors involved in histone methylation, deacetylation and ubiquitination. We also found that GTF2IRD1 interacts with multiple members of the same gene family, thus expanding its interactional network and illustrating conserved binding domains. Immunofluorescence and co-immunoprecipitation in mammalian cells support these interactions and show a specific pattern of subnuclear localization for endogenous GTF2IRD1. To assess the effect of GTF2IRD1 on gene expression, we conducted microarray transcript profile analyses in corpus striatum tissue from Gtf2ird1 KO mice followed by qRT-PCR validation. This KO mouse model has been shown to mirror many of the defects of WBS, including ataxia and abnormalities of locomotor drive. We found an increased expression of genes involved in neuronal development and a cluster of immediate-early response genes that have previously been linked with hyperactivity. Our data are consistent with a role for GTF2IRD1 as an epigenetic regulator of gene repression that coordinates interactions with transcription factors, DNA binding proteins and components of the chromatin modification machinery and provide a possible explanation for the underlying molecular cause of locomotor changes in WBS patients.

329

Targeted High-Throughput Sequencing of 220 genes identifies a high proportion of causative mutations in over 80 patients with undiagnosed intellectual disability. *C. Redin[1, 2], S. Le Gras[3], J. Lauer[4], A. Creppy[1, 4], Y. Herenger[4], V. Geoffroy[3], Y. Alembik[5], M. Doco-Fenzy[6], B. Doray[5], P. Edery[7], S. El Chehadeh[8], L. Faivre[8, 9], E. Flori[10], B. Isidor[11], G. Lesca[7], A. Masurel[8], B. Jost[3], J. Muller[1, 4], B. Gérard[4], J.L. Mandel[1, 2, 4], A. Piton[1, 2].* 1) Translational medicine & Neurogenetics, IGBMC, Illkirch, France; 2) Chaire de Génétique Humaine, Collège de France; 3) Microarray and Sequencing Platform, IGBMC, Illkirch, France; 4) Laboratoire de Diagnostic Génétique, Hôpitaux Universitaires de Strasbourg, France; 5) Service de Génétique Médicale, Hôpitaux Universitaires de Strasbourg, Strasbourg, France; 6) Biology Section, Department of Genetics, University Hospital of Reims, Reims, France; 7) Laboratoire de Cytogénétique Constitutionnelle, Service de Génétique, Centre de Biologie et de Pathologie Est, Hospices Civils de Lyon, Bron, France; 8) Centre de génétique et Centre de Référence Anomalies du développement et Syndromes malformatifs, Hôpital d'Enfants, CHU Dijon, Dijon, France; 9) Génétique des anomalies du développement (GAD) EA 4271, Faculté de Médecine, Université de Bourgogne, Dijon, France; 10) Service de cytogénétique, Hôpitaux Universitaires de Strasbourg, Strasbourg, France; 11) Service de Génétique Médicale, Centre Hospitalier Universitaire de Nantes 7, Quai Moncousu, 44000 Nantes Cedex, France.

Over 200 genes have a well-documented implication in monogenic forms of intellectual disability (ID), of which about 100 are located on the X-chromosome. While there is a good coverage for diagnostic demands in patients with evocative syndromic forms, for lesser syndromic patients the diagnostic offer is limited to Fragile-X testing and CGH-array analysis and most cases hence remain undiagnosed. We tested a targeted exon-capture approach of 220 ID genes (100 XLID genes and 120 autosomal genes mostly associated to dominant forms) coupled with NGS. We successfully analyzed 82 patients (mostly males and simplex cases) with undiagnosed ID. We identified 15 causative mutations: 9 maternally-inherited or de novo in X-linked genes (DMD, KDM5C, MAOA, MECP2, SLC9A6, FMR1, IQSEC2, IL1RAPL1, SHROOM4) and 6 de novo truncating in autosomal genes (RAI1, SLC2A1, two in DYRK1A, two in TCF4). We also detected 5 likely-causative mutations requiring additional validation analyses. Some causative mutations were rather surprising either by the nature of the mutation itself or when considering the patient's phenotype. As examples, we depicted a distal frameshift in DMD in a patient with no muscular phenotype; an extremely complex rearrangement in MECP2 in an 11 years-old boy with severe ID and speech absence; a deletion of FMR1 last exon in a severely affected male with previous extensive molecular investigation; de novo nonsenses in RAI1 and TCF4 in patients without phenotypic traits evocative of Smith-Magenis or Pitt-Hopkins syndromes. Other findings allow supporting previously proposed ID genes such as MAOA, with the first replication of its involvement in ID along with prominent behavioral disturbances (Brunner et al., 1993). Lastly, we report 2 de novo truncating mutations in DYRK1A, in patients with phenotype concordant with previous ones, delineating a peculiar syndrome with microcephaly, speech absence and feeding disorder. The identification of certainly/likely causative mutations in 18-24% of the patients proves the relevance of our strategy for the diagnosis of ID, being more cost-effective than trio-exome sequencing while leading to a comparable proportion of diagnostic results (21-41% in 151 trios, de Ligt,2012 and Rauch, 2012). Finally, the identification of causative mutations in syndromic genes in patients deviating from the typical phenotype expand the clinical spectrum associated with such genes and suggest they should be screened more systematically.

330

Genetic Analysis and New Gene Discovery in Nemaline Myopathy. *V.A. GUPTA[1], G. RAVENSCROFT[2], R. SHAHEEN[3], E.J. TODD[2], L.C. SWANSON[1], M. SHIINA[4], K. OGATA[4], C. HSU[1], N.F. CLARKE[6], B.T. DARRAS[7], M. FARRAR[8], A. HASHEM[3], N. MANTON[2], F. MUNTONI[9], K.N. NORTH[10], S. SANDARADURA[8], I. NISHINO[11], Y.K. HAYASHI[11], C.A. SEWRY[9], E. THOMPSON[2], T.W. YU[1], C.A. BROWNSTEIN[1], R. ALLCOCK[2], M.R. DAVIS[2], C. WALLGREN-PETTERSSON[12], N. MATSUMOTO[5], F.S. ALKURAYA[3], N.G. LAING[2], A.H. BEGGS[1].* 1) Genomics Program and Division of Genetics, The Manton Center for Orphan Disease Research, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA; 2) Western Australian Institute for Medical Research and the Centre for Medical Research, University of Western Australia, Nedlands, Western Australia; 3) Developmental Genetics Unit, King Faisal Specialist Hospital and Research Center, Riyad 11211, Saudi Arabia; 4) Department of Biochemistry, Yokohama City University, Graduate School of Medicine, 3-9 Fukuura, Kanazawa-ku, Yokohama 236-0004, Japan; 5) Department of Human Genetics, Yokohama City University, Graduate School of Medicine, 3-9 Fukuura, Kanazawa-ku, Yokohama 236-0004, Japan; 6) Institute for Neuroscience and Muscle Research, Children's Hospital at Westmead, Sydney, NSW2145, Australia; 7) Department of Neurology, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA; 8) Department of Neurology, Sydney Children's Hospital and School of Women's and Children Health, University of South Wales, Australia; 9) Dubowitz Neuromuscular Centre, Institute of Child Health and Great Ormond Street Hospital, London WC1N 1EH, UK; 10) Murdoch Childrens Research Institute, The Royal Children's Hospital, Flemington Road Parkville Victoria 3052, Australia; 11) 11Department of Neuromuscular Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Tokyo 187-8502, Japan; 12) The Folkhälsan Institute of Genetics and Department of Medical Genetics, Haartman Institute, University of Helsinki, Biomedicum Helsinki, Finland.

Nemaline myopathy (NM) is one of the most common congenital myopathies and is characterized by generalized skeletal muscle weakness and the presence of nemaline bodies in affected myofibers. Clinically, the NMs form a heterogeneous group of myopathies ranging from fetal akinesia, through the severe congenital forms with death from respiratory failure during first year of life, to a mild-childhood onset myopathy with survival into adulthood. Substantial progress has been made in understanding the genetic basis of these conditions. NMs with known genetic cause include mutations in *ACTA1*, *NEB*, *TPM3*, *TPM2*, *CFL2*, *TNNT1* and *KBTBD13* and the most recently identified, *KLHL40*. Mutations in these genes account for only ~ 60% of all genetically identified cases. To identify mutations in cases with unknown genetic cause, we performed next generation sequencing on 60 NM patients. Data analysis and subsequent validation by Sanger sequencing of this cohort identified novel mutations in known genes associated with NM as well as novel disease genes. We identified 26 *NEB*, 3 *ACTA1*, 1 *TPM3* and 1 *TNNT1* mutation in the 60 patients. WES also resulted in the identification of a new disease locus, *KLHL41*, in five families affected with NM. Kelch-like protein 41 (KLHL41) belongs to the family of Kelch and BTB domain containing proteins that includes two other NM loci, KLHL40 and KBTBD13. Mutations in *KLHL41* showed a clear phenotype-genotype correlation. Small deletions resulted in severe phenotypes with neonatal death, whereas missense changes resulted in an impaired motor function with survival in to late childhood and/or early adulthood. Functional studies in zebrafish showed that loss of *klhl41* results in highly diminished motor function and myofibrillar disorganization with thickened Z-lines and nemaline body formation, the pathological hallmark of NM. The BTB domain containing Kelch family members interact with Cul3 ubiquitin ligase to form a functional ubiquitination complex. The interaction of KLHL40, KLHL41 and KBTBD13 with E3 ubiquitin ligases implicates the involvement of Kelch proteins mediated ubiquitination pathway in disease pathology in NM. Further studies on the mechanism of Kelch protein regulated pathways will allow a clearer understanding of NM and may assist in future therapeutics developments.

331

Large scale meta analysis of 250,000 individuals reveals novel biological pathways involved in adult human height. *T. Esko[1], AR. Wood[2], S. Vedantam[1], J. Yang[3], TH. Pers[1], SI. Berndt[4], MN. Weedon[2], G. Lettre[5], J. O'Connell[6], DI. Chasman[7], G. Abecasis[8], ME. Goddard[3], RJF. Loos[9,10], E. Ingelsson[11], PM. Visscher[3], JH. Hirschhorn[1], TM. Frayling[2], on. behalf of GIANT Consortium[1].* 1) Divisions of Endocrinology, Boston Children's Hospital and Broad Institute, Cambridge, MA, USA; 2) Genetics of Complex Traits, Exeter Medical School, Exeter, UK; 3) University of Queensland Diamantina Institute, University of Queensland, Princess Alexandra Hospital, Brisbane, Queensland, Australia; 4) 4Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA; 5) Montreal Heart Institute, Montreal, Quebec, Canada; 6) Deptartment of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA; 7) 7Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 8) Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA; 9) MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge, UK; 10) Mount Sinai School of Medicine, New York, NY, USA; 11) Department of Medical Sciences, Uppsala University Hospital, Uppsala, Sweden.

Adult height is a highly heritable polygenic trait that reflects the outcome of childhood growth, a fundamental developmental process. Studies of height have already employed large sample sizes (~130,000) and identified more loci (180) than for any other polygenic trait or disease, but have nonetheless only accounted for one tenth of the phenotypic variation. Although further expansion of GWAS sample size is predicted to substantially augment the explained heritability, it is not yet known if newly identified variants with very small effects will keep providing novel insight to human growth biology. We report results from nearly doubling the sample size of existing height GWAS data to ~250,000 individuals of European ancestry. By applying an approximate conditional analysis we identified 697 independent variants shown to cluster into 424 genomic loci (P<1e-4). Associated variants were not only non-randomly distributed with respect to functional and putatively functional regions of the genome (nsSNPs, eQTLs and epigenetic marks) but showed high allelic heterogeneity (262 secondary signals) in both established and novel loci. Variants in strictly novel loci were more prominently enriched for eQTLs and not for nsSNPs, which suggest that associations with smaller effect sizes will increasingly point to regulatory variants. By applying both existing and novel pathway enrichment and gene prioritizing bioinformatics tools, we provide evidence that identification of many 100's and even 1000's of associated variants will continue to provide biologically relevant information. The larger number of loci highlight many more significantly enriched growth-related pathways, including signalling by insulin-like growth factors, Hedgehog, WNT, BMPs and TGF-beta, mTOR, and MAPK, as well as chromatin remodelling. Our results prioritize 639 genes (FDR <5%), which reveal enriched expression in growth-related tissues (cartilage; P<1e-12) and include many known skeletal growth syndrome genes, This list provides new candidates for regulating human growth or for underlying diseases of abnormal growth. In conclusion, larger sample sizes have identified further height associated variants that as a group provide novel and more specific biological insights into human growth. By extension, larger GWA studies from other diseases and traits, for which there are fewer associated loci than for height, are likely to continue to provide additional insights into underlying biology.

332

Opening the X files - Chromosome X-wide association study reveals new loci for fasting insulin and height and evidence for incomplete dosage compensation. *T. Tukiainen[1,2,3], M. Pirinen[2], A.-P. Sarin[2,4], C. Ladenvall[5], J. Kettunen[2,4], T. Lehtimäki[6], M.-L. Lokki[7], M. Perola[2,8,9], J. Sinisalo[10], E. Vlachopoulou[7], J.G. Eriksson[8,11,12,13,14], L. Groop[2,5], A. Jula[15], M.-R. Järvelin[16,17,18,19,20], O.T. Raitakari[21,22], V. Salomaa[8], S. Ripatti[2,4,23,24].* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland; 3) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, USA; 4) Unit of Public Health Genomics, National Institute for Health and Welfare, Helsinki, Finland; 5) Department of Clinical Sciences, Diabetes and Endocrinology, Lund University and Lund University Diabetes Centre, CRC at Skåne University Hospital, Malmö, Sweden; 6) Department of Clinical Chemistry, Fimlab Laboratories, University of Tampere School of Medicine, Tampere, Finland; 7) Transplantation Laboratory, Haartman Institute, University of Helsinki, Finland; 8) Department of Chronic Disease Prevention, National Institute for Health and Welfare, Finland; 9) Estonian Genome Center, University of Tartu, Tartu, Estonia; 10) Division of Cardiology, Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland; 11) Department of General Practice and Primary Healthcare, University of Helsinki, Finland; 12) Unit of General Practice, Helsinki University Central Hospital, Finland; 13) Folkhälsan Research Center, Helsinki, Finland; 14) Vaasa Central Hospital, Vaasa, Finland; 15) Population Studies Unit, Department of Chronic Disease Prevention, National Institute for Health and Welfare, Turku, Finland; 16) Department of Epidemiology and Biostatistics, MRC Health Protection Agency (HPA) Centre for Environment and Health, School of Public Health, Imperial College London, United Kingdom; 17) Institute of Health Sciences, University of Oulu, Finland; 18) Biocenter Oulu, University of Oulu, Finland; 19) Unit of Primary Care, Oulu University Hospital, Finland; 20) Department of Children and Young People and Families, National Institute for Health and Welfare, Oulu, Finland; 21) Department of Clinical Physiology and Nuclear Medicine, University of Turku and Turku University Hospital, Turku, Finland; 22) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku and Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland; 23) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 24) Hjelt Institute, University of Helsinki, Helsinki, Finland.

Thus far chromosome X SNPs have been largely neglected in genome-wide association studies. ChrX, however, contains approximately 5% of genomic DNA and 3% of known SNPs and thus represents one potential source for the "missing heritability" for complex phenotypes. Here we demonstrate the benefits of including chromosome X in large-scale genetic studies. In order to comprehensively survey the contribution of chrX to common quantitative phenotypes we imputed >400,000 good-quality non-pseudoautosomal chrX SNPs in 19,697 Finnish and Swedish individuals utilizing the imputation reference from the 1000 Genomes Project. We focused our screen on twelve anthropometric and cardiometabolic phenotypes for which tens of autosomal loci have been identified: height, body-mass-index, waist-hip-ratio, systolic and diastolic blood pressure, C-reactive protein, insulin, glucose, total, LDL and HDL cholesterol and triglycerides. Using a linear mixed model we estimate that common and low frequency SNPs in chrX contribute up to 1.4% of the total variance of the twelve phenotypes. The highest estimate was observed for height. Additionally, more than 0.5% of the variance in systolic blood pressure, HDL-C, fasting glucose and insulin appear to be due to X chromosome variation. In a chromosome X-wide association analysis we identify three novel loci: one for fasting insulin (rs139163435 in Xq23, P-value = 2.87×10^{-8}) and two for height (rs182838724 near *ATRX* and rs1751138 near *ITM2A*, P-values 4.69×10^{-8} and 5.97×10^{-10}, respectively). *ITM2A* is a likely functional candidate in the latter height locus: it has been implicated in cartilage development and is an eQTL in whole blood (P-value = 6.23×10^{-14}, N = 513). Evaluation of the different models of dosage compensation in the three loci shows that there is a lack of dosage compensation in females near *ITM2A*. This observation is supported by gene expression data and a previous cellular study, which showed that *ITM2A* escapes from X chromosome inactivation in the majority of women. Our findings provide a clear motivation to assess chrX associations in larger sample sizes to identify further loci for complex traits. In addition, linking phenotype information to loci that escape from the X chromosome inactivation can bring insights into the biological bases of sexually dimorphic traits and sex chromosome aneuploidies.

333

Expanded and Novel Loci for A1c Levels Identified Through a Trans-Ethnic Meta-Analysis Approach in European and African American Ancestry Samples. *E. Wheeler[1], M-F. Hivert[2,3], C-T. Liu[4] on behalf of MAGIC and AAGILE.* 1) Human Genetics, Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) Department of Medicine, Université de Sherbrooke, Sherbrooke, Quebec, Canada; 3) General Medicine Division, Massachusetts General Hospital, Boston, MA, USA; 4) Department of Biostatistics, Boston University, Boston, MA, USA.

A1c is now recommended for type 2 diabetes (T2D) diagnosis. The Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) previously undertook a genome-wide association study (GWAS) meta-analysis of A1c levels in up to 45,000 non-diabetic European individuals and revealed 10 loci ($P<5\times10^{-8}$). MAGIC has now performed an updated meta-analysis in up to 132,914 European individuals from 28 GWAS and 20 Metabochip cohorts. In parallel, the African American Glucose and Insulin Genetic Epidemiology (AAGILE) Consortium has performed a meta-analysis of GWAS for A1C in up to 7099 non-diabetic individuals from 8 cohorts. Variants were excluded if minor allele frequency (MAF)<0.01, HWE<10-4, call rate<0.95, or failed imputation. In AAGILE, a missense variation in *G6PD* (Val==>Met) previously associated with glucose-6-phosphatase dehydrogenase deficiency was strongly associated with lower A1c levels ($P=6.8\times10^{-122}$) in line with higher red cell turnover. This variant is monomorphic in populations of European descent but common in African descent populations (YRI hapmap: MAF=0.22). Combining the European and African descent samples using the Meta-ANalysis of Transethnic Association Studies (MANTRA) software, confirmed all 10 previously identified MAGIC loci and revealed an additional 37 loci including 6 loci known to be associated with T2D and/or glucose levels (eg. *CDKAL1*, $\log_{10}BF = 9.98$ and *SLC2A2*, $\log_{10}BF = 10.89$) and 7 loci previously found to be associated with red blood cell traits or iron metabolism (eg. *BET1L*, $\log_{10}BF = 8.81$). Among the loci not previously associated with A1c levels, many are located near genes of high biologic relevance, including a variant in the intronic region of *ATAD2B* ($\log_{10}BF = 13.17$). Follow-up of these newly identified loci will include fine-mapping, leveraging differences in local linkage disequilibrium structure between the ethnic groups. The newly identified loci improve our understanding of the genetic determinants of A1c, and our findings might influence the use of A1c for T2D diagnosis in specific ethnic backgrounds.

334

Genome-wide association study for serum metabolome reveals 57 associated loci for biomarkers of complex metabolic diseases. *J. Kettunen[1], T. Haller[2], A. Demirkan[3], R. Rawal[4], T. Tukiainen[5], T. Esko[2,5], L.C. Karssen[3], C. Gieger[4], H.K. Dharuri[6], J.B. van Klinken[6], K.W. van Dijk[6], M. Waldenberger[7], M. Ala-Korpela[8], P. Soininen[8], A.J. Kangas[8], T. Lehtimäki[9], M. Perola[10], C. van Duijn[3], J.G. Eriksson[11], T. Illig[12], A. Metspalu[2], A. Jula[13], M.-R. Järvelin[14], J. Kaprio[15], O. Raitakari[16], V. Salomaa[17], A. Palotie[1], S. Ripatti[1].* 1) Institute for Molecular Medicine, Helsinki, Finland, FIMM; 2) Estonian Genome Center, University of Tartu, Tartu, Estonia; 3) Genetic Epidemiology Unit, Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 4) Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 5) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, USA; 6) Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; 7) Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 8) Computational Medicine, Institute of Health Sciences, University of Oulu, Oulu, Finland; 9) Department of Clinical Chemistry, Fimlab Laboratories, Tampere University Hospital and University of Tampere, Finland; 10) Unit for Genetic Epidemiology, National Institute for Health and Welfare, Helsinki, Finland; 11) Department of General Practice and Primary Health Care, University of Helsinki and the Helsinki University Hospital, Helsinki, Finland; 12) Hannover Unified Biobank, Hannover Medical School, Hannover, Germany; 13) Department of Chronic Disease Prevention, National Institute for Health and Welfare, Turku, Finland; 14) Department of Epidemiology and Biostatistics, Imperial College, London, UK; 15) Department of Public Health, University of Helsinki, Helsinki, Finland; 16) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; 17) Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland.

Our aim was to search for genetic determinants for circulating biomarkers of complex metabolic diseases to provide tools for causal inference between biomarker and disease. We performed a genome-wide association study (GWAS) for 123 metabolites measured using nuclear magnetic resonance spectroscopy (NMR). We had 10 European study populations with both metabolite data and genome-wide SNP data available, totaling up to 17 800 individuals. The SNP arrays were extended with imputation to 42 million markers by using April 2012 release of the 1000 Genomes project. We found 57 independent loci associated with one or more metabolic variables using multiple testing corrected genome-wide significance ($p < 2.7*10^{-9}$). Out of these 57 significant loci, the lead associated measure was lipoprotein related in 24 cases, lipid related in 5 cases and small molecule or protein in 28 cases. In 7 cases the variant with strongest association was insertion or deletion. Out of the 57 significant loci, 12 loci were novel meaning that same locus had not been associated with same or similar metabolic measure before. Of the 12 novel loci, 1 was for lipoproteins, 10 were for small molecules and 1 for lipid measures. Further, we found novel loci for recently found type 2 diabetes biomarkers: One locus for phenylalanine and three for glycine. In addition, we found a new locus for creatinine, which is a biomarker for chronic kidney disease. The new loci for type 2 diabetes biomarkers harbored enzymes known to be involved in the metabolism of glycine or phenylalanine. The phenylalanine locus contained the phenylalanine hydroxylase gene (OMIM 612349), which harbors known variants cause phenylketonuria (PKU). The glycine loci contained the two genes of glycine cleavage system. Mutations in these genes have been shown to cause Mendelian disease Glycine encephalopathy (GE, OMIM 605899). PKU and GE both include patients with a spectrum of symptomatic severity. These three loci could potentially be associated with less severe forms or trait components of GE and PKU, however, screening in relevant patient samples is required. As a conclusion, this study provides new genetic determinants for several complex disease biomarkers and therefore provides better tools for evaluating causality between biomarkers and complex diseases such as type 2 diabetes and chronic kidney disease.

335

Using correlated phenotypes to functionally classify GWAS loci. *N. Eriksson, J.Y. Tung, D.A. Hinds.* 23andMe, Mountain View, CA.

While there have been thousands of genetic loci convincingly associated with hundreds of different phenotypes through GWAS, the function of only a few of these associations has been uncovered. New methods are needed in order to uncover the pathways through which these associations function. Luckily, there are a large number of pleiotropic relationships among these loci (i.e., loci that affect multiple different phenotypes). This may occur with highly correlated traits (e.g., eye color and hair color), modestly correlated traits (e.g., different autoimmune diseases) or even seemingly unrelated traits (e.g., Parkinson's disease and baldness). Here, we take advantage of these pleiotropic effects across the wide range of phenotypes collected from 23andMe customers in order to cluster GWAS SNPs by functional category. We jointly model the correlation structure between a set of phenotypes using generalized estimating equations and look across all SNPs associated with at least one of the phenotypes, allowing us to classify them according to their effect across all the phenotypes. We apply this method to three examples from the 23andMe dataset, showing how correlated phenotypes can provide insight into the function of disease associated loci. First, we look at breast cancer SNPs using breast size as a proxy. There is at most a weak correlation between breast size and breast cancer risk, however we find that many of the loci associated with breast cancer also show associations with breast size. Of these, most have the cancer risk allele leading to larger size, however, several lead to smaller size, leading to clues about the functions of these regions. Next, we study BMI and food choice. It is suspected that several BMI SNPs, such as those in FTO, are related to BMI through neurological pathways. Here, we show that indeed, several of the 32 SNPs associated with BMI show a consistent signal across several food choice phenotypes, whereas others don't, suggesting a division of BMI loci into those involved in BMI through food choice and those through other pathways. Finally, we jointly investigate hair color and basal cell carcinoma. Our method leads to the discovery of several novel associations (including TGM3, TERT, and GATA3 for skin cancer). These SNPs show a firm division of skin cancer loci into those related to pigmentation and those shared with other cancers. We believe this method will be of increasing importance as the number of phenotypes studied grows.

336

Meta-analysis of SNP associations with body mass index in >339,000 individuals gives new genetic and biological insights into the underpinnings of obesity. *E.K. Speliotes[1,2], A.E. Locke[3], S. Berndt[4], B. Kahali[1,2], A. Justice[5], T. Pers[6], J. Yang[7], F. Day[8], S. Gustafsson[9], C. Powell[1,2], S. Vedantam[5,10,11], D.C. Croteau-Chonka[12,13], T. Winkler[14], A. Scherag[15], I. Barroso[16,17], J.S. Beckmann[18,19], C.M. Lindgren[20], C.J. Willer[21], P. Visscher[7], K.L. Mohlke[12], K.E. North[5], E. Ingelsson[20,22], J.N. Hirschhorn[5,10,11], R.J.F. Loos[8,23] for the GIANT Consortium.* 1) Internal Med, Gastroenterology, University of Michigan, Ann Arbor, MI; 2) Department of Computational Medicine and Bioinformatics ,University of Michigan, Ann Arbor, Michigan, USA; 3) Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA; 4) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland 20892, USA; 5) Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514, USA; 6) Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA; 7) University of Queensland Diamantina Institute, University of Queensland, Princess Alexandra Hospital, Brisbane, Queensland, Australia; 8) MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge, UK; 9) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, 171 77 Stockholm, Sweden; 10) Divisions of Genetics and Endocrinology and Program in Genomics, Children's Hospital, Boston, Massachusetts 02115, USA; 11) Metabolism Initiative and Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA; 12) Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599, USA; 13) Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 14) Public Health and Gender Studies, Institute of Epidemiology and Preventive Medicine, Regensburg University Medical Center, Regensburg, Germany; 15) Institute for Medical Informatics, Biometry and Epidemiology, University of Duisburg-Essen, Germany; 16) Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK; 17) University of Cambridge Metabolic Research Labs, Institute of Metabolic Science Addenbrooke's Hospital, CB2 OQQ, Cambridge, UK; 18) Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois (CHUV) University Hospital, 1011 Lausanne, Switzerland; 19) Department of Medical Genetics, University of Lausanne, 1005 Lausanne, Switzerland; 20) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK; 21) Department of Cardiovascular Medicine, Genetics and Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan 48109, USA; 22) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 23) Mount Sinai School of Medicine, New York, NY.

Obesity is a heritable condition that affects more than a third of the U.S. population. It can predispose to development of metabolic disease but has few effective treatments. To elucidate the genetic underpinnings of obesity, we conducted the largest meta-analysis of SNP associations with body mass index (BMI), the most common measure of obesity, combining data from up to 339,224 individuals from 125 studies. We confirmed 40 established obesity loci and identified 57 new loci associated with BMI ($P < 5\times10-8$). We used GCTA (Yang 2011) to identify second signals ($P < 5\times10-8$) at 5 associated loci (*FANCL*, *NLRC3/ADCY9*, *GPRC5B/GP2*, *BDNF*, *MC4R*). Polygene analysis indicates that substantial additional heritability is captured by SNPs below the threshold of genome-wide significance. To identify genes and pathways that influence BMI, we use novel methods to integrate eQTL, functional variant, literature connection, gene expression, protein-protein interaction, mouse knockout phenotype, and pathway database information at associated loci. Our data confirm a prominent role for central nervous system regulation of BMI, and extend previous work by newly implicating additional biological processes as regulators of human body mass. These include CNS processes such as synaptic function, cell-cell adhesion and glutamate signaling (*PCDH9*, *CADM2*, *NRXN3*, *NEGR1*, *GRID1*), as well as peptide biology (*GRP*, *SCG3*), lipid metabolism (*NPC1*, *DGKG*), and glucose/insulin action (*RPTOR*, *FOXO3*, *TCF7L2*, *GIPR*, *IRS1*). Of note, one of the proposed mechanisms of topiramate, a component of a recently approved anti-obesity drug is an effect on CNS glutamate signaling, supporting this and other pathways that we identify as possible targets for obesity intervention. Interestingly, whereas most BMI associated loci have effects on related metabolic diseases/traits in expected epidemiological directions, some novel BMI SNPs display unexpected patterns of association with the BMI-increasing allele being protective for disease risk (e.g., *TCF7L2* for type 2 diabetes and *IRS1* for cardiovascular disease; $P < 0.0005$). Such pleiotropy begins to define a shared genetic etiology between BMI and metabolic disease, which may help explain why some but not all obese individuals develop metabolic disease. These results greatly enhance our understanding of BMI biology, open the door for further research into the etiology of obesity and to developing new ways to prevent and treat obesity and its complications.

337

Large-scale association analysis identifies novel loci associated with waist-to-hip ratio and suggests underlying biological mechanisms. *D. Shungin[1,2,3], D.C. Croteau-Chonka[5,6], A.E. Locke[7], T.W. Winkler[4], T. Ferreira[8], R. Magi[8,9], T.H. Pers[10,11,12], A.E. Justice[13], R.J. Strawbridge[14], C. Fox[15], K.E. North[13], E. Speliotes[16,17], I. Heid[4,18], I. Barroso[19,20], R.J. Loos[22], L.A. Cupples[15,23], P.W. Franks[1,2], E. Ingelsson[21], A. Morris[8], K.L. Mohlke[6], C.M. Lindgren[8,17] on behalf of the Genetic Investigation of ANthropometric Traits (GIANT) Consortium.* 1) Department of Clinical Sciences, Skane University Hospital, Lund University, Malmo, Sweden; 2) Department of Public Health and Clinical Medicine, Umea University, Umea, Sweden; 3) Department of Odontology, Umea University, Umea, Sweden; 4) Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany; 5) Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 6) Department of Genetics, University of North Carolina, Chapel Hill, NC; 7) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 8) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 9) Estonian Genome Center, University of Tartu, Estonia; 10) Division of Endocrinology, Children's Hospital Boston, Boston, MA; 11) Metabolism Initiative and Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 12) Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark; 13) Department of Epidemiology and Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, NC; 14) Cardiovascular Genetics and Genomics Group, Karolinska Institutet, Stockholm Sweden; 15) National Heart, Lung, and Blood Institute, Framingham Heart Study, Framingham, MA; 16) Department of Internal Medicine, Division of Gastroenterology, and Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 17) Broad Institute, Cambridge, MA; 18) Genetic Epidemiology, Helmholtz Zentrum Muenchen-German Research Center for Environmental Health, Neuherberg, Germany; 19) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK; 20) University of Cambridge Metabolic Research Labs, Institute of Metabolic Science Addenbrooke's Hospital, Cambridge, UK; 21) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 22) Charles R. Bronfman Institute of Personalized Medicine, Child Health and Development Institute, Department of Preventive medicine, Mount Sinai School of Medicine, New York, NY; 23) Department of Biostatistics, Boston University School of Public Health, Boston, MA.

To study the genetic and biological underpinnings of waist-to-hip ratio (WHR) adjusted for BMI, a measure of adipose tissue distribution associated with increased cardiometabolic risk independent from overall adiposity, we performed sex-combined and sex-stratified fixed effects meta-analyses in 210,087 individuals from 57 GWA studies and 40 studies genotyped on the Metabochip, discovering 35 novel and confirming 14 previously reported loci associated at $P<5\times10^{-8}$ To identify potential causal variants at these loci, we examined expression QTL data in multiple trait-relevant tissues and found that at 21 loci (12 novel and 9 reported) the WHR-associated SNP was either the strongest SNP associated in cis with significant ($P<10^{-5}$) expression of the gene transcript, or that it explained a substantial proportion of the variance in gene transcript levels when conditional analysis was performed. Out of 21 loci, 15 SNPs (9 novel) were associated with expression levels in subcutaneous adipose tissue and 8 SNPs (4 novel) were associated with expression levels in omental adipose tissue To investigate whether the loci associated with WHR might also play pleiotropic roles for other cardiometabolic traits, we tested for enrichment of directionally-consistent associations among all 49 WHR loci with 22 cardiometabolic traits. We found significant enrichment for T2D, fasting insulin adjusted for BMI (FI), and fasting glucose (13, 27, and 9 SNPs with significant ($P<6.5\times10^{-4}$) effects). In unsupervised hierarchical clustering of standardized association statistics for the 49 loci across 22 traits, we identified a cluster of 30 SNPs associated with high-density lipoprotein cholesterol, FI, and triglycerides Finally, to broadly identify biologically relevant WHR genes and pathways, we used the Data-driven Enrichment-Prioritized Integration for Complex Traits method to analyze all SNPs with $P<10^{-5}$ and identified at least one significantly prioritized gene (false-discovery rate [FDR] < 5%) at 49 WHR loci. Further, we identified 781 significant gene sets or pathways (FDR < 5%), the top ones being 'Decreased percent body fat' and 'Abnormal skeleton morphology' This effort from the GIANT consortium further elucidates the genetic architecture of fat distribution, provides biological data that may help identify the functional loci underlying the SNP associations and confirms epidemiological assumptions about the shared genetic underpinnings of related cardiometabolic traits.

338

Chipping Away At The Common Variant Genetics of Age Related Macular Degeneration. *L.G. Fritsche on behalf of The International AMD Genomics Consortium (https://amdgenetics.org/).* Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI.

Age-related macular degeneration (AMD) is a major cause of blindness in the elderly. Disease predisposition is complex and influenced by multiple environmental and genetic factors. To extend understanding of AMD genetics and biology, we set out to examine the association between common and rare genetic variation in a large set of predominantly late-stage AMD cases (N≈25,000) and appropriately age-matched controls (N≈25,000) obtained via international collaboration.

We examined the genome using a custom array, that included ≈250,000 rare/common variants, derived from prior large-scale sequencing efforts, including targeted sequencing of known AMD loci. Also included were ≈250,000 common variants distributed evenly across the genome. Comprehensive genotyping of ≈50,000 individuals is underway at CIDR (http://www.cidr.jhmi.edu/).

Here, we report an initial analysis of common variants (minor allele frequency ≥1%) using currently available genotype data for 12,030 late-stage AMD patients and 11,332 controls of European ancestry. After adjusting for population structure, age and sex, our genome-wide association study confirmed 18 of the 19 AMD loci previously reported by The AMD Gene Consortium (Fritsche *et al.*, NatGenet 2013; $P<10^{-5}$), and identified several independent signals in known loci (near *CFH*, *CFB*/*C2*, *C3*, *APOE*, *LIPC*, and *COL8A1*). Among the more interesting signals is an independently associated coding variant in *TNXB* near *CFB*/*C2*. Our analysis also suggests several novel loci with $P<5\times10^{-8}$, including loci specifically associated with the neovascular form of late-stage AMD, but not the atrophic form. In addition to loci near genes that indicate an overall enrichment of several pathways (e.g. complement cascade, arteriogenesis), we identify loci near genes without known function, such as a locus that maps near a microRNA gene (*MIR4532*).

Pooling resources and effort, our consortium is carrying out the largest ever genetic study of AMD. We identified novel loci and revealed additional signals hidden near some known loci. Careful follow-up will uncover the genetic architecture and biological significance of these findings. Not only will this experiment accelerate the acquisition of knowledge on the underlying disease causes, but also facilitate numerous subanalyses (e.g. genotype-phenotype correlations, risk model evaluation).

339

Genetics and biology of rheumatoid arthritis contribute to drug discovery. *Y. Okada[1,2,3], D. Wu[1,2,3,4], C. Terao[5,6], K. Ikari[7], Y. Kochi[8], K. Ohmura[6], A. Suzuki[8], H. Yamanaka[7], J. Denny[9], J. Greenberg[10], R. Graham[11], M. Brown[12], S. Bae[13], J. Worthington[14,15], L. Padyukov[16], L. Klareskog[16], P. Gregersen[17], P. Visscher[12,18], K. Siminovitch[19,20], R. Plenge[1,2,3], the RACI consortium, the GARNET conrotium.* 1) Division of Rheumatology, Immunology, and Allergy, BWH, Harvard Medical Schoo, Boston, MA; 2) Division of Genetics, BWH, Harvard Medical School, Boston, MA, USA; 3) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 4) Department of Statistics, Harvard University, Cambridge, MA, USA; 5) CGM, Kyoto University Graduate School of Medicine, Kyoto, Japan; 6) Department of Rheumatology and Clinical immunology, Graduate School of Medicine, Kyoto University, Kyoto, Japan; 7) Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Japan; 8) Laboratory for Autoimmune Diseases, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan; 9) Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN, USA; 10) New York University Hospital for Joint Diseases, NewYork, NY, USA; 11) Immunology Biomarkers Group, Genentech, South San Francisco, CA, USA; 12) The University of Queensland Diamantina Institute, Princess Alexandra Hospital, University of Queensland, Brisbane, Australia; 13) Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, South Korea; 14) Arthritis Research UK Epidemiology Unit, Centre for Musculoskeletal Research, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK; 15) National Institute for Health Research, Manchester Musculoskeletal Biomedical Research Unit, Central Manchester University Hospitals National Health Service Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, UK; 16) Rheumatology Unit, Department of Medicine (Solna), Karolinska Institutet, Stockholm, Sweden; 17) The Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, NY, USA; 18) Queensland Brain Institute, The University of Queensland, St Lucia 4072, Brisbane, Australia; 19) Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada; 20) Department of Medicine, University of Toronto, Toronto, Canada.

A major challenge in human genetics is to devise a systematic strategy to integrate disease-associated variants with diverse genomic and biological datasets to provide insight into disease pathogenesis and guide drug discovery. Here, we demonstrate one such strategy for a common autoimmune disease with no known cure, rheumatoid arthritis (RA). We performed a trans-ethnic genome-wide association study (GWAS) in a total of >100,000 subjects of European and Asian ancestries (29,880 RA cases and 73,758 controls). We discovered 42 novel RA risk loci at a genome-wide level of significance ($P < 5\times10^{-8}$), bringing the total to 101.

The common alleles at these RA risk loci reveal: a shared genetic architecture among individuals of European and Asian ancestry; most risk alleles alter gene expression with fewer alleles altering protein structures; two-thirds of loci have pleiotropic effects on other traits, especially disorders of the immune system; an overlap with genes that contribute to human primary immunodeficiency and hematological cancer somatic mutations; and specific cell types (e.g. overlap with H3K4me3 peaks in $CD4^+$ regulatory T cells) and molecular pathways (e.g., T cell, B cell, cytokine signaling) that contribute to RA pathogenesis.

We also demonstrate that RA risk loci are enriched for genes that are the target of approved therapies for RA (e.g., *TNF* and *IL6R*), and further suggest that drugs approved for other indications may be repurposed for the treatment of RA (e.g., CDK4/CDK6 inhibitors used in cancers).

Together, this comprehensive genetic study sheds light on fundamental pathways and genes that contribute to RA pathogenesis, and provides empirical evidence that the genetics of RA can provide important information for drug discovery efforts.

340

The Ashkenazi Jewish Genome. *S. Carmi[1], E. Kochav[1], K. Hui[2], X. Liu[3], J. Xue[1], F. Grady[1], S. Guha[4,5], K. Upadhyay[6], S. Mukherjee[4,5], B.M. Bowen[2], V. Joseph[7], A. Darvasi[8], K. Offit[7], L. Ozelius[9], I. Peter[9], J. Cho[2], H. Ostrer[6], G. Atzmon[6], L. Clark[3], T. Lencz[4,5], I. Pe'er[1,10].* 1) Department of Computer Science, Columbia University, New York, NY; 2) Department of Internal Medicine, Yale School of Medicine, New Haven, CT; 3) Department of Pathology and Cell Biology, Columbia University Medical Center, New York, NY; 4) Center for Psychiatric Neuroscience, The Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, NY; 5) Department of Psychiatry, Division of Research, The Zucker Hillside Hospital Division of the North Shore-Long Island Jewish Health System, Glen Oaks, NY; 6) Department of Genetics, Albert Einstein College of Medicine, Bronx, NY; 7) Department of Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, New York, NY; 8) Department of Genetics, The Institute of Life Sciences, The Hebrew University of Jerusalem, Givat Ram, Jerusalem, Israel; 9) Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY; 10) Center for Computational Biology and Bioinformatics, Columbia University, New York, New York.

Ashkenazi Jews (AJ) number ≈10 million individuals worldwide, mostly in the US and Israel. In accordance with historical records, recent studies showed that AJ are genetically homogeneous with mixed European and Middle-Eastern ancestry and that the AJ population had undergone a severe bottleneck around 800ya followed by an extremely rapid expansion. These characteristics make the AJ population highly attractive for genetic studies. Here, we report the sequencing of 128 complete genomes of healthy AJ individuals. Sequencing was carried out by Complete Genomics to coverage >50x, and achieved 97% call rate, Ti/Tv=2.14, and 99.7% concordance with SNP arrays. Additional cleaning further reduced the number of false positives to just ≈5000, as determined by examining runs-of-homozygosity. We show that our AJ sequencing panel is 3- fold more effective in filtering out benign variants in clinical AJ genomes than a European, non-Jewish panel. Similarly, our AJ panel reduced the inaccuracy of AJ array imputation, for both rare and common alleles, by 10-20%. Inspection of specific genes related to AJ genetic disorders identified known disease mutations as well as dozens of additional risk alleles. Population-genetic comparison of the AJ genomes to 26 Flemish genomes sequenced using the same technology revealed increased heterozygosity and less allele sharing in AJ, in accordance with the AJ admixed nature and partial Middle-Eastern origin. On the other hand, AJ showed more population-specific allele sharing, higher load of deleterious alleles, and a smaller overall projected number of variants, potentially due to the recent bottleneck. Analysis of identical-by-descent segments, which are abundant in AJ and highly informative on recent history, confirmed a recent severe bottleneck of merely ≈300-400 individuals. Using the allele frequency spectrum, which is informative on ancient history, we inferred the time of the Out-of-Africa founder event to be ≈52,000±4000ya, and the fraction of European ancestry in AJ to be ≈55±2%. We also inferred the split between the ancestral Middle-Eastern population and contemporary Europeans to be as recent as ≈11,000±500ya, suggesting the genetic origin of modern-day Europeans is predominantly Neolithic, and much later than the first dated *Homo sapiens* migration into Europe. This result, made possible by our pioneering sequencing of individuals with Middle-Eastern ancestry, resolves a long-standing debate over European origins.

341

Rare variant sharing reveals population histories. *I. Mathieson[1], G. McVean[1,2].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Department of Statistics, University of Oxford, Oxford, United Kingdom.

Sharing of rare variants is highly informative about recent ancestry at a specific locus, and the rarer the variant the more recent the implied ancestry. Genome-wide patterns of rare allele sharing at different frequencies therefore provide a rich description of the shared ancestry of a sample. We show how to use this data to make inference about human history, and compare it to other approaches using identity by descent (IBD) sharing. First we describe a practical method for extracting and using rare variant sharing information. Using a combination of sequence and array genotype data we find shared haplotypes within a sample, and then use the joint distribution of the length of these haplotypes and the number of mutations they carry to infer the shared history of the sample in terms of the distribution of coalescence times. With this approach, we detect much smaller shared haplotypes than IBD-based methods (on the order of hundreds of kb rather than Mb), and thus infer history up to thousands of generations in the past, rather than hundreds, which allows us to investigate relatedness among humans at a worldwide scale. As an illustration, we applied this method to the Phase 1 data release of the 1000 Genomes Project, identified over 3 million shared haplotypes, and fully characterised the distribution of coalescent times between populations. For example, the median age of a haplotype shared between two GBR individuals (GBR-GBR) is 123 generations, and less than 1% of such haplotypes are older than 1000 generations. In contrast, the median age of a GBR-YRI haplotype is 744 generations and 41% are older than 1000 generations. Admixed populations like ASW show distributions of coalescent times consistent with mixtures of the distributions of the admixing populations. Finally, we compare our patterns of haplotype sharing to those generated using IBD sharing, and explain both the practical and conceptual differences between these approaches and the implications for inference.

342

High risk population isolate reveals low frequency variants predisposing to intracranial aneurysms. *M.I. Kurki[1,2], E.I. Gaál[3], J. Kettunen[4,5], T. Lappalainen[6], V. Anttila[4,7,8], F.N.G van 't Hof[9], M. von und zu Fraunberg[1,2], H. Lehto[3], A. Laakso[3], R. Kivisaari[3], T. Koivisto[1], A. Ronkainen[1], J. Rinne[1], L.A.L. Kiemeney[10,11], S.H. Vermeulen[12], M. Kaunisto[4,13], J.G. Eriksson[5,14,15,16], T. Lehtimäki[17], O.T. Raitakari[18,19], V. Salomaa[5], M. Gunel[20], E.T. Dermitzakis[6], Y.M. Ruigrok[9], G.J.E. Rinkel[9], M. Niemelä[3], J. Hernesniemi[3], S. Ripatti[4,5,21], P.I.W. de Bakker[8, 22,23,24], A. Palotie[4,7,8,25], J.E. Jääskeläinen[1,2].* 1) Kuopio University Hospital, Kuopio, Finland; 2) Neurosurgery, Institute of Clinical Medicine, University of Eastern Finland, Kuopio, Finland; 3) Department of Neurosurgery, Helsinki University Central Hospital, Helsinki, Finland; 4) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 5) Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland; 6) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva 1211, Switzerland; 7) Analytical and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 8) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 9) UMC Utrecht Stroke Center, Department of Neurology and Neurosurgery, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, The Netherlands; 10) Department of Urology, Radboud University Nijmegen Medical Centre, The Netherlands; 11) Department for Health Evidence, Radboud University Nijmegen Medical Centre, The Netherlands; 12) Department of Epidemiology, Biostatistics and HTA, Radboud University, Nijmegen Medical Centre, The Netherlands; 13) Folkhälsan Research Centre, Helsinki, Finland; 14) Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland; 15) Department of Internal Medicine, Vasa Central Hospital, Vasa, Finland; 16) Unit of General Practice, Helsinki University Central Hospital, Helsinki, Finland; 17) Department of Clinical Chemistry, Fimlab Laboratories, Tampere University Hospital and University of Tampere, Tampere, Finland; 18) Department of Clinical Physiology and Nuclear Medicine, University of Turku and Turku University Hospital, Turku, Finland; 19) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku and Turku University Central Hospital, Turku, Finland; 20) Department of Neurosurgery, Department of Neurobiology and Department of Genetics, Program on Neurogenetics, Howard Hughes Medical Institute, Yale School of Medicine, New Haven, CT, USA; 21) Hjelt Institute, University of Helsinki, Finland; 22) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America; 23) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 24) Department of Epidemiology, University Medical Center Utrecht, Utrecht, The Netherlands; 25) Department of Human Genetics, The Wellcome Trust Sanger Institute, Cambridge, UK.

Population isolates have the potential to provide a shortcut for identification of low frequency and rare variants associated with common diseases. Here we use the benefits of the isolated population of Finland, a sample enriched for familial cases and dense 1000 Genomes based imputation of GWAs data, to identify low frequency variants associated to intracranial aneurysm disease. About 3% of the population develops saccular intracranial aneurysms (sIAs), a complex trait, with a sporadic and a familial form in 10% of the cases. The familial sIA patients more often have multiple aneurysms, suggesting genetic factors contributing to the aneurysm formation process. Subarachnoid hemorrhage from sIA (sIA-SAH) is a devastating form of stroke. Certain genetic variants are rare throughout Europe but are enriched in the Finns, a population isolate with a small founder population and bottleneck events. As the sIA-SAH incidence in Finland is >2x higher than worldwide, such variants may associate with sIA in the Finnish population. We tested genotyped and 1000 Genomes imputed variants for association in 760 Finnish sIA patients (40% familial cases), and in 2,513 matched controls with case-control status and with the number of sIAs. The most promising loci ($p < 5E-6$) were replicated in an additional 858 Finnish sIA patients and 4,048 controls. The frequencies and effect sizes of the replicated variants were compared to a continental European population using 717 Dutch cases and 3,004 controls. We discovered four new high-risk loci with low frequency lead variants. Three were associated with the case-control status: 2q23.3 (MAF 2.1%, OR 1.89, p 1.42 × 10-8); 5q31.3 (MAF 2.7%, OR 1.66, p 3.17 × 10-8); 6q24.2 (MAF 2.6%, OR 1.87, p 1.87 × 10-11) and one with the number of sIAs: 7p22.1 (MAF 3.3%, RR 1.59, p 6.08 x -9). Two of the loci (5q31.3, 6q24.2) were also replicated in the Dutch sample . The 7p22.1 locus was strongly differentiated, as the lead variant was more frequent in Finland (4.6%) than in the Netherlands (0.3%). Additionally, we replicated a previously inconclusively locus on 2q33.1 in all samples tested (OR 1.27, p 1.87 × 10-12). This study illustrates the utility of population isolates, familial enrichment, dense genotype imputation and alternate phenotyping in search for variants associated with complex diseases.

343

Rare variant association studies: what population genetics models teach us about power and study design. *BM. Neale[1,2], O. Zuk[1,3], E. Hechter[1,4], K. Samocha[1,2], MJ. Daly[1,2], S. Sunyaev[1,5], S. Schaffner[1], E. Lander[1,6].* 1) Broad Institute of MIT and Harvard, Cambridge MA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 3) Toyota Technological Institute, Chicago, IL; 4) Department of Mathematics, UC Berkeley, CA; 5) Division of Genetics, Brigham and Women's Hospital, Boston, MA; 6) Department of Biology, Massachusetts Institute of Technology, Cambridge MA.

Recent advances in sequencing and genotyping enable the assessment of rare variation in humans at unprecedented levels. In particular, Rare Variant Association Studies (RVAS) aim to shed light on the role of rare variants on common diseases. The allele frequency spectrum and impact of rare alleles depend on demographic history. To explore the influence of population demography on the nature of functional genetic variation in genes, we conducted a series of forward simulations and compared the results of these simulations to exome sequencing data of over 20,000 individuals. We demonstrate that population demography does not heavily influence the average number of deleterious mutations per individual. However, bottlenecks such as those that occurred in Finland and Iceland increase the variation in the number of deleterious mutations per gene. This increased variance has clear implications for gene identification and highlights the value of populations that have experienced a recent bottleneck for genetic investigation. We next study the effect of demography on our ability to detect alleles in RVAS. RVAS differ from GWAS as we need to aggregate rare variants for association tests. Rare variants form a heterogeneous group, with different effect sizes, selection coefficients and allele frequencies. We introduce a two-class mixture model for coding rare variants, where variants are assumed to be either completely harmless, or completely essential for protein function. Using this model, we contrast the power to detect association for rare coding variation under a scenario of “perfect information,” restricting the analysis to causal mutations vs. an “imperfect information” scenario, where the set of functional mutations is unknown or predicted at different levels of accuracy. These comparisons aim to understand how best to handle missense mutations in contrast to putative loss of function (LoF) variants, as LoFs are predicted to have a more consistent impact on phenotype. In conclusion, the search for rare functional variants is likely to more challenging for common complex traits than was the case for common variant identification. However, study designs of more extreme selection and the use of populations with more recent bottlenecks can dramatically improve the power to detect significant association. These results have clear implications for the design and analysis of rare variant studies and will inform the next round of genetic investigations.

344

Finnish founding bottleneck leads to excess of damaging loss-of-function variants with medically relevant associations. *E.T. Lim[1,2,3], P. Würtz[4,5], A.S. Havulinna[5], P. Palta[4,6], T. Tukiainen[1,2,3], Sequencing Initiative Suomi (SISu) Project.* 1) Analytic and Translational Genetics Unit, Mass General Hospital, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 3) Department of Genetics, Harvard Medical School, Boston, MA; 4) Institute for Molecular Medicine Finland, University of Helsinki, Finland; 5) Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland; 6) Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

Exome sequencing has made it possible to identify rare variants that are associated with complex diseases. However, these association tests have proven to be challenged by the large sample sizes needed to achieve adequate power to identify alleles with modest effects. There has been recent success in discovering rare (<1% allele frequency) and low-frequency (1-5% allele frequency) variants in various diseases such as Alzheimer’s disease and insulin processing by taking advantage of the genetics of founder populations, such as the Icelandic and Finnish populations. To further develop this strategy, an international collaboration called the SISu project (Sequencing Initiative Suomi), aims to construct near complete genome variant data from up to 200,000 Finnish genomes from the National Biobanks repository (www.nationalbiobanks.fi) and link genome variant data with multiple phenotypes, partially gathered from national health records. The SISu project includes groups from UCLA (NFBC study), University of Michigan (GoT2D study), the Broad Institute (GoT2D), Oxford University (GoT2D), The Wellcome Trust Sanger Institute (migraine, Finrisk), Lund University (GoT2D), National Institute for Health and Welfare (Finland, (Finrisk)) and University of Helsinki. As a part of the SISu project we assessed the potential of the Finnish founder population for studying low-frequency variants in complex diseases by comparing 3000 exome sequences from Finns to outbred non-Finnish Europeans (NFEs) from Germany, Britain and Sweden. We discovered that the bottleneck in Finland has driven the allele frequencies of some extremely rare and deleterious variants in NFEs to much higher allele frequencies (0.5-5%) in Finns. As a result, the average Finn has 1.4x more loss-of-function variants (nonsense and essential splice site) than an average NFE and 2x more rare complete knockouts. This suggests that rare and low-frequency variants with medically important consequences can be detected more easily using smaller sample sizes in Finns. To discover new associations, we applied a reverse genetics approach and genotyped 89 low-frequency loss-of-function variants that are at least 2x enriched across ~35,000 Finns. We report several medically important associations for these low-frequency loss-of-function variants with various complex diseases such as Vitamin B12 deficiency, lipids and blood pressure, emphasizing the utility of the Finnish population for rare variant association studies.

345

A rare functional variant in *APOC3* is associated with lipid traits and has risen in frequency in distinct population isolates. *E. Zeggini[1], G. Dedoussis[2], L. Southam[1,3], A-E. Farmaki[2], G.R.S. Ritchie[1,4], D-K. Xifara[3,5], A. Matchan[1], K. Hatzikotoulas[1], N.W. Rayner[3], Y. Chen[1], C. Kiagiadaki[6], K. Panoutsopoulou[1], J. Schwartzentruber[1], L. Moutsianas[3], E. Tsafantakis[6], C. Tyler-smith[1], G. McVean[3], Y. Xue[1], I. Tachmazidou[1], UK10K Consortium.* 1) Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom; 2) Harokopio University Athens, Athens, Greece; 3) Wellcome Trust Centre for Human Genetics, Oxford, UK; 4) European Bioinformatics Institute, Hinxton, UK; 5) Department of Statistics, Oxford, UK; 6) Anogia Medical Centre, Anogia, Greece.

Population isolates can enhance the power to detect association at low-frequency and rare sequence variation, because of potentially increased allele frequency and extended linkage disequilibrium. We have collected samples from two isolated populations in Greece (HELlenic Isolated Cohorts study, www.helic.org). All samples (n~3000) have information on a wide array of anthropometric, cardiometabolic, biochemical, haematological and diet-related traits. All individuals have been typed on the Illumina OmniExpress and HumanExome Beadchip platforms. In an analysis of HDL levels using exome chip data in 1256 individuals from the HELIC-MANOLIS study, we find striking evidence for association with common-frequency positive control variants in *CETP* (for example, rs1532624, $p=1.1\times10^{-11}$), the well-established HDL-associated locus. Despite the small sample size, we also find genome-wide significant evidence for association with a functional variant in *APOC3* (rs76353203, R19X, $p=4.6\times10^{-9}$), which explains 2.9% of the trait variance. The T allele, which is the derived allele, changes an arginine residue into a stop codon in exon 2 of all 3 coding transcripts of the *APOC3* gene. R19X is also significantly associated with decreased triglyceride levels ($p=1.1\times10^{-11}$) and explains 3.9% of trait variance. This cardioprotective effect is recapitulated in the strong association with high HDL as a dichotomised trait defined as >60mg/dl ($p=4.3\times10^{-10}$). We find that around 3.8% of samples are heterozygous for the mutation. Individuals with the R19X variant do not tend to be closely related, with 97.3% of carrier pairs having ≤10% of alleles identical by descent exome array-wide. R19X has previously been associated with HDL levels in the Amish founder population. The R19X minor allele is carried by 4 out of 3621 whole genome sequenced individuals in the UK10K study (MAF=0.05%). The increased frequency of R19X enables discovery of this signal at genome-wide significance in a small sample size. The sample size needed to detect the effect observed in HELIC at 80% power in an outbred European population is 67,000. *APOC3* R19X constitutes the first known example of a clinically important variant that was previously thought to be private to a population but which has in fact drifted in frequency in two independent population isolates and is strongly associated with traits of high medical relevance.

346

The impact of recent human demography on deleterious mutation load and the genetic architecture of disease susceptibility. *G. Sella[1, 2], Y. Simons[1], M.C. Turchin[3], J.K. Pritchard[3,4].* 1) Department of Ecology, Evolution, and Behavior, Hebrew University of Jerusalem, Jerusalem, Israel; 2) Department of Biological Sciences, Columbia University, New York, United States of America. (as of Oct 1st); 3) Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America; 4) Howard Hughes Medical Institute.

Human populations have undergone dramatic changes in population size in the past 100,000 years, including a severe bottleneck of non-African populations and recent explosive population growth. There is currently great interest in how these demographic events may have affected the burden of deleterious mutations in individuals and the allele frequency spectrum of disease mutations in populations. Here we use population genetic models to show that-contrary to previous conjectures-recent human demography has likely had very little impact on the average burden of deleterious mutations carried by individuals. This prediction is supported by exome sequence data showing that African American and European American individuals carry very similar burdens of damaging mutations. We next consider whether recent population growth has increased the importance of very rare mutations in complex traits. Our analysis predicts that for most classes of disease variants, rare alleles are unlikely to contribute a large fraction of the total genetic variance, and that the impact of recent growth is likely to be modest. However, for diseases that have a direct impact on fitness, strongly deleterious rare mutations likely do play important roles, and the impact of very rare mutations will be far greater as a result of recent growth. In summary, demographic history has dramatically impacted patterns of variation in different human populations, but these changes have likely had little impact on either genetic load or on the importance of rare variants for most complex traits.

347

Inferring ancient demography using whole-genome sequences from multiple individuals. *M. Steinruecken[1], J. Kamm[2], Y. Song[1,2].* 1) EECS, University of California, Berkeley, CA; 2) Statistics, University of California, Berkeley, CA.

Uncovering the demographic history of present-day populations, especially of humans, has received a lot of attention, since knowledge of this history is necessary to correctly interpret the results of association studies, or the action of evolutionary forces that shaped the genetic variation observed today. Quantities of interest include the divergence times of species or subpopulations together with the intensity and duration of subsequent gene flow; as well as the sizes of ancestral populations. Despite significant progress in coalescent theory in the last decades, full-likelihood inference from a sample of DNA sequences under a suitable structured coalescent model is still infeasible. Some existing methods incorporate linkage information, but are limited to a small sample size, prohibiting accurate inference in the recent or very distant past. Other studies used large sample sizes to enable inference about the more recent past, but this prohibited the inclusion of linkage.

To benefit from linkage information and larger sample sizes, we developed a method based on the conditional sampling distribution (CSD). The CSD describes the distribution of an additionally sampled haplotype conditional on having already observed a given set of sequences. Combining ideas from the structured coalescent and the Sequential Markov Coalescent, we devised a Hidden Markov model (HMM) that can be used to efficiently and accurately approximate the true CSD. This approximate CSD can be applied in suitable composite likelihood frameworks to approximate the probability of observing a given set of sequences under a certain demographic scenario. The fact that our model can be cast as an HMM allows for efficient inference of demographic parameters using an Expectation-Maximization approach.

We demonstrate the performance of our inference procedure through extensive simulations. We show that our method can accurately recover biologically relevant demographic parameters like population divergence times, migration rates, or ancestral population sizes from simulated datasets. We apply our method to human genomic sequence data to demonstrate its utility in learning about human demographic history. Applying our CSD in frameworks for phasing genotypes or imputation of missing sequence data would make it possible to account for substructure in the underlying population, thus potentially increasing accuracy.

348

Inferring human population history and gene flow from multiple genome sequences. *S. Schiffels, R. Durbin.* Wellcome Trust Sanger Institute, Cambridge, Cambridgeshire, United Kingdom.

The availability of complete human genome sequences from populations across the world has given rise to new population genetic inference methods such as the Pairwise Sequentially Markovian Coalescent (PSMC). While PSMC infers demographic history for times between 50kya and 2mya with good resolution, neither the more recent evolutionary history nor migration patterns across populations can be addressed. Here we present a new method that overcomes both of these shortcomings. The Multiple Sequentially Markovian Coalescent (MSMC) infers more recent evolutionary history within and across populations up to a few thousand years ago. MSMC models the pattern of mutations in multiple genome sequences under the sequentially Markovian coalescent with recombination. It fits local genealogical trees to the observed pattern, focussing on the first coalescence among any two individuals. We apply our method to genome sequences from several family trios from the 1000 Genomes project with African, European, Asian and Native American ancestry, inferring population sizes and migration rates as a function of time with high resolution. In particular, our method gives information about the separation of non-African populations after the out-of-Africa event, and observes the bottleneck involved in the peopling of the Americas.

349

A haplotype map derived from whole genome low-coverage sequencing of over 25,000 individuals. *J. Marchini on behalf of the Haplotype Consortium.* Department of Statistics, University of Oxford, United Kingdom.

Several whole-genome low-coverage sequencing studies of population and disease cohorts are currently underway or nearing completion. Each of these studies will produce haplotype sets that can be used for downstream imputation into other GWAS cohorts. The goal of the Haplotype Consortium is to combine the data across 18 studies totaling more than 25,000 predominantly European samples, resulting in a reference set of haplotypes that will act as a central resource for imputation and population genetics studies. The large sample size of the combined study will improve the accuracy of the genotype calling and phasing in each individual study. This will limit the need for groups to choose between reference sets or carry out imputation from multiple reference sets and, importantly, should improve imputation of rare variants. There are several challenges in achieving the goals of the Haplotype Consortium. For example, the haplotype sets produced by each individual study are not easily combined due to differences between sequence analysis protocols and resulting lists of polymorphic sites identified in each study. We have investigated strategies for combining haplotype sets using phasing and imputation methods, by combining data at the haplotype or genotype level. In addition, we have investigated approaches that use genotype likelihoods at a union set of polymorphic sites across studies. I will present the results of these comparisons together with our plans for how this resource can be used by the community for imputation based analysis, as well as our future plans for extending the size of this resource to include other ethnicities.

350

Statistical estimation of haplotype sharing from unphased genotype data. *D. Xifara[1], I. Mathieson[1], I. Tachmazidou[2], G. Dedoussis[3], L. Southam[1,2], K. Panoutsopoulou[2], K. Hatzikotoulas[2], E. Zeggini[2], G. McVean[1].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) Wellcome Trust Sanger Institute, Hinxton, UK; 3) Harokopio University Athens, Athens, Greece.

In large-scale population genomic data sets, individual chromosomes are likely to share extended regions of haplotype identity with others in the sample. Patterns of local haplotype sharing are informative about many processes including demography, selection and recombination. However, in outbred diploid populations, the identification of extended shared haplotypes is not straightforward, particularly in the presence of low levels of genotyping error. Here, we introduce a model-based method for detecting extended haplotype sharing between sets of individuals that provides accurate estimates from unphased genotype data. We also describe an implementation of the algorithm that can be applied to data sets consisting of thousands of samples. By applying the method to dense SNP data from 5,144 samples from the UK we show that the median extent of maximal haplotype sharing between unrelated samples is 1.7 cM, implying that even variants at frequencies of 1 in 10,000 within the UK are likely to be over 50 generations old (1,000 - 1,500 years). Moreover, we show that these data are consistent with a model in which explosive growth within the UK dates to 100 generations ago. In contrast, within a Greek population isolate (the MANOLIS cohort; part of the HELIC project) the median extent of maximal haplotype sharing within a sample of 754 'unrelated' individuals is 15 cM, implying approximately 4 generations (80-100 years) until the closest common-ancestor. By assessing the size and geographical distribution of maximal haplotype sharing within and between all cohorts of the HELIC project, we can characterise factors influencing local ancestry and begin to date connections between populations.

351

HapFABIA: Identification of very short segments of identity by descent (IBD) via biclustering. *S. Hochreiter, G. Povysil.* Institute of Bioinformatics, Johannes Kepler University Linz, Linz, Austria.

Identity by descent (IBD) can be detected reliably for long shared DNA segments which are found in related individuals. However, many studies contain cohorts of unrelated individuals that share only short IBD segments. New sequencing technologies facilitate identification of short IBD segments through rare variants which convey more information on IBD than common variants. Current IBD detection methods, however, are not designed to utilize rare variants for the detection of short IBD segments. Short IBD segments reveal genetic structures at high resolution. Therefore, they can help to improve imputation and phasing, to increase genotyping accuracy for low-coverage sequencing, and to increase the power of association studies. Since short IBD segments are further assumed to be old, they can shed light on the evolutionary history of humans. We propose HapFABIA, a computational method that applies a biclustering technique to identify very short IBD segments characterized by rare variants. HapFABIA significantly outperformed competing algorithms at detecting short IBD segments on artificial and simulated data with rare variants. HapFABIA identified t short IBD segments characterized by rare variants with a median length of 25 kbp in data for chromosome 1 from the 1000 Genomes Project. IBD segments that match the Denisovan or the Neandertal genomes (archaic genomes) are either shared by a very low or a very high proportion of Africans. IBD segments that match archaic genomes are enriched at lengths in the ranges of 0 to 12 kbp (about 130 kyr in the past) and 38 to 60 kbp (13 - 20 kyr). IBD segments that match an archaic genome and are of length 0 - 12 kbp are overrepresented in Africans, while those of length 38 - 60 kbp are mainly found in Asians or Europeans. Both the distributions of proportions as well as the IBD segment lengths hint to two events: (1) an admixture of humans and archaic genomes outside of Africa and (2) an admixture of humans and archaic genomes within Africa or survival of ancient DNA segments in the African population.

352

A new method for genotype calling and phasing for the 1000 Genomes Project leads to improved downstream imputation accuracy. *O. Delaneau[1], A. Menelaou[2], J. Marchini[1], The 1000 Genomes Project Consortium.* 1) Department of Statistics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) University Medical Centre Utrecht, Utrecht, Netherlands.

The 1000 Genomes Project has pioneered the use of low coverage sequencing, followed by LD-based genotype refinement, for the construction of comprehensive haplotype reference panels. This approach has become popular in many other studies of population and disease cohorts. We have made clear improvements to this strategy by developing a new LD-based genotype refinement approach. Firstly, we take advantage of genome-wide SNP chip genotypes available on the project samples. We first phase these genotypes using SHAPEIT2 to create a dense haplotype scaffold across the genome. Since a large proportion of the samples are part of trios and duos the haplotype scaffold is very accurate. In the second step, we use the low-coverage sequencing data and phase each novel variant site onto the haplotype scaffold. To do this we have extended the SHAPEIT2 approach to work with low-coverage sequencing data and to accommodate a phased haplotype scaffold. We have applied this approach to the Phase 1 dataset, as well as new, larger sets of project samples. On the Phase 1 dataset we produce genotype callsets with lower error rates than other methods by at least 25%. More importantly, our new haplotype reference panel leads to improved downstream imputation accuracy in GWAS samples. For example, for SNPs with a MAF of 1% we observe an increase of 0.1 on the R2 scale when we compared imputed genotypes to validation genotypes obtained from high-coverage Complete Genomics sequencing data. A key advantage of this scaffold-based approach is that other variant types such as indels, deletions, STRs and multi-allelic variants can also be phased onto the haplotype scaffold in a highly parallel scheme. For example, we have also developed new methods that can phase deletions, which have variable ploidy, and multi-allelic variants, such as small regions containing SNPs and indels, onto the scaffold. Overall, this strategy is being adopted to process the final 1000 Genomes Project release.

353

Identification of genetic epistasis in regulation of gene expression via variance expression quantitative trait loci. *A. Brown[1], A. Buil[2,3,4], M.N. Davies[5], A. Viñuela[5], T. Lappalainen[2,3,4], H.F. Zheng[6], J.B. Richards[5,6], K.S. Small[5], T.D. Spector[5], E.T. Dermitzakis[2,3,4], R. Durbin[1].* 1) Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) Dep. Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 3) Institute of Genetics and Genomics in Geneva, University of Geneva, Geneva, Switzerland; 4) Swiss Institute of Bioinformatics, Switzerland; 5) Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom; 6) Departments of Medicine, Human Genetics, Epidemiology and Biostatistics, McGill University, Montreal, Canada.

Epistasis, in which the effect of a variant on a phenotype is modulated by one or more other variants, is frequently observed in model organisms but has proved difficult to isolate from human GWAS. One issue is the vast search space of all pairwise combinations of SNPs: variance QTL have been suggested as an alternative to an exhaustive search. Variance QTL are SNPs that effect the variance of a trait rather than the mean, and are plausible candidates to be involved in gene by environment interactions (GxE) and epistasis. We use variance QTL affecting gene expression (ve-QTL) to discover epistasis using RNA-seq from lymphoblastoid cell lines (LCLs), subcutaneous adipose tissue, skin and whole blood from ~800 female twins from the TwinsUK cohort. Using this v-eQTL approach, we reduce the search space from a quadratic scan to two linear scans. This space is further reduced by concentrating on the cis-window (1MB around transcription start site). In addition, our twin design allows exploration of variance explained by cis-trans epistasis, as well as an analysis of discordance within monozygotic (MZ) twins pairs to infer presence of GxE. The analysis of 765 LCL samples found variance in expression of 501 genes to be affected by a SNP in the cis window (a v-eQTL). Looking at discordance within MZ twin pairs we found evidence that GxE could contribute to 63% of the v-eQTL. A further scan for SNPs in epistasis with these v-eQTL discovered epistatic interactions affecting expression of 170 genes. We replicated 35 of the epistatic interactions, 30 with the same direction of effect, in a RNA-seq dataset of 465 samples from 1000 Genomes. Moreover, leveraging available sequence data from the 1000 genomes project, we were able to test for possible confounding by eQTL on rare haplotypes. We found that in 10 cases, a secondary eQTL would not explain the interaction. One example of epistasis affecting expression was the gene *TRIT1*, for which we found that 8.5% of variance in expression was explained by an interaction between two loci on the boundaries of two distinct enhancer regions. We have successfully exploited v-eQTL to discover replicated epistasis affecting gene expression. Our study design also suggests some of the variants were involved in GxE. We are currently exploring other tissues (adipose, skin and whole blood), focusing in these cases on GxE as environmental factors are known to play a larger role here than in LCLs.

354

Association and replication of SNP-SNP interactions for hundreds of gene expression phenotypes. *A. Fish, W. Bush.* Center for Human Genetics Research, Vanderbilt University, Nashville, TN.

SNP-SNP interactions are thought to play a major role in the etiology of many complex phenotypes; however, they have not been extensively investigated due to computational and statistical limitations, such as the large number of tests necessary and the difficulty in replicating complex genetic effects. Gene expression is an extremely intricate process, with multiple protein complexes interacting with DNA to facilitate transcription. The role of gene regulation in the etiology of disease been increasingly recognized, as the majority of significant findings from GWAS occur in gene regulatory regions. Correspondingly, SNP-SNP interactions within the cis-regulatory regions may play an important role in disease susceptibility. In this study, we investigated interactions been single nucleotide polymorphisms (SNPs) affecting gene expression in 210 HapMap samples with both genotype information and gene expression data in lymphoblastoid cell lines. We limited our analysis to pairs of SNPs which were previously identified as expression quantitative trait loci (eQTLs) with nominal significance ($p < 0.05$). Amongst these eQTLs with significant main effects, we created all possible pair-wise combinations for each gene, resulting in a total of approximately 30 million tests. We created models containing both additive and dominant components to avoid a systematic bias inherent to standard multiplicative interaction terms, in which deviations from additivity in single SNP effects are falsely identified as significant interaction terms. Using a false discovery rate (FDR) correction level of 0.05, we identified 4,284 significant interactions, which corresponded to 123 distinct genes. Using an additional set of 466 multiethnic HapMap samples, we investigated 596 of the initial 4,284 significant interactions. Of these 596 interactions, 281 replicated using a FDR correction of 0.05. Functional annotation of the specific SNPs was also performed, in order to determine potential biological mechanisms which might be predisposed to gene-gene interactions. These results indicate that non-linear interactions among cis-regulatory variants may be a common component to the architecture of gene expression.

355

Gene-gene Interaction Analysis for Next-generation Sequencing. *J. Zhao[1], Y. Zhu[1], M. Xiong[2].* 1) Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA; 2) Human Genetics Center, the University of Texas School of Public Health, Houston, TX.

Traditional gene-gene interaction analysis was originally designed to test pair-wise genetic interactions among common variants. These methodologies are difficult to apply to rare variants because of prohibitive computational time, large number of tests and low statistical power. Rare variants generated by next-generation sequencing (NGS) pose great challenges for genetic interaction analysis due to the following reasons: (1) the demands in the paradigm of changes in interaction analysis; (2) the severe multiple-testing problems, and (3) the expensive computations. To meet these challenges, here we propose a novel statistical method that shifts the paradigm from interaction between two SNPs to the interaction between two genomic regions. In other words, we treat a gene, instead of a SNP, as the unit of analysis and use functional data analysis techniques as dimensional reduction tools to collectively test interactions between all possible SNP pairs within two genomic regions, including both common and rare variants. Through intensive simulation analyses, we demonstrated that this novel approach has correct type 1 error rates and higher power in detecting genetic interactions compared to other existing methodologies. The proposed statistic was applied to several real NGS datasets of cardiovascular disease, including the Wellcome Trust Case Control Consortium (WTCCC) study, the Framingham Heart Study (FHS), and the NHLBI's Exome Sequencing Project. Of the 27 significantly interacting gene pairs identified in the FHS, 6 interacting pairs were able to be replicated in the WTCCC study and 24 pairs were able to be confirmed in the EOMI study after accounting for multiple testing by Bonferroni correction, indicating that the proposed novel statistic has a great potential in genetic interaction analysis for NGS data.

356

Capturing the geographic and genetic components controlling individual genetic regulation of cardio-metabolic quantitative traits. *Y. Idaghdour[1,2], JP. Goulet[1,2], JC. Grenier[1], E. Gbeha[1], A. Hodgkinson[1], V. Bruat[1], T. de Malliard[1,2], J. Hussin[1,2], E. Hip-Ki[1], P. Awadalla[1,2].* 1) Pediatrics, Sainte-Justine Research Center, Montreal, QC, Canada; 2) CARTaGENE, Sainte-Justine Research Center, Montreal, QC, Canada.

Study of how environment and genome interact to shape gene expression variation is relevant to understanding the genetic architecture of complex phenotypes. To quantify the relative magnitudes of environmental and genetic effects on human transcriptome and determine how do gene expression profiles correlate with cardiometabolic phenotypes, we generated whole-genome genotypic data and whole-blood RNASeq gene expression profiles from 1000 deeply endophenotyped individuals with high and low cardiovascular disease risk scores. These individuals were sampled in three geographic locations from an aging cohort of 20,000 participants in Quebec, Canada. We documented strong effects of the environment on the transcriptome implicating modulation of core circulating cell functions. We detected transcriptional signatures that best correlate with cardiometabolic phenotypes and show how genotypic regulatory variation is a major determinant of these effects. Furthermore, we tested the hypothesis that gene expression traits associated with cardiometabolic phenotypes can be dependent on individual's regulatory genotypes and show examples of gene-environment interactions at the transcriptional level.

357

Identification of a Set of Highly Constrained Genes from Exome Sequencing Data. *K.E. Samocha[1,2], E.B. Robinson[1], B.M. Neale[1,2], M.J. Daly[1,2].* 1) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 2) Program in Medical and Population Genetics and Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA.

With a rising number of studies identifying *de novo* mutations, it is important to be able to evaluate the findings, especially when prioritizing genes for further study. Since a mutation in a gene under evolutionary constraint may be more likely to contribute to disease, we sought to identify a set of such genes based on a large collection of exome sequence data.

We developed a sequence context based model of *de novo* mutations to create per-gene probabilities of mutation. We noticed a high correlation (0.94) between the probability of a synonymous mutation in a gene and the number of rare synonymous variants identified in that same gene first using the NHLBI's Exome Sequencing Project data (evs.gs.washington.edu), then with >25,000 exomes analyzed simultaneously (see abstract by MacArthur et al). We predicted the number of variants that we would expect to see in the dataset and, in order to quantify deviations, created a Z score of the chi-squared difference between observation and expectation for both synonymous and missense variation. While the distribution of these Z scores for the synonymous variants was normal, there is a marked shift in the missense distribution towards having fewer variants than predicted.

We identified a list of excessively constrained genes representing roughly 5% of all genes. This set of genes showed enrichment for entries in the Online Mendelian Inheritance in Man (OMIM) database. Roughly half of the top 41 constrained genes - for which deviation from the expected number of missense variants was significant at $p < 10^{-6}$ - have entries in OMIM with dominant or *de novo* inheritance patterns. By contrast, a set of genes for which the missense variants were very close to expectation ($n = 235$, $-0.05 < Z < 0.05$) had only 9 *de novo* or dominant inheritance entries in OMIM, which was significantly different than the number in the top 41 constrained genes ($p < 10^{-16}$).

This list of constrained genes showed significantly more overlap with genes containing a *de novo* loss of function mutation in both autism and intellectual disability ($p < 0.0001$ for both), but not with those genes with *de novo* loss of function mutations in controls ($p = 0.66$), indicating that this approach can effectively prioritize genes in which mutations can strongly predispose to disease.

358

Utility of a Strategic Next Generation Sequencing Approach to Genomic Diagnosis of Patients with Neurodevelopmental Disorders. *S. Soden[1,2,4], C. Saunders[1,2,3,4], E. Farrow[1,2,4], N. Miller[1,2], L. Smith[1,2,4], D. Dinwiddie[1,2,3,4,5,6], A. Atherton[1,2,4], J. LePichon[1,2,4], B. Heese[1,2,4], A. Abdelmoity[1,2,4], N. Safina[1,2,4], A. Modrcin[1,2,4], L. Willig[1,2,4], S. Kingsmore[1,2,3,4].* 1) Center for Pediatric Genomic Medicine, Children's Mercy Hospital, Kansas City, Missouri 64108, USA; 2) Department of Pediatrics, Children's Mercy Hospital, Kansas City, Missouri 64108, USA; 3) Department of Pathology, Children's Mercy Hospital, Kansas City, Missouri 64108, USA; 4) School of Medicine, University of Missouri-Kansas City, Kansas City, Missouri 64108, USA; 5) Deparment of Pediatrics, University of New Mexico Health Science Center, Albuquerque, New Mexico 87131, USA; 6) Clinical Translational Science Center, University of New Mexico, Albuquerque, New Mexico 87131, USA.

Background: Nearly 50% of patients with neurodevelopmental disorders do not receive an etiologic diagnosis, despite often lengthy and costly evaluations. Next-generation sequencing (NGS) has the potential to identify a molecular diagnosis in such patients, resulting in improved patient care. Objectives: We present a novel approach to whole genome (WGS), whole exome (WES), and customized NGS panel testing of 107 patients with neurodevelopmental disorders at the Children's Mercy Hospital Center for Pediatric Genome Medicine (CPGM). Methods: Under a research protocol, patients with neurodevelopmental disabilities were nominated for NGS by treating neurologists, geneticists, rehabilitation specialists, and neurodevelopmental pediatricians. Nominations were reviewed to designate participants for 1 of 3 tests: WGS, WES, or CPGM-developed customized panel. Test selection was based on a structured analysis of clinical features, prior testing, family history, and availability of parents for enrollment. A trio of novel software applications, developed in-house, was used for analysis: Sign and Symptom Assisted Genome Analysis (SSAGA), Rapid Understanding of Nucleotide variant Effect Software (RUNES) and Variant Integration and Knowledge in Genomes (VIKING). These tools tailor variant results by customizing the list of genes analyzed for each patient based on phenotype. Diagnostic genotypes were confirmed with Sanger sequencing and reported to referring physicians. Results: 107 patients with neurodevelopmental disorders, primarily intellectual disability, epilepsy, weakness and regression, were enrolled. Molecular diagnoses were made in 40 of 107 (37%) subjects enrolled: 3 of 4 tested with WGS, 27 of 56 tested with WES, and 10 of 47 tested with the customized panel. Candidate genes were identified in 9 additional subjects by WES, for which functional testing has been initiated. Among patients who received a diagnosis, average age of symptom onset was 12.9 (range 0-90) months, age at enrollment was 84.8 (2-252) months, and age at genomic diagnosis was 99.5 (10-262) months. Referring physicians reported that diagnosis changed management in 10 cases, namely initiation of new treatments, discontinuation of unnecessary medications, and avoidance of invasive procedures. Conclusion: A strategic NGS approach to genomic diagnosis of patients with neurodevelopmental disabilities increases diagnostic yield, is comparatively rapid, and can impact patient care.

359

De novo mutations in autism spectrum disorders revealed by whole genome sequencing. *Y.H Jiang[1], R. Yuen[2], X. Jin[3,4], M. Wang[4], N. Chen[4], X. Wu[4], J. Ju[4], J. Mei[4], Y. Shi[4], L. Zwaigenbaum[8], M.T Carter[5], C. Chrysler[6], L. Drmic[6], L. Lau[2], D. Mercio[2], A. Thompson[6], M. Uddin[2], B. Thiruvahindrapuram[2], E. Anagnostou[7], S. Walker[2], R. Ring[9], J. Wang[4], C. Lajonchere[9], J. Wang[4,10], A. Shihi[9], P. Szatmari[6], H. Yang[4], G. Dawson[9,11], Y. Li[4], S.W. Scherer[2,12].* 1) Department of Pediatrics, Duke University School of Medicine, Durham, NC; 2) Program in Genetics and Genome Biology, The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario M5G 1L7, Canada; 3) BGI@CHOP, Children's Hospital of Philadelphia, Philadelphia, PA, 19104; 4) BGI-Shenzhen, Bei Shan Road, Yantian, Shenzhen, 518083, China; 5) Division of Clinical and Metabolic Genetics, Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 6) Department of Psychiatry and Behavioural Neurosciences, Offord Centre for Child Studies, McMaster University, Hamilton, Ontario L8S 4K1, Canada; 7) Bloorview Research Institute, University of Toronto, Toronto, ON M4G 1R8, Canada; 8) Department of Pediatrics, University of Alberta, Edmonton, Alberta T5G 0B7, Canada; 9) Autism Speaks, New York, NY, USA; 10) The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health Sciences, and Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Demark; 11) Department of Psychiatry, University of North Carolina at Chapel Hill, NC 27510, USA; 12) McLaughlin Centre, University of Toronto, Toronto, Ontario M5S 1A1, Canada.

Autism Spectrum Disorder (ASD) demonstrates high heritability and, familial clustering and ~4:1 male to female bias, yet the genetic causes are only partially understood, due to extensive clinical and genetic heterogeneity. Whole genome sequencing (WGS) promises added value to identify novel ASD risk genes as well as new mutations in known loci, but an assessment of its full utility in an ASD group has not been performed. We have initiated an international endeavor called the Autism Genome 10K-Project, which aims to sequence the genomes from 10,000 individuals belonging to Autism Speaks Autism Genetic Resource Exchange(AGRE). In a pilot study, we used WGS to examine 32 families with ASD to detect de novo or rare inherited genetic variants predicted to be deleterious (loss-of-function and damaging missense mutations). Among ASD probands, we identified deleterious de novo mutations in 6 of 32 (19%) and X-linked or autosomal inherited alterations in 10 of 32 (31%) families (some had combinations of mutations). The proportion of families identified with such putative mutations was larger than has been previously reported, a yield that is in part due to the more comprehensive and uniform coverage afforded by WGS. Deleterious mutations variants were found in four novel, 9 known, and 8 candidate ASD risk genes. Examples include CAPRIN1 and AFF2 (both linked to FMR1 involved in fragile X syndrome), VIP (involved in social-cognitive deficits), and other genes such as SCN2A and KCNQ2 (also linked to epilepsy), NRXN1, and CHD7, which causes ASD-associated CHARGE syndrome. Taken together, these results suggest that WGS and thorough informatic analyses for de novo and rare-inherited mutations will improve the detection of genetic variants likely associated with ASD or its accompanying clinical symptoms. We have just completed WGS of second phase of 100 ASD trios. The progress of analyzing these families will also be presented.

360

Recurrently mutated genes contribute to the risk for developing sporadic autism spectrum disorder. *B.J. O'Roak[1], E.A. Boyle[1], K.T. Witherspoon[1], B. Martin[1], C. Lee[1], L. Vives[1], E. Karakoc[1], J. Hiatt[1], D.A. Nickerson[1], R. Bernier[2], J. Shendure[1], E.E. Eichler[1,3].* 1) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA; 2) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA; 3) Howard Hughes Medical Institute, Seattle, WA.

Recent exome sequencing of families with simplex or "sporadic" autism spectrum disorder (ASD) has highlighted the importance of *de novo* mutations and led to the discovery of novel autism candidate genes. However, most genes are only disrupted once in >1,000 of these simplex trios (single affected child-two unaffected parents). To address the problem of locus heterogeneity, we have optimized an ultra-low-cost (amortized reagent costs <$1/sample/gene) resequencing method based on modified molecular inversion probes (MIPs) and a statistical framework for evaluating the likelihood of recurrent mutations at individual genes (O'Roak *et al.*, *Science*, 2012). We initially applied this approach to a set of 44 genes, demonstrating that six are recurrently mutated in ~1% (24/2,573) of simplex ASD probands. Here we extend this paradigm further by screening 64 genes (202 kilobases/sample) in >4,000 ASD probands and >2,200 unaffected siblings. These candidate genes were selected based on *de novo* mutations discovered in the exomes of 1,100 probands with ASD or intellectual disability—442 sequenced by us, including unpublished results from an additional 233 ASD families (932 exomes). In this study design, we enriched for genes with the most severe events, network association, chromatin function, recurrence in exome data, or novel candidates from our unpublished data. We excluded genes likely to be highly mutable or shown to carry an excess of rare severe mutations in the Exome Sequencing Project database. We made a number of protocol improvements allowing for enhanced design metrics, molecular tagging (Hiatt *et al.*, *Genome Research*, 2013), and a 35% increase in target capture and evaluation (without increasing reads/sample). On average, >95% of the target regions are uniquely captured 10 or more times, including many regions that are poorly captured using current exome protocols. Preliminary results show severe mutations in ~3% of ASD probands and implicate several novel risk loci, including *CACNA1D*, *NCKAP1*, *SETDB2*, and a second gene in the Down's syndrome critical region. Our data strongly support a major role for recurrently disrupted genes in sporadic ASD, implicate a β-catenin/chromatin remodeling network in its etiology, and provide a model for discovering and rigorously validating bona fide genetic risk loci for sporadic, genetically heterogeneous neuropsychiatric disorders.

361

Identification of Biological Pathways Associated with Phenotypically-defined Subgroups of Autism Spectrum Disorders. *O.J. Veatch[1], B.L. Yaspan[2], M.A. Pericak-Vance[3], J.L. Haines[1].* 1) Ctr Human Gen, Vanderbilt Med Ctr, Nashville, TN; 2) Genentech, Inc., South San Francisco, CA; 3) Hussman Institute for Human Genomics, Miller School of Medicine, Miami, FL.

Autism Spectrum Disorders (ASD) are complex neurodevelopmental disorders with strong evidence for genetic susceptibility. However, effect sizes for implicated loci are small and these loci do not explain the majority of ASD heritability. Difficulty in identifying genetic variation with strong effects may arise from the wide trait variability being explained by underlying genetic heterogeneity. Minimizing phenotypic heterogeneity and applying pathway-based analysis to GWAS data are ways to address these obstacles. We used the Autism Genetic Resource Exchange dataset for our initial modeling. Unsupervised clustering was used to group cases using behavioral and biomarker information, and genetic contributions to cluster assignment were evaluated. We analyzed genome-wide SNP data with the Family-Based Association Test, performing separate analyses based on cluster assignment. Further we used the PARIS pathway analysis program to identify biological pathways of interest. We validated our results using an independent dataset derived from the Autism Genome Project. The phenotypic analysis generated two main clusters based on overall trait severity. We see increased odds for siblings being assigned to the same phenotype cluster ($OR \approx 1.4$, $p < 0.00001$). We also see that cases in a given cluster are more genetically related when compared to the unclustered dataset ($Fst \approx 0.17 \pm 0.26$). We identified 149 genes associated ($p < 0.001$) with the "less severe" ASD cluster and 166 genes associated ($p < 0.001$) with the "more severe" cluster. There is minimal overlap (~7%) when comparing genes associated with different clustering-defined subgroups. Genes associated with the "less severe" cluster relate to immune function, while genes associated with the "more severe" cluster relate to cellular growth, survival, and mobility. We replicated clustering results in the AGP dataset and see again that unique biological mechanisms are implicated when comparing genes associated with either the "more severe" or "less severe" AGP clusters. Our results suggest that meaningful phenotypic subgroup definitions can help clarify the underlying genetic etiology of Autism Spectrum Disorders.

362

***De novo* mutation in the dopamine transporter gene associates dopamine dysfunction with autism spectrum disorder.** *NG. Campbell[1], PJ. Hamilton[1], S. Sharma[2], K. Erreger[2], FH. Herborg[3], C. Saunders[4], AN. Belovich[4], EH. Cook[5], U. Gether[3], HS. Mchaourab[2], HJ. Matthies[2], A. Galli[1,2,4], JS. Sutcliffe[1,2,6].* 1) Center for Molecular Neuroscience, Vanderbilt Univ, Nashville, TN; 2) Mol. Physiol. and Biophysics, Vanderbilt Univ, Nashville, TN; 3) Neuroscience and Pharmacology, Univ. of Copenhagen, Copenhagen, Denamrk; 4) Pharmacology, Vanderbilt Univ, Nashville, TN; 5) Psychiatry, Univ. of Illinois at Chicago, Chicago, IL; 6) Psychiatry, Vanderbilt Univ, Nashville, TN.

Risk for autism spectrum disorder (ASD) is largely genetically determined, however, its etiology is highly complex. Studies focusing on rare DNA copy number variation (CNV) point to *de novo* mutation as one significant class of genetic liability and recent whole exome sequencing studies has supported a role for coding *de novo* mutations (DNMs). Among the first DNMs identified by the NIH ARRA Autism Sequencing Consortium was a missense substitution (T356M) in the *SLC6A3* gene encoding the dopamine (DA) transporter (DAT). DAT is a presynaptic regulator of dopaminergic tone in the central nervous system by mediating the high-affinity re-uptake of synaptically released DA. Here, we report the first functional, structural, and behavioral characterization of an ASD associated DNM in the dopamine transporter. Expression of T356M DAT in CHO cells revealed a near absence of DAT-dependent DA uptake relative to wildtype ($p<0.001$). Patch-clamp experiments pre-loading cells with DA revealed significantly diminished AMPH-induced DA efflux by T356M DAT relative to wildtype ($p<0.01$). Most notably, amperometry experiments revealed that mutant DAT constitutively leaks DA from the cell under basal conditions, contrary to wildtype transporter. In the bacterial homolog leucine transporter, substitution of A289 (the homologous site to T356) with a Met, promotes an outward-facing conformation upon substrate binding, a conformation required for substrate efflux. Lastly, Drosophila containing the T356M knock-in mutation demonstrates significantly increased basal locomotion, a trait associated with DA dysfunction and ASD ($p<0.05$). In conclusion, we have characterized a novel *de novo* mutation in DAT that demonstrates profound functional abnormalities. Given the powerful constitutive efflux of DA and virtual absence of DA uptake activity caused by the mutation, we consider it likely that this variant is a significant ASD risk factor. Taken together with prior association between abnormal DAT function and ADHD, these observations may provide a link between ASD risk and pathophysiology and ADHD, which co-occurs in ~40 percent of subjects with ASD. These studies also more broadly implicate altered regulation of DA homeostasis as a potential mechanism underlying part of the overall liability to ASD.

363

Analysis of synaptic function during neurogenesis and maturation in homogeneous populations of autism-affected GABAergic and glutamatergic neurons. *B.A. DeRosa[1, 2, 3], K.C. Belle[1, 2, 3], J.M. Van Baaren[1, 3], J.M. Lee[1, 3], M.L. Cuccaro[1, 2, 3], J.M. Vance[1, 2, 3], M.A. Pericak-Vance[1, 2, 3], D.M. Dykxhoorn[1, 2, 3].* 1) John P. Hussman Institute for Human Genomics; 2) Dr. John T. Macdonald Foundation Department of Human Genetics; 3) University of Miami Miller Schoold of Medicine, Miami, Fl.

Autism spectrum disorder (ASD) is a heterogeneous neurodevelopmental condition characterized by deficiencies in social interaction, verbal and non-verbal communication, and repetitive stereotyped behaviors. The lack of genetically relevant human disease models of ASD that can be used to study the pathogenic mechanisms that underlie this disorder has significantly impeded the development of novel therapeutic interventions. A number of recent studies have utilized post-mortem brain tissue obtained from individuals affected with ASD to study the biological mechanisms that give rise to the disorder. However, these samples represent a single time point, when the disease is fully manifested, limiting what can be learned about the disease process. With the use of patient-specific induced pluripotent stem cells (iPSCs), it is possible to generate large quantities of neurons with genetic backgrounds that are known to result in ASD. Furthermore, a number of methods are available that enable the differentiation of these iPSCs into the specific neuronal subtypes that are most likely to be affected by ASD-specific genetic backgrounds. The in vitro differentiation of the specific neuronal subtypes closely mimics in vivo neurogenesis and synaptic maturation allowing us to characterize the impact of autism on neurodevelopmental processes. For this purpose, we have generated a panel of iPSC lines from peripheral blood samples obtained from 7 ASD patients. These ASD iPSCs are able to be differentiated into functionally active GABAergic and glutamatergic cortical-like neurons. Furthermore, we have developed a set of lentiviral-based human promoter-driven fluorescent reporters of GABAergic neurons, glutamatergic neurons, and development-stage in neurogenesis. These cell type-specific fluorescent reporters allow the isolation of these specific neuronal cell types from heterogeneous cell populations in culture via fluorescence-activated cell sorting. A comparative analysis of the homogeneous populations of ASD-affected and unaffected control iPSC-derived neurons show abnormalities in expression patterns of genes involved in synaptic function and pathways that include genes previously associated with ASD. The use of ASD-specific iPSC-derived neurons provides a novel approach to understand the complex cellular and molecular processes that govern these disorders opening up new opportunities for the development of novel therapeutic approaches.

364

Disruption of the ASTN2 / TRIM32 locus at chr9q33.1 in gender modulated risk for autism, ADHD and other neurodevelopmental phenotypes. *K. Tammimies[1], A.C. Lionel[1,2], A.K. Vaags[3], J.A. Rosenfeld[4], J.W. Ahn[5], A. Noor[6], C.K. Runke[7], V. Pillalamarri[8], M.T. Carter[6], C. Fagerberg[9], B.R. Lowry[10], M.J. Gazzellone[1,2], R.K.C. Yuen[1], S. Walker[1], B.A. Fernandez[11], D. Tolson[12], D.S. Cobb[12], P.A. Arnold[13], P. Szatmari[14], R. Schachar[15], C.R Marshall[1,2], C. Brasch-Andersen[9], M. Speevak[16], M. Fichera[17,18], C.M. Ogilvie[5], D.J. Stavropoulos[19], Y. Shen[20], J.C. Hodge[7], M.E. Talkowski[8,20,21], S.W. Scherer[1,2].* 1) The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; 2) Department of Molecular Genetics and the McLaughlin Centre, University of Toronto, Toronto, ON, Canada; 3) Department of Anatomic Pathology, Cancer Cytogenetics Laboratory & Alberta Children's Hospital, Calgary, AB, Canada; 4) Signature Genomic Laboratories, Perkin Elmer, Spokane, WA, USA; 5) Cytogenetics Department, Guy's and St Thomas' NHS Foundation Trust, London, UK; 6) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada; 7) Department of Medical Genetics, Mayo Clinic, Rochester, MN, USA; 8) Program in Medical and Population Genetics, Broad Institute of Harvard and M.I.T., Cambridge, MA, USA; 9) Department of Clinical Genetics Odense University Hospital, Odense C, Denmark; 10) Departments of Medical Genetics and Pediatrics, Alberta Children's Hospital, University of Calgary, Calgary, AB, Canada; 11) Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St. John's, NL, Canada; 12) Pediatric genetics Madigan Army Medical Center, Tacoma , WA, USA; 13) Department of Psychiatry and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; 14) Department of Psychiatry and Behavioural Neurosciences, Offord Centre for Child Studies, McMaster University, Hamilton, ON, Canada; 15) Department of Psychiatry and Neurosciences and Mental Health Program, Research Institute, The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada; 16) Genetics Department, Credit Valley Hospital, Mississauga, ON, Canada; 17) Laboratorio Genetica Medica, I.R.C.C.S. Associazione Oasi Maria Santissima, Troina, Italy; 18) Genetica Medica, Università di Catania, Italy; 19) Department of Pediatric Laboratory Medicine, Cytogenetics Laboratory, Hospital for Sick Children, Toronto, ON, Canada; 20) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 21) Departments of Genetics and Neurology, Harvard Medical School, Cambridge, MA, USA.

Genomic studies have begun to uncover the architecture of genetic risk for different neurodevelopment disorders (NDDs) including Autism Spectrum Disorder (ASD). The data have highlighted rare de novo and inherited copy number variants (CNVs) and single nucleotide variants impacting several genes encoding cell-adhesion and scaffolding proteins at the neuronal synapse. Recently, CNVs disrupting ASTN2 and/or TRIM32, a small gene intronic to ASTN2, have been detected by genome-wide studies in a few individuals with different NDDs. The vertebrate-specific astrotactins, ASTN2 together with its closely related paralog, ASTN1, have critical role in neuronal migration during brain development. The genetic findings together with a consistent functional role of the astrotactins necessitate the need to determine the prevalence of mutations in NDDs including ASD. Here, we present a screening of the ASTN2/TRIM32 locus at chr9q33.1 and the ASTN1 locus at chr1q25.2 for CNVs in clinical microarray data from 89,318 individuals, including 63,901 NDD subjects. We identified 45 deletions and 11 duplications affecting the ASTN2/TRIM32 locus. The strongest signal for significant enrichment was found for deletions at the 3' end that disrupted all isoforms of ASTN2 and/or included TRIM32 in the cases with NDDs (p-value = 0.006) compared to population-based controls containing 41,536 individuals. More strikingly a greater penetrance of the deletions was observed in males compared to females. The most common traits observed in 45 individuals with ASTN2 deletions included ASD, ADHD, anxiety, OCD and speech delay. Deletions affecting ASTN1 were rarer in occurrence than those at ASTN2 and were predominantly of de novo origin. In addition, we measured the expression of ASTN2 isoforms in human brain samples showing that the short isoforms transcribed from the significant 3' region were expressed. The spatiotemporal expression profiles of the astrotactins in human brain revealed dynamic expression pattern for ASTN2 in the cerebellar cortex and neocortex and a steady, high level of expression for ASTN1. Future studies are required to investigate the molecular basis of the findings presented here including the gender-biased penetrance and phenotypic heterogeneity of the ASTN2/TRIM32 deletions, the functional relevance of the ASTN2 shorter isoforms and the complex interplay of the astrotactins in brain development and psychopathology.

365

TBC1D24, responsible for early-onset epilepsies associated with intellectual disabilities, plays a role in the formation and maturation of cerebral cortex. *A. Falace[1,5], E. Buhler[3], M. Fadda[5,6], F. Watrin[1,2], P. Lippiello[5], E. Pallesi-Pocachard[4], P. Baldelli[5,6], F. Benfenati[5,6], F. Zara[7], A. Represa[1,2], A. Fassio[5,6], C. Cardoso[1,2].* 1) INSERM U901, Parc Scientifique de Luminy, Marseille, France; 2) Aix-Marseille University, Marseille, France; 3) Postgenomic Platform, INMED, INSERM, Parc Scientifique de Luminy, Marseille, France; 4) Molecular and Cellular Biology Platform, INMED, INSERM, Parc Scientifique de Luminy, Marseille, France;; 5) Department of Experimental Medicine, University of Genova, Italy; 6) Department of Neuroscience and Brain Technology, Italian Institute of Technology, Genova, Italy;; 7) Laboratory of Neurogenetics, Gaslini Insitute, Genova, Italy.

The comorbidity of intellectual disabilities (ID) and epilepsy is well recognized but incompletely understood. Many disorders with obvious structural or metabolic effects on the brain directly cause both ID and epilepsy. In certain syndromes, notably the early-onset epileptic encephalopathies, seizures themselves contribute to cognitive deterioration. Genetic knowledge about early epileptic encephalopathies and idiopathic epilepsy has revolutionized the diagnostic approach to these disorders, and an increasing number of gene mutations have been related to their pathogenesis. TBC1D24 is a novel epilepsy-related gene mutated in rare early-onset forms of epilepsy including idiopathic infantile myoclonic epilepsy and severe, drug resistant early onset epilepsy with ID. Additional mutations in TBC1D24 were recently described in siblings affected by malignant migrating partial seizures in infancy. TBC1D24 is mainly expressed in cerebral cortex and hippocampus. TBC1D24 interacts with ARF6, a main regulator of endocyting recycling at plasma membrane and is involved in neurite outgrowth. To get insight into the role of TBC1D24 in cortical development, we used the in utero RNA interference approach, to knockdown TBC1D24 expression in rat embryonic neuronal progenitor cells. Using two different shRNAs, we knocked-down TBC1D24 expression in rat brains at embryonic day 15.5 (E15.5) and showed a delay in radial migration of neurons into the cortical plate at E20, with most of cells showing morpho-functional abnormalities. We also showed that this abnormal migration pattern is not due to defects in the neuronal differentiation. Concomitant expression of wild-type TBC1D24 and shRNA prevented migration delay, whereas pathogenic variants failed to complement RNAi effect, suggesting a loss-of-function mechanism in patients. Analysis at postnatal stage revealed that TBC1D24-knockdown neurons reached appropriate cortical layers, although featuring significant defects in morphological maturation. Next, we showed that concomitant expression of an inactive form of ARF6 was able to rescue the TBC1D24 knockdown phenotype in vivo suggesting that developmental abnormalities of TBC1D24 knockdown model are primary due to an impairment of ARF6 endocyting trafficking. In conclusion, our data revealed a developmentally regulated role of TBC1D24 in cortical formation suggesting that mutations in TBC1D24 may cause subtle alterations in cortical development leading to epilepsy and ID.

366

Transcriptional consequences of 16p11.2 microdeletion/microduplication syndrome in human lymphoblasts and mouse cortex. *I. Blumenthal[1], A. Ragavendran[1], S. Erdin[1], L. Klei[2], J. Guide[1], M. Stone[1], C. Ernst[1], J. Levin[3], V. Wheeler[1], K. Roeder[4], B. Devlin[2], J.F. Gusella[1,3,5], M.E. Talkowski[1,3,5].* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA; 3) Broad Institute of MIT and Harvard, Cambridge, MA; 4) Department of Statistics, Carnegie Mellon University, Pittsburgh, PA; 5) Department of Neurology, Harvard Medical School, Boston, MA.

Autism spectrum disorder (ASD) is a heterogeneous developmental disorder affecting 1-2% of the population, with recurrent reciprocal copy number variants (CNVs) representing one of the largest known risk factors. We evaluated the global transcriptional consequences of one of these recurrent CNV regions, 16p11.2, which accounts for ~1% of all ASD cases, but also confers risk to diverse phenotypic outcomes despite recurrent breakpoints mediated by non-allelic homologous recombination, suggesting modifying factors that affect penetrance and expressivity. We performed a customized strand-specific RNA sequencing protocol (dUTP) on lymphoblasts from 7 multiplex ASD families (n=35 individuals), each harboring a segregating 16p11.2 CNV and displaying heterogeneity of both genotype (ASD individuals discordant for CNV) and phenotype (subjects with CNV without phenotype). We also sequenced RNA from the cortex of 8 mice with deletion or duplication of the 16p11.2 syntenic region (7qF3) and 8 gender-matched wild type littermates. For genes within the CNV region, we observed a highly significant and consistent effect in both human and mouse datasets associated with reduced expression in deletions and increased expression in duplications. We also detected presumed positional effects in the CNV's proximity as well as dysregulation of genes elsewhere in the genome. In the humans, 27 genes associated with ASD (from SFARIgene and AutismKB) were nominally significant and 9 appeared in a protein interaction network (generated with DAPPLE) connecting them to 3 genes in the CNV. Gene ontology of these results identified pathways important to chromatin structure and remodeling. The mouse cortex data identified 42 ASD associated loci among the nominal results, and a DAPPLE interaction network revealed a sub-network connecting seven of the CNV region genes to 16 ASD loci outside the region. Gene ontology analysis of the sub-network highlighted WNT signaling, MAPK signaling, and melanogenesis. Interestingly, there were 126 genes differentially expressed in the same direction as a consequence of both deletion and duplication. Our findings reveal interconnected networks of genes whose expression is disrupted by 16p11.2 dosage imbalance, suggesting potential convergence on common ASD pathways due to different mutational mechanisms, as well as a potential common mechanism of ASD risk conferred by both deletion and duplication through genes that require tight dosage control.

367

A homozygous mutation in Smoothened, a member of the Sonic Hedgehog (SHH)-GLI pathway is involved in human syndromic atrioventricular septal defect. *W.S. Kerstjens-Frederikse[1], Y. Sribudiani[2], M.E. Baardman[1], L.M.A. Van Unen[2], R. Brouwer[2], M. van den Hout[2], C. Kockx[2], W. Van IJcken[2], A.J. Van Essen[1], P.A. Van Der Zwaag[1], G.J. Du Marchie Sarvaas[3], R.M.F. Berger[3], F.W. Verheijen[2], R.M.W. Hofstra[2].* 1) Dept Gen, Univ of Groningen, Univ Med Ctr Groningen, Netherlands; 2) Dept Gen, Erasmus Med Ctr, Rotterdam, Netherlands; 3) Dept Ped Cardiol, Univ of Groningen, Univ Med Ctr Groningen, Netherlands.

Introduction: Atrioventricular septal defect (AVSD) is a common congenital heart disease with a high impact on personal health. It is often accompanied by other congenital anomalies and in many of these syndromic AVSDs, defects in the sonic hedgehog (SHH)-GLI signalling pathway have been detected. SMO codes for the transmembrane protein smoothened (SMO), which is active in cells with a primary cilium and is located on the ciliary membrane. SMO is a key protein in the SHH-GLI signaling cascade. Methods: Two probands, a twin boy and girl, presented with an AVSD, large fontanel, postaxial polydactyly and skin syndactyly of the second and third toes of both feet. The boy also had hypospadias. The parents were consanguineous and they had one healthy older child. Karyotyping was normal and Smith-Lemli-Opitz syndrome (SLOS) was excluded. Exome sequencing was performed and candidate variants were validated by Sanger sequencing. Results: A novel homozygous missense mutation c.1725C>T (p.R575W) in SMO (7q32.3) was detected. Functional studies in fibroblasts of the patients showed normal expression of SMO protein but an abnormal localization of SMO, outside the cilia. Moreover we show severely reduced downstream GLI1 mRNA expression after stimulation with the SMO agonist purmorphamine. These results, together with the previously described association of SHH signalling defects with AVSD and SLOS, suggest that this SMO mutation is involved in syndromic AVSD in these patients. Conclusion: We present the first reported smoothened mutation in humans, in two patients with an AVSD and a phenotype resembling Smith-Lemli-Opitz syndrome.

368

Identification of *PRDM16* as a disease gene for left ventricular noncompaction and the efficient generation of a personalized disease model in zebrafish. *A.-K. Arndt[1,2], S. Schaefer[3], R. Siebert[4], S.A. Cook[5], H.-H. Kramer[2], S. Klaassen[6], C.A. MacRae[1].* 1) Cardiovascular Division, Brigham and Women's Hospital, Boston, MA; 2) Department of Congenital Heart Disease and Pediatric Cardiology, University Hospital of Schleswig-Holstein, Kiel, Germany,; 3) Max-Delbruck-Center for Molecular Medicine, Berlin, Germany; 4) Institute of Human Genetics, University Hospital Schleswig-Holstein, Kiel, Germany; 5) National Heart Centre, Singapore; 6) Department of Pediatric Cardiology, Charité, Berlin, Germany.

Using our own data and publically available array comparative genomic hybridization data, we identified the transcription factor *PRDM16* (PR domain containing 16) as a causal gene for the cardiomyopathy associated with monosomy 1p36, and confirmed its role in individuals with non-syndromic left ventricular noncompaction cardiomyopathy (LVNC) and dilated cardiomyopathy (DCM). In a cohort of 75 non-syndromic patients with LVNC we detected 3 sporadic mutations, including 1 truncation mutant, 1 frameshift null mutation, and a single missense mutant. In addition, in a series of cardiac biopsies from 131 individuals with DCM, we found 5 individuals with 4 previously unreported non-synonymous variants in the coding region of *PRDM16*. None of the *PRDM16* mutations identified were observed in over 6500 controls. *PRDM16* has not previously been associated with cardiovascular disease. Modeling of *PRDM16* haploinsufficiency and a human truncation mutant in zebrafish resulted in impaired cardiomyocyte proliferation with associated physiologic defects in cardiac contractility and cell-cell coupling. Using a phenotype-driven screening approach in the fish, we have identified 5 compounds that are able to rescue the physiologic defects associated with mutant or haploinsufficient *PRDM16*. Notably, all of the compounds had the capacity to restore cardiomyocyte proliferation and to prevent apoptosis in the model. Wildtype zebrafish also demonstrated a significant increase in cardiomyocyte numbers after treatment with the compounds suggesting a pro-proliferative effect of the compounds. In addition, the compounds also rescued the contractile and electrical defects observed in these disease models. These findings underline the importance of personalized disease models for specific pathways, to accelerate the exploration of disease biology and the development of innovative therapeutic approaches.

369

Mutation and copy number variation of FOXC1 causes cerebral small vessel disease. *C.R. French[1], S. Seshadri[2], A.L Destefano[3], M. Fornage[4], D.J. Emery[5], M. Hofker[6], J. Fu[6], A.J. Waskiewicz[7], O.J. Lehmann[1, 8].* 1) Ophthalmology, University of Alberta, Edmonton, AB, Canada; 2) Department of Neurology, Boston University, Boston, MA, U. S. A; 3) School of Public Health, Boston University, Boston, MA, U. S. A; 4) Institute of Molecular Medicine and School of Public Health, University of Texas Health Sciences Center, Houston, TX, U.S.A; 5) Department of Radiology, University of Alberta, Edmonton, AB, Canada; 6) Department of Medical Genetics, University Medical Center Groningen, Groningen, The Netherlands; 7) Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada; 8) Department of Medical Genetics, University of Alberta, Edmonton, AB, Canada.

Cerebral small vessel disease (CSVD) represents a major risk factor for stroke and cognitive decline in the elderly. The ability to readily visualize its microangiopathic features by magnetic resonance imaging provides opportunities for using markers of CSVD to identify novel stroke associated pathways. Using targeted genome-wide association analysis we identified CSVD associated single nucleotide polymorphisms (SNPs) adjacent to the forkhead transcription factor FOXC1, and using eQTL analysis in two independent data sets, demonstrate that such SNP's are associated with FOXC1 expression levels. We further demonstrate, using magnetic resonance imaging, that patients with either FOXC1 mutation or copy number variation exhibit CSVD. These findings, present in patients as young as two years of age and observed with missense and nonsense mutations as well as FOXC1-encompassing segmental deletion and duplication, demonstrate FOXC1 dysfunction induces cerebral small vessel pathology. A causative role for FOXC1 in the development and maintenance of cerebral vasculature is supported by the cerebral hemorrhage generated by morpholino-induced suppression of FOXC1 orthologs in a zebrafish model system. Furthermore, in vivo imaging demonstrates profoundly impaired migration of neural crest cells and their subsequent association with nascent vasculature, a process required for the differentiation of perivascular mural cells. In addition, foxc1 inhibition reduces the expression of pdgfrα, a gene critically required for vascular stability via its role in mural cell recruitment. Taken together, these data support a requirement for Foxc1 in stabilizing newly formed vasculature via recruitment of neural crest derived mural cells, and define a casual role for FOXC1 in cerebrovascular pathology.

372

Genetic influence on LpPLA2 activity at baseline as evaluated in the exome chip-enriched GWAS study among ~13600 patients with chronic coronary artery disease in the STABILITY (STabilisation of Atherosclerotic plaque By Initiation of darapLadIb TherapY) trial. *L. Warren[1], L. Li[1], D. Fraser[1], J. Aponte[1], A. Yeo[2], R. Davies[3], C. Macphee[3], L. Hegg[3], L. Tarka[3], C. Held[4], R. Stewart[5], L. Wallentin[4], H. White[5], M. Nelson[1], D. Waterworth[3].* 1) GlaxoSmithKline, Res Triangle Park, NC; 2) GlaxoSmithKline, Stevenage, UK; 3) GlaxoSmithKline, Upper Merion, Pennsylvania, USA; 4) Uppsala Clinical Research Center, Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 5) 5Green Lane Cardiovascular Service, Auckland Cty Hospital, Auckland, New Zealand.

STABILITY is an ongoing phase III cardiovascular outcomes study that compares the effects of darapladib enteric coated (EC) tablets, 160 mg versus placebo, when added to the standard of care, on the incidence of major adverse cardiovascular events (MACE) in subjects with chronic coronary heart disease (CHD). Blood samples for determination of the LpPLA2 activity level in plasma and for extraction of DNA was obtained at randomization. To identify genetic variants that may predict response to darapladib, we genotyped ~900K common and low frequency coding variations using Illumina OmniExpress GWAS plus exome chip in advance of study completion. Among the 15828 Intent-to-Treat recruited subjects, 13674 (86%) provided informed consent for genetic analysis. Our pharmacogenetic (PGx) analysis group is comprised of subjects from 39 countries on five continents, including 10139 Whites of European heritage, 1682 Asians of East Asian or Japanese heritage, 414 Asians of Central/South Asian heritage, 268 Blacks, 1027 Hispanics and 144 others. Here we report association analysis of baseline levels of LpPLA2 to support future PGx analysis of drug response post trial completion. Among the 911375 variants genotyped, 213540 (23%) were rare (MAF < 0.5%). Our analyses were focused on the drug target, LpPLA2 enzyme activity measured at baseline. GWAS analysis of LpPLA2 activity adjusting for age, gender and top 20 principle component scores identified 58 variants surpassing GWAS-significant threshold (5e-08). Genome-wide stepwise regression analyses identified multiple independent associations from PLA2G7, CELSR2, APOB, KIF6, and APOE, reflecting the dependency of LpPLA2 on LDL-cholesterol levels. Most notably, several low frequency and rare coding variants in PLA2G7 were identified to be strongly associated with LpPLA2 activity. They are V279F (MAF=1.0%, P= 1.7e-108), a previously known association, and four novel associations due to I1317N (MAF=0.05%, P=4.9e-8), Q287X (MAF=0.05%, P=1.6e-7), T278M (MAF=0.02%, P=7.6e-5) and L389S (MAF=0.04%, P=4.3e-4). All these variants had enzyme activity lowering effects and each appeared to be specific to certain ethnicity. Our comprehensive PGx analyses of baseline data has already provided great insight into common and rare coding genetic variants associated with drug target and related traits and this knowledge will be invaluable in facilitating future PGx investigation of darapladib response.

373

Genome-wide association study identifies common and rare genetic variants in caspase-1-related genes that influence IL-18 regulation in patients with Acute Coronary Syndrome. *A. Johansson[1, 2], N. Eriksson[1], E. Hagström[1,3], C. Varenhorst[1,3], A. Åkerblom[1,3], M. Bertilsson[1], T. Axelsson[4], B.J. Barratt[5], R.C. Becker[6], A. Himmelmann[7], S. James[1,3], H.A. Katus[8], G. Steg[9], R.F. Storey[10], A. Syvänen[4], L. Wallentin[1,3], A. Siegbahn[1,11].* 1) Uppsala Clinical Research Center, Uppsala University, Sweden; 2) Department of Immunoloy, Genetics and Pathology, Uppsala University, Sweden; 3) Department of Medical Sciences, Cardiology, Uppsala University, Sweden; 4) Department of Medical Sciences, Molecular Medicine, Science for Life Laboratory, Uppsala University, Sweden; 5) AstraZeneca R& D, Alderley Park, Cheshire, UK; 6) Duke Clinical Research Institute, Duke University Medical Center, Durham, North Carolina, USA; 7) AstraZeneca Research and Development, Mölndal, Sweden; 8) Medizinishe Klinik, Universitätsklinikum Heidelberg, Heidelberg, Germany; 9) INSERM-Unité 698, Paris, France; Assistance Publique-Hôpitaux de Paris, Hôpital Bichat, Paris, France; Université Paris-Diderot, Sorbonne-Paris Cité, Paris, France; 10) Department of Cardiovascular Science, University of Sheffield, Sheffield, UK; 11) Department of Medical Sciences, Clinical Chemistry, Uppsala University, Sweden.

Interleukin 18 (IL-18) levels are increased in patients with acute coronary syndromes (ACS) and correlated with myocardial injury. We performed a genome-wide association study (GWAS) to identify genetic determinants of IL-18 levels in patients with ACS. In the PLATelet inhibition and patient Outcomes (PLATO) trial, enrolling a broad selection of ACS patients, baseline plasma IL-18 levels were measured in 16633 patients. Of these, 9340 were successfully genotyped using Illumina HumanOmni2.5 or HumanOmniExpressExome BeadChip and SNPs imputed using 1000 Genomes Phase I integrated variant set. Seven independent associations, in five chromosomal regions, were identified. The first region, with two independent (r^2 = 0.11) association signals (rs34649619, p = $1.17*10^{-50}$ and rs360718, p = $2.03*10^{-12}$), is located within *IL18*. Both top SNPs are located in predicted promoter regions, and the insertion polymorphism rs34649619 (T/TA) disrupts a transcription factor binding site for FOXI1, FOXD3 and FOXA2. The second region, also represented by two independent (r^2 = 0.003) association signals (rs385076, p = $6.99*10^{-72}$ and rs149451729, p = $3.79*10^{-16}$), is located in *NLRC4*. While rs385076 overlaps with a regulatory region, rs149451729 is a rare coding variant resulting in an amino acid substitution, predicted to be deleterious. The third region is located upstream of *CARD16, CARD17*, and *CARD18* and one of the top SNPs (rs17103763, p = $6.19*10^{-9}$) has previously been associated with expression levels of *CARD16*. The two remaining chromosomal regions are located within *GSFMF/MROH6* (rs2290414, p = $5.66*10^{-17}$) and *RAD17* (rs17229943, p = $5.00*10^{-12}$). While the latter genes have not been associated with IL-18 production previously, others are known to be involved in IL-18 release. NLRC4 is an inflammasome that activates the inflammatory cascade in the presence of bacterial molecules. It recruits and activates procaspase-1, which in its turn is responsible for the maturation of pro-IL-18. *CARD16-18*, also known as *COP1, INCA* and *ICEBERG*, encode caspase inhibitors, known to bind to and prevent procaspase-1 activation. Our results suggest that SNPs in *IL18* and caspase-1-associated genes are important for IL-18 production. By combining the identified SNPs in a Mendelian randomization study, the causal effect of IL-18 on clinical endpoints could be further evaluated in a longitudinal study.

370

Genetic association of common variants with a rare cardiac disease, the Brugada Syndrome, in a multi-centric study. *C. Dina[1,2], J. Barc[3], Y. Mizusawa[3], C.A. Remme[3], J.B. Gourraud[1,2], F. Simonet[1], P.J. Schwartz[4], L. Crotti[4], P. Guicheney[5], A. Leenhardt[6], C. Antzelevitch[7], E. Schulze-Bahr[8], E.R. Behr[9], J. Tfelt-Hansen[10], S. Kaab[11], H. Watanabe[12], M. Horie[13], N. Makita[14], W. Shimizu[15], P. Froguel[16], B. Balkau[17], M. Gessler[18], D. Roden[19], V.M. Christoffels[3], H. Le Marec[1,2], A.A. Wilde[3], V. Probst[1,2], J.J. Schott[1,2], R. Redon[1,2], C.R. Bezzina[3].* 1) Thorx Inst, INSERM UMR 1087, CNRS, Nantes, France; 2) CHU Nantes, l'institut du thorax, Nantes, France; 3) Heart Failure Research Center, Academic Medical Center, Amsterdam, Netherlands; 4) University of Pavia, Pavia, Italy,; 5) Inserm UMR 956, UPMC, Paris, France,; 6) Cardiology Unit, Hôpital Bichat, Assistance Publique-Hôpitaux de Paris, Nantes, France,; 7) Department of Experimental Cardiology, Masonic Medical Research Laboratory, Utica, NY, United States,; 8) Department of Cardiovascular Medicine, University Hospital, Münster, Germany,; 9) Cardiovascular Sciences Research Centre, St George's University, London, United Kingdom,; 10) Laboratory of Molecular Cardiology, University of Copenhagen, Copenhagen, Denmark,; 11) 1Department of Medicine I, Ludwig-Maximilians University, Munich, Germany,; 12) Department of Cardiovascular Biology and Medicine, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan,; 13) Department of Cardiovascular and Respiratory Medicine, Shiga University of Medical Science, Otsu, Japan,; 14) Department of Molecular Physiology, Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, Japan,; 15) Division of Arrhythmia and Electrophysiology, Department of Cardiovascular Medicine, National Cerebral and Cardiovascular Center, Suita, Osaka, Japan,; 16) CNRS UMR 8199, Pasteur Institute, Lille, France,; 17) Inserm UMR 1018, Centre for research in Epidemiology and Population Health, Villejuif, France,; 18) Theodor-Boveri-Institute, University of Wuerzburg, Wuerzburg, Germany,; 19) Department of Medicine and Pharmacology, Vanderbilt University School of Medicine, Nashville, TN, United States.

The Brugada Syndrome (BrS) is considered as a rare Mendelian disorder with autosomal dominant transmission. BrS is associated with an increased risk of sudden cardiac death and specific electrocardiographic features consisting of ST-segment elevation in the right precordial leads. Loss-of-function mutations in SCN5A, encoding the pore-forming subunit of the cardiac sodium channel (Nav1.5), are identified in ~20% of patients. However, studies in families harbouring mutations in SCN5A have demonstrated low disease penetrance and in some instances absence of the familial SCN5A mutation in some affected members. These observations suggest a more complex inheritance model. To identify common genetic factors modulating disease risk, we conducted a genome-wide association study on 312 individuals with BrS and 1115 ancestry-matched controls. Two genomic regions displayed significant association. Both associations were replicated on two independent case/control sets from Europe (598/855) and Japan (208/1016) and a third locus emerged, all three with extremely significant p-values (1.10-14 down to 1.10-68). To our knowledge, this is the first time that several common variants are associated with a rare disease, with very high effect (Osdds-ratio) ranging from 1.58 to 2.55. While two loci displaying association hits had already been shown to influence ECG parameters in the general population, the third one encompasses a transcription factor which had never been related to cardiac arrhythmia. We showed that this factor regulates Nav1.5 channel expression in hearts of homozygous knockout embryos and influence cardiac conduction velocity in adult heterozygous mice. At last, we found that the cumulative effect of the 3 loci on disease susceptibility was unexpectedly large, indicating that common genetic variation may have a strong impact on predisposition to rare disease.

371

Loss-of-Function Mutations in *CECR1*, Encoding Adenosine Deaminase 2,Cause Systemic Vasculopathy with Fever and Early Onset Strokes. *Q. Zhou[1], A. Zavialov[2], M. Boehm[3], J. Chae[1], M. Hershfield[4], R. Sood[5], S. Burgess[6], A. Zavialov[2], D. Chin[1], C. Toro[7], R. Lee[8], M. Quezado[9], A. Ombrello[1], D. Stone[1], I. Aksentijevich[1], D. Kastner[1].* 1) Inflammatory Disease Section, NHGRI, Bethesda, USA; 2) Turku Centre for Biotechnology, University of Turku, Turku, Finland; 3) Laboratory of Cardiovascular Regenerative Medicine, NHLBI, Bethesda, USA; 4) Department of Medicine, Duke University Medical Center, Durham, USA; 5) Zebrafish Core, NHGRI, Bethesda, USA; 6) Developmental Genomics Section, NHGRI, Bethesda, USA; 7) NIH Undiagnosed Diseases Program, NIH, Bethesda, USA; 8) Translational Surgical Pathology Section, NCI, Bethesda, USA; 9) General Surgical Pathology Section, NCI, Bethesda, USA.

We recently observed 5 unrelated patients with fevers, systemic inflammation, livedo reticularis, vasculopathy, and early-onset recurrent ischemic strokes. We performed exome sequencing on affected patients and their unaffected parents. The 5 patients shared 3 missense mutations in *CECR1*, encoding adenosine deaminase 2 (ADA2), with the genotypes A109D/Y453C, Y453C/G47A, G47A/H112Q, R169Q/Y453C, and R169Q/28kb genomic deletion encompassing the 5'UTR and first exon of *CECR1*.All mutations are either novel or present at low frequency (<0.001) in several large databases, consistent with the recessive inheritance. The Y453C mutation was present in 2/13004 alleles in an NHLBI database. Both alleles are found in 2 affected siblings who suffered from late-onset ischemic stroke, indicating that heterozygous mutations in ADA2 might be associated with susceptibility to adult stroke. Computer modeling based on the crystal structure of the human ADA2 suggests that CECR1 mutations either disrupt protein stability or impair ADA2 enzyme activity. All patients had at least a 10-fold reduction in serum and plasma concentrations of ADA2, and reduced ADA2-specific adenosine deaminase activity. Western blots showed a decrease in protein expression in supernatants of cultured patients' cells. ADA2 is homologous to ADA1, which is mutated in some patients with SCID. In contrast to ADA1, ADA2 is expressed predominantly in myeloid cells and is a secreted protein, and its affinity for adenosine is much less than ADA1. Animal models suggest that ADA2 is the prototype for a family of growth factors (ADGFs).Although there is no mouse homolog of *CECR1*, there are 2 zebrafish homologs, *Cecr1a* and *Cecr1b*. Using morpholino technology to knock down the expression of the ADA2 homologs, we observed intracranial hemorrhages in approximately 50% of the zebrafish embryos harboring the knockdown construct, relative to 3% in controls. Immunohistochemical studies of endothelial cells from patients' skin biopsies demonstrate a diffuse systemic vasculopathy characterized by impaired endothelial integrity, endothelial cellular activation, and a perivascular infiltrate of CD8 T-cells and CD163-positive macrophages. ADA2 is not expressed in the endothelial cells. Our data suggest that ADA2 may be necessary for vascular integrity in the developing zebrafish as an endothelial cell-extrinsic growth factor, and that the near absence of functional ADA2 in patients may lead to strokes by a similar mechanism.

374

Prevalence and Predictors of Pneumothorax in Patients with Connective Tissue Disorders Enrolled in the GenTAC (National Registry of Genetically Triggered Thoracic Aortic Aneurysms and Cardiovascular Conditions) Registry. *J.P. Habashi[1], G.L. Oswald[2], K.W. Holmes[1,5], E.M. Reynolds[10], S. LeMaire[3], W. Ravekes[1], N.B. McDonnell[4], C. Maslen[5], R.V. Shohet[6], R.E. Pyeritz[7], R. Devereux[8], D.M. Milewicz[9], H.C. Dietz[2], GenTAC Registry Consortium.* 1) Dept Pediatric Cardiology, Johns Hopkins Univ, Baltimore, MD; 2) Dept. Medical Genetics, Johns Hopkins Univ, Baltimore, MD; 3) Baylor College of Medicine, Houston TX; 4) NIA at Harbor Hospital, Baltimore, MD; 5) Oregon Health & Science University, Portland, OR; 6) Queen's Medical Center, Honolulu, HI; 7) The University of Pennsylvania, Philadelphia, PA; 8) Weill Cornell Medical College of Cornell University, New York NY; 9) University of Texas Medical School at Houston, Houston, TX; 10) University of Maryland, Baltimore, MD.

Spontaneous pneumothorax—described as escape of air into the pleural space surrounding the lung in the absence of traumatic injury—is a rare occurrence in the general population (0.1-0.5%), however is well recognized in Marfan syndrome (MFS)(4-5%). Associations between pneumothorax and other connective tissue disorders (CTDs) are less well recognized. We sought to examine potential associations of pneumothorax with MFS, vascular Ehlers-Danlos syndrome (vEDS) and other CTDs. Phenotypic data were analyzed on all GenTAC patients with confirmed diagnoses of MFS, vEDS, Loeys-Dietz syndrome (LDS), bicuspid aortic valve with aortic enlargement (BAVe) or familial thoracic aortic aneurysm and dissection (FTAAD) to assess the prevalence of pneumothorax and associated features (1918 total pts). Of 695 patients with Ghent criteria-confirmed MFS, 73 had experienced a spontaneous pneumothorax (prevalence 10.5%), higher than reported in the literature. The frequency of pneumothorax in vEDS patients (16/107, 15%) was similar to the frequency in the MFS group. The prevalences of pneumothorax in LDS (4/73, 5.5%), FTAAD (13/237, 5.5%), and BAVe (19/806, 2.4%) were significantly less than that for MFS and vEDS ($p<0.001$), yet greater than reported for the general population. In MFS patients with a pneumothorax, there was a three-fold increase in reported skeletal features of pectus carinatum, pectus excavatum, scoliosis and/or kyphosis compared to those without pneumothorax. Similarly, in vEDS, there was a four-fold increase in pectus carinatum, scoliosis and kyphosis in those patients with a pneumothorax compared to those without pneumothorax. In a subset of patients with self-reported data (n=846), smoking was not associated with increased prevalence of pneumothorax. Gender was not a predictor of pneumothorax in any of the diagnostic categories analyzed despite literature reports of increased prevalence in males. In patients enrolled in the GenTAC registry with a diagnosis of MFS, vEDS, BAVe, FTAAD or LDS, the prevalence of pneumothorax was significantly increased in all CTDs analyzed as compared to the general population. The prevalence of pneumothorax was significantly higher in patients with MFS or vEDS than in the other CTDs. These data suggest that skeletal features may be a predictor for pneumothorax. Patients presenting with a spontaneous pneumothorax should be evaluated for several potential CTDs; such an evaluation could unmask an undiagnosed aortic aneurysm.

375

Surprising clinical lessons from targeted next generation sequencing of thoracic aortic aneurysmal genes. *B. Loeys[1,2], D. Proost[1], G. Vandeweyer[1], S. Salemink[2], M. Kempers[2], G. Oswald[3], H. Dietz[3], G. Mortier[1], L. Van Laer[1].* 1) Center for Medical Genetics, University of Antwerp/ Antwerp University Hospital, Antwerp, Belgium; 2) Department of Genetics, Radboud University Medical Center, Nijmegen, The Netherlands; 3) Mc Kusick Nathans Institute for Genetic Medicine, Johns Hopkins University Hospital, Baltimore, USA.

Thoracic aortic aneurysm/dissection (TAA), an important cause of death in the industrialized world, is genetically heterogeneous and at least 14 causative genes have been identified, accounting for both syndromic and non-syndromic forms. The diagnosis is not always straightforward because a considerable clinical overlap exists between patients with mutations in different genes, and mutations in the same gene cause a wide phenotypic variability. Molecular confirmation of the diagnosis is becoming increasingly important for gene-tailored patient management but consecutive, conventional molecular TAA gene screening is expensive and labor-intensive. To shorten the turn-around-time, to increase mutation-uptake and to reduce the overall cost of molecular testing, we developed a TAA gene panel for next generation sequencing (NGS) of 14 TAA genes (ACTA2, COL3A1, EFEMP2, FBN1, FLNA, MYH11, MYLK, NOTCH1, SKI, SLC2A10, SMAD3, TGFB2, TGFBR1 and TGFBR2). We obtained enrichment with Haloplex technology and performed 2x150 bp paired-end runs on a Miseq sequencer in a series of 57 consecutive TAA patients, both syndromic and non-syndromic. The sensitivity and false positive rate were previously shown to be 100% and 3%, respectively. Applying our NGS approach, we identified a causal mutation in 16 patients (28%). This uptake is really high as on average one molecular study per patient (range 0-6) was performed prior to inclusion in this study. One mutation was found in each of the 6 following genes: ACTA2, COL3A1, TGFBR1, MYLK, SMAD3, SLC2A10 (homozygous); two mutations in NOTCH1 and eight in FBN1. An additional 6 variants of unknown significance were identified: 2 in FLNA, 2 in NOTCH1, 1 in FBN1 and 1 heterozygous in EFEMP2. All variants were confirmed by Sanger sequencing. Remarkably, from the eight FBN1 positive patients, three patients had previously been tested FBN1 negative by certified labs, indicating that the sensitivity of Sanger sequencing is not 100%. Interestingly, in two FBN1 mutation positive patients the clinical diagnosis of Marfan syndrome was unsuspected. Similarly, the clinical diagnosis of vascular Ehlers-Danlos syndrome (COL3A1) had not been made. Finally, the ACTA2 mutation was identified postmortem from paraffin-embedded extracted DNA. We conclude that our NGS approach for TAA genetic testing overcomes the intrinsic hurdles of Sanger sequencing and becomes a powerful tool in the elaboration of clinical phenotypes assigned to different genes.

376

Noninvasive Fetal Trisomy Test (NIFTY): A Large-Scale Clinical Practice in 78,289 cases. *F. Chen, XY. Pan, XC. Li, HJ. Ge, SP. Chen, H. Jiang.* BGI-shenzhen, Shenzhen, Guangdong, China.

Objectives To report the performance of noninvasive fetal trisomy test based on massively parallel sequencing (MPS) from maternal plasma in a large-scale clinical practice in China.

Methods The noninvasive fetal trisomy test (NIFTY) was offered as a prenatal screening test in 49 medical centers. Participants received NIFTY as well as prenatal screening for Down's syndrome based on maternal serum and ultrasound were recruited from December 2010. Performance of NIFTY and prenatal screening was compared, and pregnancies with full karyotyping were used to evaluate sensitivity and specificity of NIFTY in trisomy 21, 18 and 13. Follow-up investigation was accessed after the expected date of confinement to offer a complete understanding.

Results Among 78,289 cases recruited, 53,477 were at high risk for Down's syndrome according to maternal serum and ultrasound screening, while only 510 of which were identified as trisomy 21 in NIFTY. In 2125 cases with full karyotyping, 419 were classified as positive in NIFTY, including 320 cases of trisomy 21, 79 cases of trisomy 18 and 20 cases of trisomy 13, with 98.97% sensitivity and 98.10% specificity for all three trisomies. During follow-up investigation, fetal outcome of 21,839 cases without karyotyping were obtained, providing a consistent result with NIFTY identification.

Conclusions Our large-scale clinical practice proved that MPS-based test is of high sensitivity and specificity in detecting fetal common trisomies. Compared to traditional prenatal screening, this MPS-based test could avoid about 99.05% of invasive prenatal diagnostic procedures.

377

Comparison of three single-cell whole genome amplification (WGA) methods for detection of genomic aberrations by array CGH: A step towards noninvasive prenatal diagnosis using intact fetal cells. *A. Breman[1], W. Bi[1], C.A. Shaw[1], I. Van den Veyver[1,2], C.J. Shaw[1], A. Stubbs[1], M. Withers[1], G. Fruhman[1,2], A. Patel[1], J.R. Lupski[1], A. Beaudet[1].* 1) Dept of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Dept of Obstetrics & Gynecology, Baylor College of Medicine, Houston, TX.

Detection of genomic copy number abnormalities in a single cell using array CGH offers a promising non-invasive alternative for prenatal diagnosis. Since fetal cells are extremely rare in the maternal peripheral blood circulation, the genetic status of the fetus has to be derived from the analysis of a few or even a single fetal cell. We previously showed (PMID 22470934) that array CGH on whole genome amplification (WGA) DNA from a single cell could detect >1 Mb copy number changes using a custom array with focused regions of increased probe coverage. Recently, several new WGA technologies were described. The goal of this study was to compare three WGA methods for their ability to produce optimal array CGH results from single cells using a catalog array. Single cells from lymphoblastoid cell lines carrying various genomic abnormalities were analyzed using genome-wide oligonucleotide-based array CGH analysis with a SurePrint G3 Human CGH 180K Agilent array. Detection of copy number changes was compared across multiple WGA methodologies including PicoPLEX (Rubicon Genomics), Ampli1 (Silicon Biosystems) and Multiple Annealing and Looping Based Amplification Cycles (MALBAC; PMID 23258894). Our previously optimized conditions for single cell array CGH also included a two-step DNA purification method and closely matched pooled reference DNAs. We analyzed the single cell array data from each WGA method for the ability to detect copy number differences for the X chromosome with gender-mismatched cells as well as trisomy 21, DiGeorge syndrome deletion, CMT1A duplication and *MECP2* duplication. We consistently detected dosage difference in sex chromosomes for gender mismatched hybridizations and for chromosome 21 in trisomy 21 cells. The 2.5 Mb DiGeorge syndrome deletion was also detectable in a single cell using all 3 WGA platforms, whereas detection of the 1.5 Mb CMT1A duplication and the 0.6 Mb *MECP2* duplication was dependent on the algorithm used to identify copy number changes. These data suggest that aneuploidy and other genomic imbalances in a single cell can be detected by oligo-based array CGH when the WGA, purification and hybridization conditions are optimized, but that increased probe density may be necessary to consistently detect small copy number changes.

378

Next-generation sequencing based preimplantation genetic testing of 24-chromosome aneuploidy and monogenic disorders. *N.R. Treff[1,2,3], X. Tao[1], A. Fedick[1,2], D. Taylor[1,2,3], K.H. Hong[1,2], E.J. Forman[1,2], R.T. Scott Jr.[1,2].* 1) Reproductive Medicine Associates of New Jersey, Basking Ridge, NJ., USA; 2) Dept. of Obstertics, Gynecology, and Reproductive Sciences, UMDNJ-Robert Wood Johnson Medical School, Basking Ridge, NJ, USA; 3) Dept. of Genetics, Rutgers-The State University of New Jersey, Piscataway, NJ, USA.

Comprehensive chromosome screening (CCS) of preimplantation embryo aneuploidy has recently been shown to improve IVF outcomes in multiple randomized controlled trials. Next-generation sequencing (NGS) may provide a unique opportunity to perform CCS and monogenic disorder testing in parallel with high throughput and reduced costs. However, the unique and major challenge of applying NGS to preimplantation genetic testing (PGT) is maintaining accuracy on the limited quantity of starting material from an embryo, while also controlling costs per case (since multiple embryos are evaluated). This study develops and validates a new cost-effective methodology for NGS-based detection of monogenic disorders and 24-chromosome aneuploidy in the human blastocyst. Phase I of development involved the characterization of 5-cell samples from positive control cell lines in order to mimic the number of cells obtained in a typical trophectoderm biopsy. These samples included a variety of cell lines with known aneuploidy, triploidy, or mutation carrier status. NGS was performed using semiconductor sequencing on either the Ion PGM or Proton with molecular barcoding for analysis of multiple samples per chip. Specific disease related mutations, highly polymorphic SNPs, and repetitive sequences with chromosome specific alignment capability were targeted during PCR based capture. Sample processing and data analysis criteria were established to provide 100% consistency with the known genotypes and karyotypes of each control sample. Phase II involved blinded analysis of trophectoderm biopsies from previously characterized human blastocysts and additional cell line samples. Up to 1010 bases were sequenced with up to 96 samples on an individual P1 chip (4,608 chromosomes in a single run). 80 embryo and 240 cell line samples were evaluated. NGS resulted in 100% consistency with known karyotypes and genotypes. Targeted sequencing operating costs were 1/3rd lower than the most efficient of the currently used methodologies. In conclusion, targeted NGS provides an accurate and high-throughput methodology for preimplantation 24-chromosome aneuploidy screening and monogenic disorder diagnosis. This NGS method also represents a more cost-effective solution for PGT and may soon become a standard in reproductive medicine.

379

Genetic normalization of differentiating aneuploid cleavage stage embryos. *P.R. Brezina[1], R. Ross[2], A.T. Benner[3], R.P. Dicky[4], R. Kaufmann[2], R. Anchan[5], Y. Zhao[6], A. Barker[7], K.J. Tobler[6], G.R. Cutting[8], W.G. Kearns[3,6].* 1) Fertility Associates of Memphis, Memphis, TN; 2) Fort Worth Fertility, Fort Worth, TX; 3) Center for Preimplantation Genetics, LabCorp, Rockville, MD; 4) The Fertility Institute of New Orleans, Mandeville, LA; 5) Department of Obstetrics and Gynecology, Division of Reproductive Endocrinology and Infertility, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 6) Department of Gynecology and Obstetrics, Division of Reproductive Endocrinology and Infertility, Johns Hopkins Medical Institutions, Baltimore, MD; 7) Arizona Center for Fertility Studies, Scottsdale, AZ; 8) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins Medical Institutions, Baltimore, MD.

We determined if aneuploid embryos can undergo genetic normalization during differentiation into blastocysts. Forty patients underwent 40 in vitro fertilization cycles and preimplantation genetic screening secondary to repeat pregnancy loss or unexplained infertility. IRB approval was obtained. Single blastomeres from 325 cleavage stage embryos were biopsied and underwent DNA extraction and amplification by a modified multiple displacement amplification protocol and single nucleotide polymorphism (SNP) microarrays or a random priming DNA amplification protocol and comparative genomic hybridization (aCGH) arrays. All embryos were cultured to Day-5 or Day-6 following oocyte fertilization. If aneuploid embryos differentiated into a blastocyst, they underwent surgery to separate the inner cell mass (ICM) from the trophectoderm (TE). The ICM and TE cells were then confirmed by immunocytochemistry using Oct3/4 and Cdx2 respectively. Following cell type confirmation SNP or aCGH microarrays were performed. Forty-eight percent (156/325) of cleavage stage embryos were euploid. Of these, 70% (109/156) differentiated into blastocysts. Of these patients, 60% (24/40) achieved a clinical pregnancy with fetal cardiac activity. In contrast, 52% (169/325) of cleavage stage embryos were aneuploid and only 39% (66/169) differentiated into a blastocyst. Of these 66 blastocysts, 36% (24/66) remained aneuploid in both the ICM and TE cells, 5% (3/66) of the blastocysts had a euploid ICM and an aneuploid TE and 2% (1/66) of aneuploid blastocysts had a euploid TE but an aneuploid ICM. Remarkably, 58% (38/66) of the cleavage stage embryos with an aneuploid blastomere, that differentiated to the blastocyst stage of development, were euploid for all ICM and TE cells analyzed. No uniparental disomy was observed. These results demonstrate that a substantial fraction of cleavage stage embryos, with identified aneuploidy in a single blastomere, can normalize during differentiation to the blastocyst stage of development. We hypothesize that the aneuploid blastomeres from cleavage stage embryos are marginalized during differentiation. We propose that mosaic aneuploidy in early embryo development and subsequent embryo normalization to form euploid embryos is not an uncommon event in human embryo differentiation.

380

Maternal age dependent loss of SMC1β transcripts in human oocytes. *V. Jobanputra[1], S.K. Nurudeen[2], M. Shirazi[1], R.W. Prosser[2], A. Naini[1], J.K. Kline[3], M.V. Sauer[2], D. Warburton[1,4].* 1) Pathology; 2) Obstetrics and Gynecology; 3) Epidemiology; 4) Genetics and Development, Columbia University, New York, NY.

The pathogenic mechanisms underlying the well-established association of advancing maternal age with increased risk of meiotic nondisjunction are unknown. Studies of human oocytes obtained in assisted reproductive settings suggest that premature chromatid separation may be more common than previously thought. One hypothesis is that age-related meiotic errors are the result of failure of cohesins to maintain the association of chromatids during either the resting period in female meiosis or the completion of MI or MII. Evidence supporting this hypothesis is based mainly on studies in the mouse showing that cohesion binding to chromosomes decreases in oocytes of older mothers and that mutations in cohesins lead to errors in meiosis. We set out to examine this hypothesis in humans by measuring the transcript levels of SMC1B, one of the meiosis-specific cohesins, in human oocytes. Immature oocytes which were not used for intracytoplasmic sperm injection ICSI were collected from women undergoing in vitro fertilization (IVF). A total of 90 oocytes were collected from 67 women (age range 21 to 43 years). We extracted RNA from the oocytes and used quantitative real-time PCR to measure the relative amount of SMC1β mRNA. We classified women in four age groups: < 30years (n=27), 30-34 years (n=26), 35-39 years (n=23) and > 40 years (n=14). SMC1β transcripts were detectable in all oocytes. Maternal age was negatively associated with SMC1β mRNA levels. When comparing the transcript levels by age groups, women over 40 years of age had significantly lower SMC1β transcript levels compared with women < 30 years (p=0.0234). Our results implicate an age-dependent loss in the transcript levels of SMC1β in human oocytes. To our knowledge, this is the first study to demonstrate a relation between maternal age and cohesion level in human oocytes.

381

A clinical algorithm for efficient, high-resolution cytogenomic analysis of uncultured perinatal tissue samples; study of more than 700 cases. *G. Maire[1,2], A. Xuan Tong Yu[1,2], E. Kolomietz[1,2].* 1) Mount Sinai Hospital, Toronto, Ontario, Canada; 2) University of Toronto, Laboratory Medicine and Pathobiology.

Genetic testing for chromosomal abnormalities is an indispensible part in perinatal care, reproductive planning and detection of gestational trophoblastic disease. Standard chromosomes analysis by karyotyping is hampered by high culture failure rate, maternal cells overgrowth, culture artifacts and very low resolution. A two steps algorithm based on DNA testing was developed to resolve these problems. Direct extraction of genomic DNA (without culture) was followed by quantitative fluorescence polymerase chain reaction (QF-PCR) to assess i) the fetal origin by genotyping and ii) the common live birth aneuploidies (trisomy 13, 18, 21), sex chromosome aneusomies and triploidy. Subsequently, specimens with normal QF-PCR profile were tested by array-comparative genomic hybridization (array-CGH). Specimens submitted to the laboratory between July 2011 and May 2013 were collected (727). Data on success rate, causes of failure, abnormalities detected and gestation times were collected. DNA extraction performed on products of conception, skin, cartilage or umbilical cord was successful for 93% of the cases. Abnormalities were detected in 32%, of which 60% were detected by QF-PCR alone. Abnormalities of pathogenic clinical significance were found in 30% while only 2 (0.3%) were of unknown clinical significance. Two CNVs were likely benign and 7 were likely pathogenic. In regards to the type of abnormality, autosomal trisomies represented 60%, monosomy × 10% and triploidy 12%. Also, 13% of the abnormalities were copy number variation, of which 57% would have been missed had karyotyping been done instead of array-CGH (based on a theoretical karyotype resolution of 5Mb). In this study we have shown the higher success rate and diagnostic accuracy of the combination of QF-PCR and array-CGH algorithm over the standard karyotyping. Although performed with a high resolution array-CGH platform, only 0.3% showed the problematic variation of unknown clinical significance, while at least 9% of the pathogenic abnormalities would have been missed by G-banding analysis. Overall, the algorithm proved to be an effective and financially viable alternative to conventional karyotype testing for perinatal tissue samples.

382

Whole genome oligonucleotide-SNP arrays in prenatal diagnosis: advancement in identification of clinically significant chromosomal abnormalities. *T. Sahoo, L.P. Ross, K.A. Kopita, L.W. Mahon, J-C.J. Wang, M. Hemmat, B.T. Wang, F.Z. Boyar, M. Haddadin, M.M. Elnaggar, R. Owen, A. Anguiano.* Quest Diagnostics Nichols Institute, San Juan Capistrano, CA.

Identification of chromosomal abnormalities has been significantly enhanced by the utilization of microarray-based technologies. Chromosomal microarray-based (CMA) testing of postnatal samples for detection of pathogenic copy number alterations in patients with developmental delay, congenital anomalies, autism and intellectual disability has proven invaluable. The introduction of CMA for prenatal diagnosis has been more recent. Since the implementation of CMA for prenatal testing at Quest Diagnostics, over 1000 samples have undergone testing. Of the 1000 samples being reported here, 350 were tested utilizing a targeted BAC arrayCGH platform and the remaining 650 with Affymetrix 6.0 or the Cytoscan HD platform. Abnormal ultrasound findings were the predominant clinical indication (67% of cases). Clinically significant copy number alterations (CNAs) were reported in 92 cases (9.2%), findings of unclear clinical significance in 94 (9.4%) and 814 (81.4%) cases with no reportable copy number alterations. Significantly, 60 of the 92 cases (65%) with clinically significant CNA had an abnormal chromosome karyotype and 26 (28%) had normal karyotype analysis; in contrast only 8 of the 94 with CMA findings of unclear significance had an abnormal karyotype. Clinically significant CMA findings, in addition to confirming whole chromosome aneuploidies, provided accurate and explicit identification of marker chromosomes (11% cases), accurate characterization of unbalanced translocations (17% cases), identification and structure of submicroscopic deletions and duplications including complex mosaic abnormalities (20% cases) and, most significantly, enabled identification of recurrent pathogenic microdeletion-microduplication syndromes in 19 cases (21%; 19 of the 26 cases with a normal karyotype analysis). Four of the 94 cases with findings of unclear significance harbored multiple segments of homozygosity suggestive of parental relatedness and predicting an increased risk for recessive Mendelian disorders. Therefore, implementation of CMA as a primary diagnostic tool in prenatal testing has profoundly improved our ability to detect clinically significant alterations, particularly in cases with abnormal ultrasound findings. Our experience confirms and incrementally improves upon the findings of other laboratories, lending further support to CMA as a first-tier test for prenatal diagnosis in high-risk pregnancies, especially those with abnormal ultrasound results.

383

The fetal *FMR1* premutation phenotype: clues from the amniotic fluid transcriptome. *L.M. Zwemer[1], S.L. Nolin[2], P. Okamoto[3], M. Eisenberg[4], D.W. Bianchi[1].* 1) Mother Infant Research Institute, Tufts Medical Center, Boston, MA, USA; 2) New York State Institute for Basic Research in Developmental Disabilities, Staten Island, NY, USA; 3) Integrated Genetics/Laboratory Corporation of America® Holdings, Westborough, MA, USA; 4) Laboratory Corporation of America® Holdings, Research Triangle Park, NC, USA.

Second trimester amniotic fluid supernatant (AFS) cell-free fetal (cff) RNA is a pure source of functional and developmental information about many organs, including the fetal brain. Prior studies have demonstrated transcriptome-wide differences in gene expression between unaffected and affected fetuses for both genetic and non-genetic diseases. Trinucleotide repeat expansion in the *FMR1* (fragile X mental retardation 1) gene results in a diagnosis of either a premutation (55-200 CGG repeats) or full mutation (>200 repeats). Little is known about the human fetal phenotype for either full mutation or premutation alleles; here we focused on fetuses diagnosed with the premutation allele, which is frequently associated with adult-onset diseases. Residual AFS from six pairs (four male, two female) of age- and gender-matched premutation and unaffected fetuses was collected after diagnostic testing for fragile X syndrome. cff RNA was extracted and prepared for hybridization to the Affymetrix® HG-U133 Plus 2.0 GeneChip® expression microarray. Seven hundred fifty-seven (757) probe sets representing 587 genes were significantly differentially expressed consistently in the six pairs (two-sided paired t-test with Benjamini-Hochberg (BH) adjusted $p<0.05$). This includes several genes specifically expressed in the testis, the immune, and central nervous systems, as defined by the BioGPS GeneAtlas® database. Manual curation of the top differentially expressed genes using Online Mendelian Inheritance in Man® identified several candidate genes of interest, including: *CDKL5*, *CDNF*, and *CPLX2*, which have all previously been implicated in neurological dysregulation. Among them, *CDKL5* is involved in multiple disorders that map to the X-chromosome or are involved in intellectual disability. Functional analysis using Ingenuity Pathway Analysis® showed that several pathways of interest were enriched at a false discovery rate of <0.25 (right-tailed Fisher's exact test with BH correction) including: cell-mediated immune response, nervous system and endocrine system development. This pilot study provides novel transcriptomic evidence to suggest that a molecular phenotype already exists in fetuses that are carriers of the *FMR1* premutation allele, including dysregulation of several candidate genes. The differential regulation of tissue-specific genes is consistent with the later involvement of these organ systems in adult-onset diseases associated with the *FMR1* premutation.

384

The wide spectrum of alpha and beta-Tubulinopathies in foetus : From microlissencephaly to asymmetrical multifocal polymicrogyria. *N. Bahi-Buisson[1], K. Poirier[2], C. Fallet-Bianco[3], Y. Saillour[2], S. Valence[1], N. Lebrun[2], M. Ossando[4], F. Razavi[4], T. Attie Bittach[4], F. Guimot[5], S. Blesson[6], B. Doray[7], B. Lhermitte[7], E. Andrini[8], P.S. Jouk[8], C. Rouleau[9], M.C. Addor[9], F. Jossic[10], P. Marcorelles[11], L. Loeuillet[12], A. Gelot[13], A. Laquerriere[14], L. Pinson[15], P. Loget[16], F. Chapon[17], P. Dias[18], N. Revencu[19], F.J. Fourniol[20], C. Beldjord[21], J. Chelly[2].* 1) INSERM U781, Institut IHU Imagine, Paris, France; 2) INSERM U1016 Institut Cochin, Paris France; 3) Neuropathologie et Foetopathologie GH Cochin Port Royal; 4) Foetopathologie GH Necker Enfants Malades; 5) Foetopathologie Hopital Robert Debré; 6) Clinical Genetics CHU Tours; 7) Clinical Genetics CHU Strasbourg; 8) Clinical Genetics CHU Grenoble; 9) Clinical Genetics, Lausanne, Switzerland; 10) Clinical Genetics CHU Nantes; 11) Clinical Genetics, CHU Brest; 12) Foetopathologie Hopital Poissy Saint Germain; 13) Foetopathologie Hopital Trousseau; 14) Foetopathologie, CHU Rouen; 15) Clinical Genetics, CHU Montpellier; 16) Clinical Genetics CHU Rennes; 17) CHU Caen; 18) Clinical Genetics, Lisboa, Portugal; 19) Clinical Genetics CHU Louvain, Belgium; 20) Institute of Structural Molecular Biology, Birkbeck College, London; 21) Molecular Biology, GH Cochin, Paris.

Complex cortical malformations associated with mutations in tubulin genes commonly referred to as Tubulinopathies are a heterogeneous group of conditions. Reported phenotypes range from microlissencephaly (MicroLIS) usually diagnosed antenatally to relatively milder with simplified gyral pattern. To further define the phenotypes of fetal tubulinopathies, we studied a cohort of 60 fœtal cases. After sequencing the four tubulin genes, a causal mutation was found in 24/60 of cases interrupted between 16 and 37.8 weeks of gestation; TUBA1A mutations in 17 cases, TUBB2B in six and TUBB3 in a single case. On the basis of a ratio of severity, three subtypes clearly emerged. MicroLIS with corpus callosum agenesis and pontocerebellar hypoplasia represent the most important group (n=11). Cortical lamination is always abnormal either absent cortical plaque (3), reduced to 2-layered cortex (5) or poorly differentiated with 4-layered individualized (3). Neuronoglial overmigration over the pial membrane are occasional (4/11) in particular in the most severe forms and 7/11 showed ectopic neurons within the white matter. All cases demonstrated voluminous enlarged germinal zones. 8/11 had TUBA1A mutations, 2 had TUBB2B and one TUBB3 mutation. Lissencephaly either classic (6/8) or with cerebellar hypoplasia (2/8) are corpus callosum agenesis was the second group. None were microcephalic. Abnormal lamination consisted of either 4 layered cortex with large band of neurons that failed to migrated (3/6) or poorly differentiated cortex (3/6). Neuroglial overmigrations were never observed while ectopic neurons were constant. Brainstem and cerebellum were mildly hypoplasic for all cases except two cases. Polymicrogyria (5) was the third group consisting of focal or multifocal polymicrogyria. None were microcephalic. The majority (4/5) had corpus callosum agenesis and two cases showed pontocerebellar hypoplasia. Two cases had TUBA1A mutations and 3 had TUBB2B mutation. Compared with the phenotypes of children with tubulinopathies, these prenatally diagnosed fetal cases represent the severe end of the tubulinopathies spectrum. The TUBA1A tubulinopathies is mostly represented by microLIS or classic LIS. By contrast, the TUBB2B tubulinopathies is either a PMG usually asymetric or microLIS or LIS. This study emphasizes the importance of neuropathological examinations in diffuse cortical dysgenesis for improving our knowledge of the distinct pathogenetic mechanisms.

385

Practical assessment of incidental finding recommendations for use in clinical exome testing. *M.C. Dulik[1], E.T. DeChene[1], L.K. Conlin[1], S. Mulchandani[1], A. Santani[1], J.L. Abrudan[1], M.J. Italia[2], M. Sarmady[2], J.C. Perin[2], B. Bernhardt[3], C.A. Stolle[1], R.E. Pyeritz[3], A. Wilkens[4], S.E. Noon[4], P.S. White[2,4], I.D. Krantz[4], N.B. Spinner[1].* 1) Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Center for Biomedical Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

The utilization of next generation sequencing (NGS) for diagnostics is leading to increased identification of incidental findings (IFs). IFs are clinically relevant results not related to the reason for the clinical test. Identification of IFs may significantly impact patient care if the findings lead to early diagnosis and improved treatment or management. As part of the NHGRI-funded Clinical Sequencing Exploratory Research Program, the Pediatric Genetic Sequencing (PediSeq) Project at The Children's Hospital of Philadelphia and the University of Pennsylvania is identifying best practices for introducing exome and genome sequencing into pediatric clinical diagnostics. In addition to ascertaining the molecular cause of the patient's primary findings, we are exploring various methods to effectively and efficiently identify IFs, which we have categorized into immediately medically actionable (IMA), childhood or adult onset medically actionable and carrier status. With the recent release of the American College of Medical Genetics and Genomics (ACMG) recommendations for reporting IFs in clinical exome and genome sequencing, we compared the results obtained using the 57 ACMG recommended genes with results obtained from our PediSeq IF analysis of 1953 disease genes with available clinical sequencing tests. Samples were processed for exome capture and sequencing, followed by alignment and variant calling using standard methods. Filtering strategies were applied to the variant calling files using both the PediSeq and ACMG IF gene lists. Most variants were present in control databases and not considered pathogenic (~93% of variants identified in the ACMG genes and ~77% of variants identified with the PediSeq IF framework). Minor allele frequency cut offs were established by identifying well-known, common pathogenic mutations. We identified an average of three potentially pathogenic variants per individual (range: 1-6) for follow up interpretation using the ACMG recommendations, and an average of 77 variants (range: 69-84) with the PediSeq gene list. Although the smaller number of variants identified using the ACMG gene list lessens the overall burden of interpreting IFs, we suggest that only returning IFs from these genes would limit the amount of potentially useful information available to families to a restricted, though important, set of conditions.

386

Individual expectations for return of secondary results from exome sequencing. *H.K. Tabor[1,2], J. Crouch[1], A.A. Lemke[1], K.M. Dent[3], A.G. Shankar[2], S.M. Jamal[2], J.H. Yu[2], M.J. Bamshad[2,4].* 1) Treuman Katz Ctr Pediatric Bioethics, Seattle Children's Hospital, Seattle, WA; 2) Department of Pediatrics, University of Washington, Seattle, WA; 3) Department of Pediatrics, University of Utah, Salt Lake City, UT; 4) Department of Genome Sciences, University of Washington, Seattle, WA.

A major obstacle to taking full advantage of exome and genome sequencing (ES/WGS) data in a clinical setting is how to effectively manage incidental or secondary results. A key step in overcoming this obstacle is understanding the expectations for return of results of people who undergo ES/WGS and how they evaluate possible risks and benefits. To this end, we conducted semi-structured interviews with 42 individuals who had undergone ES in studies of Mendelian and complex diseases and who were offered secondary results. The vast majority (95%) of individuals wanted unrestricted access to all available secondary results. They sought to prepare for the future and thought that all results including the bad, unexpected, uncertain and non-actionable, have a fundamental potential for benefit. While most individuals recognized the psychosocial risks of receiving results, including anxiety, worry and fear, they stated that they would "work through it" and that the benefits of genetic health information far outweighed any potential harms. The majority saw themselves as resilient and able to adapt to volumes of complex, uncertain, and unpleasant information, drawing on their prior experiences receiving unexpected and uncertain medical information. They described the potential benefits of results to include medically actionable information, knowledge about their/their child's disease or prognosis, ability to plan for the future and improve quality of life, and reassurance about negative results. Individuals also wanted the ability to learn more about the range of available secondary results in order to make more nuanced choices and prioritize their preferences for return. Potential challenges they acknowledged included risks to privacy and insurance, temporary anxiety, sharing results with family members, and possible need for further health care interventions. These results suggest that individuals want unfettered access to secondary results from ES and that the perceived benefits of these results are highly contextualized. Furthermore, "all or none" approaches to return of secondary ES/WGS results based on traditional models of actionability appear to be at odds with the demands of healthcare consumers. Accordingly, our results underscore the need to develop and test innovative approaches to ES/WGS results management that allow for greater autonomy and choice over preferences for return of secondary results.

387

The benefits and risks of wanting it all: how parents plan to manage their children's exome sequencing results. *J. Yu[1], J. Crouch[2], A.A. Lemke[2], A.G. Shankar[1], K.M. Dent[3], M.J. McMillin[1], S.M. Jamal[1], M.J. Bamshad[1,4,5], H.K. Tabor[1,2].* 1) Department of Pediatrics, University of Washington, Seattle, WA; 2) Seattle Children's Research Institute, Seattle, WA; 3) Department of Pediatrics, University of Utah, Salt Lake City, UT; 4) Seattle Children's Hospital, Seattle, WA; 5) Department of Genome Sciences, University of Washington, Seattle, WA.

Increasing clinical and research use of exome sequencing (ES) and whole genome sequencing (WGS) make it possible to identify all known pathogenic and risk variants in a child, including those for adult onset conditions. This raises questions about how to manage these results, including which information parents want, how they would like to manage it, what they plan to use it for, and whether they plan to share the information with their children. We conducted semi-structured interviews with 26 parents (one parent per family) of children who had undergone ES as a part of research studies of both Mendelian and complex conditions, and who were enrolled in an ongoing study on return of secondary results. The majority of parents (92%) wanted all available secondary results. Most parents expressed awareness of the potential negative psychosocial impacts of receiving "bad" or unexpected news and the uncertain meaning of some genetic information. They reasoned that access to potentially informative results trumped these risks and that receiving such results was "part of being a good parent." Specific motivations for receiving secondary results included mitigating future disease risks through health care or lifestyle changes, managing potential adverse reactions to medications, planning for future disease, even if no intervention was available, and planning for their children's future quality of life and social services. Most parents also planned to use secondary results for reproductive planning for themselves, their affected and healthy children, and extended family members. Parents planned to disclose positive results to their children during childhood, but deferred disclosure of results about adult-onset conditions to protect their children from psychosocial harms and preserve their autonomy. Involving their children in results disclosure now and again in the future was perceived as an opportunity to partner in managing their child's health and, overtime, shift responsibilities from parent to child. These findings demonstrate that parents can weigh the risks and benefits of receiving secondary results about their children and plan to manage their children's results to maximize benefits and minimize risks for their children. These results challenge conventional thinking and guidelines about the ability of parents to manage and accommodate results from their children and in turn underscore the importance of using empirical data to inform policy.

388

Evaluation of Clinical Utility of Whole Genome Sequencing: the WGS500 Programme. *J. Taylor[1], G. McVean[2], A. Wilkie[3], J. Bell[4], P. Ratcliffe[5], D. Bentley[6], P. Donnelly[2], WGS500 Consortium.* 1) Oxford Biomedical Res Ctr, Wellcome Trust Genetic Ctr, Oxford, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom; 3) Weatherall Institute of Molecular Medicine, Oxford, United Kingdom; 4) Office of Regius Professor, Richard Doll Building, Oxford, United Kingdom; 5) Nuffield Department of Medicine, Henry Wellcome Building for Molecular Physiology, Oxford, United Kingdom; 6) Illumina Inc, Chesterford Research Park, Little Chesterford, Essex, United Kingdom.

Whole genome sequencing (WGS) has been successfully used in a research setting to identify novel genes underpinning disease, but evaluation of the utility of WGS in the clinic is in its infancy. However, WGS has the potential to improve diagnosis and treatment of a range of medical conditions and offers many advantages over current approaches of targeted and exome sequencing, particularly the comprehensive interrogation of all types of genetic variation it provides. Nonetheless, significant challenges must be addressed, including mechanisms for establishing pathogenicity for variants of unknown significance at scale and managing incidental findings. In order to evaluate the potential clinical utility of WGS, we sequenced 500 genomes from patients with 42 different disease types encompassing Mendelian & immunological disorders and cancer. Patients were pre-screened for known candidate genes and chromosomal imbalances as appropriate. Sequencing was conducted on HiSeq 2000 with bioinformatics analysis using a bespoke standardised pipeline to detect, genotype and deeply annotate variants, identify broad scale copy number variation and homozygosity. To date, this study has led to the identification of (i) several novel disease genes, such as POLD & POLE for inherited colorectal cancer, TCF12 for craniosynostosis and a gene of unknown function, C15orf41, for congenital dyserythropoietic anaemia; (ii) identification of novel disease mechanisms for epilepsy phenotypes; (iii) pathogenic variants in regulatory regions; (iv) complex structural rearrangements; and (v) novel phenotypes for known genes. WGS has had substantial clinical impact, including new molecular diagnoses for patients (25% Mendelian cases), more accurate reproductive risk counselling, change of patient diagnosis leading to altered treatment and inclusion of new genes on targeted diagnostic gene panels. We will describe the general lessons learned from clinical sequencing on this scale, from bioinformatics and analysis, through the challenges of establishing pathogenicity, to clinical impact and the practicalities of introducing WGS into the clinic, including the ethical, legal and health economic dimensions, and will illustrate these with case studies. The UK Government recently announced the whole genome sequencing of 100,000 patients in the National Health Service (the UK's single healthcare provider). Our study represents a natural pilot of this approach in the UK context.

389

International views on sharing incidental findings from whole genome research. *A. Middleton[1], M. Parker[2], E. Bragin[1], CF. Wright[1], HV. Firth[1], M. Hurles[1] on behalf of the DDD study.* 1) Wellcome Trust Sanger Institute, Cambridge, Cambridgeshire, United Kingdom; 2) Ethox Centre, University of Oxford, Oxford, United Kingdom.

Whilst whole genome/exome sequencing in a research setting may be used to explore the genetic basis of a phenotype it also offers the chance to opportunistically screen for additional results unrelated to the research project but relevant to the participants' future medical health (termed 'incidental findings', IFs).There is a wealth of medical and ethics literature supporting the feedback of IFs, yet there are limited empirical studies offering a voice from both professional and public stakeholders directly affected by this. The ACMG recommendations on sharing IFs offer advice for the clinic; could these be applied to research sequencing, where scientists may have different obligations and duties, for example to funders? We have gathered the attitudes of more than 5000 people from across the world (e.g. UK, USA, Canada, Australia, S.Africa, Brazil, The Netherlands, Germany, India, Russia, China etc) towards opportunistic genomic screening in a research setting. We created a novel quantitative online survey (www.genomethics.org) to gather attitude data and integrated 10 films within this that describe various ethical issues raised by genome sequencing in a research setting. Genetic health professionals (including clinical geneticists, genetic counsellors and diagnostic lab staff) (53%) were the least likely to think information relating to life-threatening, untreatable conditions should be shared, compared to the public (69%), genomic researchers (72%) and other health professionals (77%) ($p<0.0001$). All groups, apart from genetic health professionals, thought that if research participants wanted it, they should be able to receive their raw sequence data ($p<0.0001$). When asked to consider whether genomic researchers should actively search for IFs not relevant to their original research question, the majority of all groups said no ($p<0.0001$). All stakeholders valued the rights of research participants to have access to incidental genomic data. However, genetic health professionals were the most conservative. This may be due to their appreciation of the complexities involved in translating genomic data in the clinic. There is much anecdotal support in the literature for sharing IFs in a research setting, and our participants agreed with this principle. However, they also thought that genomic researchers should be able to focus on the aims of their study without being forced to actively search for IFs, potentially at the expense of attaining those aims.

390

No evidence for increase in screening among women given report of moderately higher than average risk for breast cancer from personal genomics services: The PGen Study. *S.W. Gray[1], H.Q. Rana[1,2], S. Gollust[3], C.A. Chen[4], S. Kalia[2], J. Mountain[5], T. Moreno[6], J.S. Roberts[7], R.C. Green[2] for the PGen Study Group.* 1) Division of Medical Oncology and Population Sciences, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA; 2) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 3) Division of Health Policy and Management, University of Minnesota School of Public Health, Minneapolis, MN; 4) Data Coordinating Center, Boston University School of Public Health, Boston, MA; 5) 23andMe, Inc., Mountain View, CA; 6) Pathway Genomics, San Diego, CA; 7) Department of Health Behavior and Health Education, University of Michigan School of Public Health, Ann Arbor, MI.

Background: With the growing popularity of personal (direct-to-consumer) genomic testing, customers are receiving cancer risk assessments, but whether modestly increased risk influences screening behaviors is unclear. Responses from the Impact of Personal Genomics (PGen) Study were used to assess whether screening increased within 6 months of receiving modestly elevated breast cancer genetic risk information. Methods: The PGen Study is a web-based survey of consented personal genomics service customers from 23andMe and Pathway Genomics prior to and after receiving genetic risk information. We analyzed responses from women who did not have breast cancer and did not receive positive BRCA variant results (offered by one company). Risk results using common SNP variants were dichotomized as "above average" or "average and below average" as the predictor variable. Primary outcomes were self-reported use of clinical breast exam (CBE), mammography, and breast MRI within 6 months of return of results. Associations between predictor and outcome variables were evaluated using chi square testing and logistic regression. Results: Of 1050 women who took the baseline pre-test survey between March-July 2012, data from the 6 month survey was available for 55% at the time of analysis. Demographics of the 311 women over age 40 in our cohort were mean age 56 years (40-81); 90% White, 3% Black, 3% Hispanic; 74% college educated or above; 35% with family history of breast cancer; 96% insured; 11% above average risk (AAR). Women who received results from common variants indicating AAR did not report significantly more screening behaviors than those with average or below average risk results (ABR): CBE (AAR/ABR) 50%/49%, p=0.92; mammography (AAR/ABR): 41%/48%, p=0.43; breast MRI (AAR/ABR): 5.9%/4.0%, p=0.60. After controlling for socioeconomic characteristics, family history of breast cancer, health insurance, and testing company, there were no significant differences in the use of CBE (adjusted odds ratio (aOR) 1.41, 95% CI 0.65-3.04), mammography (aOR 0.93, CI 0.43-2.02) or breast MRI (aOR 3.74 CI 0.61-23.1) by risk group. Exploratory analysis for women under age 40 revealed similar findings. Conclusion: At 6 month follow-up, female customers of personal genomic testing who received common variant results of moderately above average breast cancer risk did not differ in breast cancer screening practices from customers who received average and below average risk results.

391

Context is complex: attitudes to incorporating genomic risk profiling into population screening programs. *S.G. Nicholls[1], H. Etchegary[2], J.C. Carroll[3], D. Castle[4], L. Lemyre[5], B.K. Potter[1], J. Little[1,6], B.J. Wilson[1] on behalf of the CIHR Emerging Team in Genomics and Screening.* 1) Department of Epidemiology & Community Medicine, University of Ottawa, ON, Canada; 2) Clinical Epidemiology, Memorial University Newfoundland, St John's, NL, Canada; 3) Family Medicine, Mount Sinai Hospital, University of Toronto, ON, Canada; 4) ESRC Innogen Centre, University of Edinburgh, Edinburgh, UK; 5) School of Psychology & Institute of Population Health, University of Ottawa, ON, Canada; 6) Canada Research Chair in Human Genome Epidemiology.

Background: The potential inclusion of genomic risk profiling into population-based screening programs may enable more targeted use of different types and intensities of screening and follow-up. However, the use of genomic profiling within screening will depend in part on public attitudes. It is, therefore, important to engage in a scientific understanding of public attitudes and respond to their concerns to ensure the effective and appropriate implementation of these applications. **Objectives**: To identify (i) general reactions to the idea of incorporating genomic risk profiling into routine screening activities, and (ii) responses to specific issues that may require consideration before implementing genomic risk profiling into routine screening using colorectal cancer (CRC) and newborn bloodspot screening (NBS) as examples. **Methods**: Eight workshops (5 CRC, 3 NBS) were conducted in Ontario and Newfoundland and Labrador, Canada. Participants were provided information regarding the possible personal, health system and societal implications of the technology. Discussions were recorded using field notes and written responses. Participants also completed a structured survey and indicated, from a list, words that best described their attitudes toward the technology. **Results**: Participants (N=170; 120 CRC and 50 NBS) varied in attitude; more participants in the CRC groups indicated that they were "enthusiastic" or "optimistic", while participants in the NBS workshops were more "worried" or had "mixed feelings". More NBS participants were concerned that a result indicating increased risk would cause extra worry compared to those in the CRC workshops (91% vs 49% respectively; $p<0.01$). Other differences included the relevance of results for their family (89% CRC vs 61% NBS; $p<0.01$). In both contexts concern was expressed regarding who had access to test results and implications for insurance. **Conclusion**: Consistent with previous studies in genetic testing, public attitudes appeared to be contingent on the context in which the genomic risk profiling technology was applied. Public attitudes to genomic technologies must, therefore, be considered in a nuanced, context-specific manner and broad brush assessments avoided. A one-size generic policy is unlikely to be satisfactory. That being said, concerns over access to test results and implications on for insurance were common to both contexts.

392

How do citizens balance the benefits and burdens of newborn screening? A choice experiment. *F.A. Miller[1], R.Z. Hayeems[1,2], Y. Bombard[3], C. Cressman[1], C.J. Barg[1], J.C. Carroll[4], B. Wilson[5], J. Little[5], D. Avard[6], J. Allanson[7], P. Chakraborty[7], Y. Giguere[8], D.A. Regier[9].* 1) Institute of Health Policy, Management & Evaluation, University of Toronto, Toronto, Canada; 2) Institute for Clinical and Evaluative Sciences, Toronto, Canada; 3) Yale University, Department of Public Health and Epidemiology; Center for Health Policy and Outcomes, Memorial Sloan Kettering Cancer Center, New York, USA; 4) Department of Family and Community Medicine, Mount Sinai Hospital, University of Toronto, Canada; 5) Department of Epidemiology and Community Medicine, University of Ottawa, Canada; 6) Centre for Genomics and Policy, Department of Human Genetics, McGill University, Montreal, Canada; 7) Department of Genetics, Children's Hospital of Eastern Ontario, and Department of Pediatrics, University of Ottawa, Canada; 8) Department of Medical Biology, Centre Hospitalier, Universitaire de Quebec, University of Laval, Canada; 9) Canadian Centre for Applied Research in Cancer Control, BC Cancer Agency, Vancouver, Canada.

Introduction: Newborn screening (NBS) programs have expanded in recent years with increased debate about the appropriate balance between their benefits and burdens. How the public values and trades-off the effects of NBS, including those impacting affected infants and their families (clinical, informational) and those imposed on the rest of the population (false positive results, overdiagnosis) can influence NBS policy and uptake. **Methods:** In January 2013, a bilingual (French and English) Internet survey was administered to a representative sample of Canadians to assess preferences for NBS across 5 attributes using a discrete choice experiment (DCE). Introduced through a detailed training module, the attributes included: clinical benefits of improved health (none, moderate, significant), earlier time to diagnosis (1 week to 4 years), reproductive risk information (available, not), false positive (FP) results (1-40 per affected baby), and overdiagnosed (OD) infants (0-2 per affected baby). Data were analyzed with a mixed logit model to identify preference heterogeneity among respondents. **Results:** The survey participation rate was 94%; 1220 completed responses met quality criteria (52% completion rate). Respondents prioritized clinical benefits over all other attributes; all respondents positively valued reproductive risk information, while 65% positively valued earlier diagnosis (35% negatively valued this effect). All respondents had a negative preference for FP results, and 98% negatively valued OD (2% positively valued this effect). Nevertheless, respondents were willing to accept large numbers of FP results and some OD infants to achieve moderate clinical benefit for one affected baby (33 FP, 2 OD), and higher numbers to achieve significant clinical benefit (49 FP, 3 OD). **Conclusions:** We report a novel approach to exploring public preferences for the complex trade-offs obliged by population screening programs. Our findings point to broad public understanding of the negative implications of false positive results and overdiagnosis, along with a willingness to tolerate some burdens in the pursuit of clinical benefits.

393

Identifying Genetic Relatives without Compromising Privacy. *E. Eskin, D. He, N. Furlotte, R. Ostrovsky, A. Sahai.* Dept Computer Sci, Univ California, Los Angeles, Los Angeles, CA.

The development of high throughput genomic technologies has impacted many areas of genetic research. While many applications of these technologies focus on the discovery of genes involved in disease from population samples, applications of genomic technologies to an individual's genome or personal genomics have recently gained a lot of interest. One such application is the identification of relatives from genetic data. In this application, genetic information from a set of individuals is collected in a database and each pair of individuals is compared to identify close genetic relatives. An inherent problem with this application and other applications of personal genomics is the issue of privacy. In this project, we propose a method for implementing personal genomics applications without compromising privacy by taking advantage of novel cryptographic techniques customized for secure and private comparison of genetic information. We demonstrate the utility of these techniques by allowing a pair of individuals to discover whether or not they are related without compromising their genetic information or revealing it to a third-party. The idea is that individuals only share enough special-purpose cryptographically protected information with each other to identify whether or not they are relatives, but not enough to expose any information about their genomes. While we focus on identification of relatives, these techniques can be applied to many areas of personal genomics.

394

Targeted sequencing of GPI anchor synthesis pathway genes identifies a new causal genes of hyperphosphatasia with mental retardation. *P. Krawitz[1], Y. Murakami[2], A. Riess[4], M. Hietala[3], U. Krueger[1], N. Zhu[1], T. Kinoshita[2], S. Mundlos[1], J. Hecht[1], P. Robinson[1], D. Horn[1].* 1) Medical Genetics, Charité, Berlin, Berlin, Germany; 2) Department of Immunoregulation, Research Institute for Microbial Diseases, Osaka, Japan; 3) Medical Biochemistry and Genetics, University, Turku, Turku, Finland; 4) Institute for Human Genetics, Eberhard Karls University Tuebingen, Tuebingen, Germany.

Recently, genes involved in the biosynthesis of the glycosylphosphatidylinositol (GPI) anchor came into the spotlight as a new subclass of congenital disorders of glycosylation (CDG) with a characteristic spectrum of clinical features. Up to date mutations in six genes of the GPI-anchor synthesis pathway, PIGA, PIGL, PIGM, PIGN, PIGO, and PIGV have been identified in patients with severe neurological features including seizures, muscular hypotonia, and intellectual disability. We developed a diagnostic gene panel for targeting all known genes of the GPI-anchor synthesis pathway to screen patients matching these features. By this means we detected three missense mutations in PGAP2, c.46C>T, c.380T>C, and c.479C>T in two unrelated patients with hyperphosphatasia mental retardation syndrome (HPMRS). These mutations cosegregated in the investigated families. PGAP2 is a gene coding for an acyltransferase that is involved in fatty acid remodeling of the GPI-anchor that is required for Golgi transport of GPIlinked substrates. Transfection of the mutant constructs p.Arg16Trp, p.Leu127Ser, and p.Thr160Ile into PGAP2-null cells showed only partial restoration of GPI-anchored marker proteins, CD55 and CD59, on the cell surface. In this work we show that also an impairment of GPI-anchor remodeling causes HPMRS and conclude that targeted sequencing of the GPI-anchor pathway genes is an effective diagnostic approach for a subclass of CDGs.

395

1000 Trio exomes; Insights into severe developmental disorders. *M. Van KOGELENBERG, T. Fitzgerald, W. Jones, J.C. Barret, M. Hurles on behalf of the DDD project.* The Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

The UK-based Deciphering Developmental Disorders (DDD) project aims to delineate the genetic architecture in children with undiagnosed severe developmental disorders. The study employs a high throughput approach that aims to recruit 12.000 children with developmental disorders and their parents via all of the 24 Regional Genetics Services in the UK and republic of Ireland.

The clinically diverse patient cohort in this study requires the consideration of both inherited and *de novo* disease models. Our approach is focused on detecting sequence variants in all coding exons, known enhancers, and the most highly conserved non-coding element. *De novo* exome variants where systematically validated by capillary sequencing which detected ~0.88 coding or splice *de novo* mutations (DNM) per trio.

Here we intend to present an overview of the exome data analysis in 1000 patient-parent trios. This includes various types of burden analysis such as the detection of a significant enrichment of loss of function DNM variants, in addition to DNM protein altering variants being more frequently present in dosage sensitive genes. Further, we will discuss the significance of identifying variants in known monoallelic disease genes, as well as the detection of recurrent functional DNM in both known disease genes (19) and genes not previously associated with human disease (26).

396

Atypical Rett Syndrome: Is it really more common in females? *K. Cusmano-Ozog, P. Tanpaiboon, L. Harris, J. Turner, L. Kehoe, T. Biagi, B. Lanpher.* Genetics & Metabolism, CNMC, Washington, DC.

Atypical Rett syndrome is a neurodevelopmental disorder identified in children who exhibit a Rett-like phenotype but do not meet the clinical criteria for Rett syndrome. A variety of genetic alterations have been associated with this phenotype including duplications of MECP2, alterations in another X-linked gene, CDKL5, as well as mutations or deletions of FOXG1 on 14q12. In the last two years at CNMC we have identified 10 probands with a Rett-like phenotype who also had a pathogenic copy number change detected by chromosomal microarray. Five males have a duplication of MECP2, two males have a deletion of FOXG1 and one male and two females have a deletion of CDKL5. Age of diagnosis varied between 1.5 and 19 years with an average age of six. Of the 10 individuals, four have a seizure disorder (infantile spasms, complex partial, myoclonic, multifocal) and is suspected but not proven in a fifth individual. EEG abnormalities with diffuse background slowing suggestive of encephalopathy were noted in four. Brain MRI studies were abnormal in six revealing delayed myelination, posterior periventricular gliosis, bilateral choroid plexus cysts, prominent ventricles and/or volume loss. Three presented with microcephaly; both males with FOXG1 deletions and one female with CDKL5 deletion. None of the males with MECP2 duplication were microcephalic. Asymmetric spasticity was identified in one male with MECP2 duplication, otherwise hypotonia was a common finding. Feeding and swallowing issues with failure to thrive were noted in all but one and six require G-tube feeds; one each also has recurrent pneumonia or sleep apnea. Additional problems identified in single individuals include: cataract, sensory neural hearing loss and coarctation, coronary artery fistula, diaphragmatic hernia, and congenital hip dysplasia. Our findings indicate that atypical Rett syndrome is a common condition that may be readily identified by chromosomal microarray. It has been thought that atypical Rett syndrome occurs more commonly in girls; however, the majority of our cases (80%) are male suggesting this may be under diagnosed in the general population. Additional studies are indicated to better determine prevalence, gender distribution and genotype-phenotype correlations.

397

Common molecular networks in Rett, Angelman, Smith-Magenis, Potocki-Lupski, Pitt-Hopkins, and chromosome 2q23.1 deletion syndromes contribute to intellectual disability, seizures, sleep, language, behavior and autism spectrum disorder. *S.V. Mullegama[1,2], B. Burns[2], Z. Shah[2], R. Tahir[2], W-H. Tan[3], S.H. Elsea[1,2,4].* 1) Human and Molecular Genetics, Baylor College of Medicine, Houston , TX; 2) Department of Human and Molecular Genetics, Virginia Commonwealth University School of Medicine, Richmond, VA; 3) Division of Genetics, Boston Children's Hospital, Boston, MA, USA; 4) Department of Pediatrics, Virginia Commonwealth University School of Medicine, Richmond, VA, USA.

Chromatin modifying genes play important roles in the genetic etiology of neurodevelopmental disorders (NDs). Many monogenic NDs are caused by altered dosage of chromatin modifying genes, including Rett (RTT, *MECP2*), Angelman (AS, *UBE3A*), Smith-Magenis (SMS, *RAI1*), Potocki-Lupski (PTLS, *RAI1*), Pitt-Hopkins (PTHS, *TCF4*), and 2q23.1 del/dup syndromes (*MBD5*). NDs are genetically and phenotypically heterogeneous but share common features that include intellectual disability with severe language delay, autism spectrum disorder, sleep disturbances, and behavioral difficulties; however, the molecular mechanisms responsible for these associations remain unknown. To investigate whether common networks are dysregulated resulting in these overlapping features, we evaluated expression of the 5 causative genes for these disorders in disorder-specific cell lines and patient lymphocyte samples, as well as gene expression microarrays of cell lines with gene-specific knockdown. QPCR studies revealed altered expression of *MECP2*, *UBE3A*, *RAI1*, *MBD5*, and *TCF4* in these cell lines. All genes had significantly reduced expression in RTT lines, which suggests a hierarchy of genes, with *MECP2* dosage highly critical for normal expression of these genes. Reduced expression of *UBE3A* in SMS lines and its overexpression in PTLS lines suggests *UBE3A* expression is sensitive to RAI1 dosage, which is corroborated by our ChIP-chip studies that indicate RAI1 directly binds to the *UBE3A* promoter. Expression of *RAI1* is reduced in both 2q23.1 disorders suggesting possible regulation of *RAI1* by MBD5. Furthermore, knockdown of *MBD5* and *RAI1* in neuroblastoma cell lines revealed dysregulation of known autism-linked genes, including *MLL3*, *PLAUR*, *TBX1*, *HDAC10*, *SMARCB1*, and *ABAT*, as well as other chromatin modifying genes. Pathway analyses showed that MBD5 and RAI1 function in chromatin remodeling, neuronal development, and cell growth/survival pathways. Two pathways disrupted by haploinsuffiiciency of *RAI1* and *MBD5* were the mTOR and circadian rhythm signaling pathways associated with autism and sleep phenotypes, respectively. Overall, these data suggest the presence of interconnected neurodevelopmental networks that when impacted by changes in gene expression may result in overlapping features across these syndromes. Identifying common points of regulation in these pathways may lead to therapeutic intervention for treatment of the common phenotypes, including sleep and behavioral problems.

398

***MBD5* deletion disrupts circadian gene expression and is associated with sleep disturbance in the 2q23.1 deletion syndrome.** *S.H. Elsea[1,2], S.V. Mullegama[1,2], Z. Shah[2], R. Tahir[2], L. Pugliesi[2].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 2) Department of Human and Molecular Genetics, Virginia Commonwealth University School of Medicine, Richmond, VA, USA.

2q23.1 deletion syndrome is a complex neurodevelopmental syndrome that includes intellectual disability, epilepsy, speech impairment, and autism spectrum disorder (ASD) and is caused by haploinsufficiency of *MBD5*. Sleep disturbance, while previously reported, has not been assessed in this population. Since phenotypically overlapping disorders like Smith-Magenis syndrome (SMS) and ASD have sleep phenotypes that impact daily living, we sought to clinically and molecularly characterize the circadian deficits in the 2q23.1 deletion syndrome. Caregiver surveys were used to gather information on sleep and activity habits of 20 children aged 9 mon to 11 y with 2q23.1 deletions. Caregivers reported many sleep concerns for their child, including difficulty falling asleep (46%), multiple nighttime wakings and early morning waking (74%), and use of medications to improve sleep (44%), features commonly observed in both autism spectrum disorder and SMS. Data from the Epworth Sleepiness Scale revealed 44% of school-age children with excessive daytime sleepiness, and sleep problems were reported to affect daily functioning of the child and the caregiver, thus impacting the overall well-being of the family. We previously showed that RAI1 directly modulates the transcription of *CLOCK* and that multiple circadian genes are dysregulated in SMS cell lines, so we used gene expression and ChIP-chip data to molecularly characterize the impact of *MBD5* deletion on sleep/circadian rhythm pathways. Data show that *RAI1*, *NR1D2*, and *PER3* were dysregulated in 2q23.1 deletion patient cell lines, similar to findings in SMS. Pathway analyses from gene expression microarrays of knockdown of *MBD5* in SH-SY5Y cell lines confirmed patient cell line data, revealing dysregulation of circadian rhythm genes, including *ATF2*, *ATF4*, *CLOCK*, and *PER3*. ChIP-chip studies also suggest that MBD5 directly regulates the transcription of *NR1D2*. These data support MBD5 as a regulator of circadian rhythm gene expression, providing a molecular basis for the sleep phenotype present in 2q23.1 deletion syndrome. Overall, delineation of a sleep phenotype in 2q23.1 deletion syndrome and molecular data to support the phenotypic findings places this MBD5-associated disorder into a category of neurodevelopmental syndromes where sleep behaviors significantly impact daily function and identify pathways for targeted therapeutic intervention that may be common to several neurodevelopmental syndromes.

399

Lysyl-tRNA Synthetase (KARS) Mutations Cause Autosomal Recessive Nonsyndromic Hearing Impairment DFNB89. *R. Santos-Cortez[1], K. Lee[1], Z. Azeem[2,3], P.J. Antonellis[4,5], L.M. Pollock[4,6], S. Khan[2], . Irfanullah[2], P.B. Andrade-Elizondo[1], I. Chiu[1], M.D. Adams[6], S. Basit[2], J.D. Smith[7], D.A. Nickerson[7], B.M.Jr. McDermott[4,5,6], W. Ahmad[2], S.M. Leal[1], University of Washington Center for Mendelian Genomics.* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA; 2) Department of Biochemistry, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan; 3) Department of Biochemistry, AJK Medical College, Muzaffarabad, Azad Jammu & Kashmir, Pakistan; 4) Department of Otolaryngology - Head and Neck Surgery, Case Western Reserve University, Cleveland, Ohio 44106, USA; 5) Department of Biology, Case Western Reserve University, Cleveland, Ohio 44106, USA; 6) Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, Ohio 44106, USA; 7) Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA.

Previously a novel autosomal recessive nonsyndromic hearing impairment (ARNSHI) locus DFNB89 was mapped to chromosome 16q21-q23.2 in three unrelated, consanguineous Pakistani families. Through whole-exome sequencing of a hearing-impaired individual from each family, novel missense mutations were identified at highly conserved residues of lysyl tRNA-synthetase (KARS), namely one family with the c.1129G>A (p.Asp377Asn) variant and two families with the c.517T>C (p.Tyr173His) variant. Both variants were predicted to be damaging by multiple bioinformatics tools. The two variants both segregated with the nonsyndromic hearing impairment phenotype within the three families and neither mutation was identified in ethnically matched controls or within variant databases. Individuals who were homozygous for KARS mutations had symmetric, severe hearing impairment across all frequencies, but did not show evidence of auditory or limb neuropathy. It is demonstrated that KARS is expressed in hair cells of zebrafish, chicken, and mouse. Moreover, KARS has strong localization to the spiral ligament region of the cochlea, and also to the sulcus epithelium, Deiters' cells, basilar membrane, and the surface of the spiral limbus. It is hypothesized that KARS variants affect aminoacylation in inner ear cells by interfering with binding activity to tRNA or p38 and with tetramer formation. The identification of rare KARS variants in ARNSHI families defines a novel gene for ARNSHI.

400

Multiple De-Novo Variants Resulting in Combined Axial Hypotonia with Dyskinesia and Facial Myokymia. *A. Torkamani[1], J. Friedman[2], C.S. Bloss[1], S. Topol[1], E.J. Topol[1], Q. Chen[4], N.J. Schork[1], W.H. Raskind[3], A. Torkamani[1].* 1) Scripps Translational Science Institute, Scripps Health and The Scripps Research Institute, San Diego, CA; 2) Department of Neurosciences, University of California at San Diego, San Diego, CA; 3) School of Medicine, University of Washington, Seattle, WA; 4) Cellular Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA.

Purpose: Recent genome sequencing studies have demonstrated the potential for rapid and relatively inexpensive molecular diagnosis and treatment prioritization in genetically heterogeneous, previously characterized diseases and diagnostic odyssey cases. Of the ~100 patients successfully diagnosed by the NIH Intramural Undiagnosed Disease Program - ~83% correspond to known disease-gene relationships, ~15% correspond to novel gene associations with described diseases, and only 2% correspond to previously unknown diseases. Herein, we present the successful molecular diagnosis of a previously unknown and complex movement disorder through combined whole exome (WES) and whole genome (WGS) sequencing and confirm the functional role of ADCY5 mutations in dyskinesia with facial myokymia. Furthermore, treatment informed by the molecular diagnosis led to partial resolution of the disorder. Methods: Combined whole genome sequencing (WGS) and whole exome sequencing (WES) was performed on an affected 15-year-old female and her unaffected parents in order to identify the genetic cause of a previously undiagnosed condition involving a lifelong history of hypotonia, motor delays, abnormal involuntary movements, and sleep disturbances. A combination of inheritance-based, population-based, functional-impact-based and variant annotation-based filters applied to the WES data, then validated by WGS, resulted in three genes bearing potential candidate variants consistent with the disease. Subsequent literature-based investigation led to the conclusion that de novo nonsynonymous mutations in ADCY5 and DOCK3, as well as a rare maternally inherited nonsynonymous mutation in DOCK3 were the likely cause of the complex phenotype. Results: Mice deficient in Adcy5 develop a Parkinsonian-like phenotype and mice deficient in Dock3 develop central axonal dystrophy and sensorimotor dysfunction. Functional validation confirmed ADCY5 gain of function and immunohistochemistry of peripheral nerves confirmed neurofilament aggregates characteristic similar to those observed in Dock3 deficient mice. Motivated by the previously observed efficacy of acetazolamide in the treatment of autosomal dominant familial dyskinesia and facial myokymia due to ADCY5 mutation, initiation of acetazolamide treatment in this individual resulted in improvement of the movement disorder.

401

Dominant β-catenin mutations cause intellectual disability with recognizable syndromic features. *T. Kleefstra[1], V. Tucci[2,3], M.H. Willemsen[1], A. Hardy[3], I. Heise[3], S. Maggi[2], W. Wissink-Lindhout[1], A. Vulto-van Silfhout[1], B.B.A. deVries[1], Z. Iqbal[1], H.G. Brunner[1], W.N. Nillesen[1], H.G. Yntema[1], H. Hilton[3], M. Simon[3], S. Tsaftaris[4,5], H. van Bokhoven[1], A. Constestabile[2], T. Nieus[2], A. Raimondi[2], B. Greco[2], D. Cantatore[2], L. Gasparini[2], L. Berdondini[2], A. Bifone[5], J. Veltman[1], L. Peart-Vissers[1], A. Gozzi[5], S. Wells[3], P.M. Nolan[3].* 1) Human Genetics, Radboud University Medical Center, Nijmegen, Netherlands; 2) Department of Neuroscience and Brain Technologies - Istituto Italiano di Tecnologia, via Morego, 30, 16163 Genova; 3) MRC Harwell, Harwell Science and Innovation Campus, Oxfordshire, OX11 0RD, UK; 4) Instituto Italiano di Tecnologia, Center for Nanotechnology Innovation @NEST, Piazza San Silvestro 12, Pisa, Italy; 5) IMT Institute for Advanced Studies Lucca, Piazza San Ponziano 6, Lucca, Italy.

The recent identification of multiple dominant mutations in both humans and mice has enabled us to explore the molecular and cellular basis of β-catenin function in cognitive impairment. Mutations in human CTNNB1 have been identified as causative in a spectrum of neurodevelopmental disorders. In identifying de novo CTNNB1 mutations in patients with intellectual disability and careful characterization of their phenotype, we were able to define a recognisable intellectual disability syndrome. In parallel, the characterization of a chemically-mutagenized mouseline displaying features that are similar to these human mutations hads enabled us to investigate the consequences of β-catenin dysfunction through development and into adulthood. The mouse mutant displays a reduced affinity for membrane-associated cadherins in association with this decreased cadherin interaction. We find that the mutation results in decreased intrahemispheric connections with deficits in dendritic branching long-term potentiation and cognitive function. For the first time in vivo we show how dominant mutations in CTNNB1 underlie losses in its adhesion-related functions leading to severe consequences including intellectual disability, progressive spasticity and abnormal craniofacial features in humans.

402

Defective Initiation of Glycosaminoglycan Synthesis due to B3GALT6 Mutations Causes a Pleiotropic Ehlers-Danlos Syndrome-like Connective Tissue Disorder. *F. Malfait[1], A. Kariminejad[2], T. Van Damme[1], C. Gauche[3], D. Syx[1], F. Merhi-Soussi[3], S. Gulberti[3], S. Symoens[1], S. Vanhauwaert[1], A. Willaert[1], B. Bozorgmehr[2], M. Kariminejad[2], I. Hausser[4], S. Fournel-Gigleux[3], A. De Paepe[1].* 1) Center for Medical Genetics Ghent University Hospital, Gent, Belgium; 2) Kariminejad-Najmabadi Pathology & Genetics Center, 1143 Med. 4th Str. Third Phase, Shahrak Gharb, 14656 Tehran, Iran; 3) UMR 7365 CNRS-Université de Lorraine (Ingénierie Moléculaire et Pharmacologie Articulaire, IMoPA), MolCelTEG Team, Biopôle UL, Faculté de Médecine, 54505 Vandoeuvre-lès-Nancy, France; 4) Department of Dermatology, University of Heidelberg, 69120 Heidelberg, Germany.

Proteoglycans are important components of cell plasma membranes and extracellular matrices of connective tissues. They consist of glycosaminoglycan chains attached to a core protein via a tetrasaccharide linkage, whereby the addition of the third residue is catalyzed by galactosyltransferase II (β3GalT6), encoded by B3GALT6. Homozygosity mapping and candidate gene sequence analysis in three independent families, presenting a severe autosomal-recessive connective tissue disorder characterized by skin fragility, delayed wound healing, joint hyperlaxity and contractures, muscle hypotonia, intellectual disability, and a spondyloepimetaphyseal dysplasia with bone fragility and severe kyphoscoliosis, identified biallelic B3GALT6 mutations, including homozygous missense mutations in family 1 (c.619G>C [p.Asp207His]) and family 3 (c.649G>A [p.Gly217Ser]) and compound heterozygous mutations in family 2 (c.323_344del [p.Ala108Glyfs*163], c.619G>C [p.Asp207His]). The phenotype overlaps with several recessive Ehlers-Danlos variants and spondyloepimetaphyseal dysplasia with joint hyperlaxity. Affected individuals' fibroblasts exhibited a large decrease in ability to prime glycosaminoglycan synthesis together with impaired glycanation of the small chondroitin/dermatan sulfate proteoglycan decorin, confirming β3GalT6 loss of function. Dermal electron microcopy disclosed abnormalities in collagen fibril organization, in line with the important regulatory role of decorin in this process. A strong reduction in heparan sulfate level was also observed, indicating that β3GalT6 deficiency alters synthesis of both main types of glycosaminoglycans. In vitro wound healing assay revealed a significant delay in fibroblasts from two index individuals, pointing to a role for glycosaminoglycan defect in impaired wound repair in vivo. Our study emphasizes a crucial role for β3GalT6 in multiple major developmental and pathophysiological processes.

403

Whole exome sequencing of 2126 African American prostate cancer cases and controls from the Multiethnic Cohort. *K.A. Rand[1,4], N. Rohland[2,3,4], A. Tandon[2,3], R. Do[3], X. Sheng[1], D.V. Conti[1], B.E. Henderson[1], C.A. Haiman[1,4], D. Reich[2,3,4].* 1) Department of Preventive Medicine, Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA 90033, USA; 2) Department of Genetics, Harvard Medical School, Boston, MA 02115, USA; 3) Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142, USA; 4) These authors contributed equally to this work.

Prostate cancer disproportionately affects African American men, with more than a two-fold increased risk when compared to other racial/ethnic populations. Previous studies have identified common genetic variants associated with prostate cancer, as well as regions of the genome that explain a portion of the excess risk observed in African Americans; however, a large proportion of the heritability of prostate cancer remains unexplained. Furthermore, these studies have been limited in their ability to adequately assess the contribution of risk from rare variants (minor allele frequency (MAF) < 0.01), which have been hypothesized to have large effects. Whole exome sequencing provides the opportunity to directly examine rare functional mutations within the coding region of the genome. To further explore both common and rare genetic variation associated with prostate cancer, we sequenced the exomes of 2126 African American prostate cancer cases and controls from the Multiethnic Cohort. A barcoded library preparation was utilized, allowing for highly multiplexed target enrichment, with a 51Mb targeted region encompassing 20,965 genes and 334,278 exons. On average, these individuals were sequenced at 8.2x coverage (+/- 4.2x). After initial filtering, we identified 486,143 variants in coding regions, with 44,366 singletons. We observed 286,015 nonsynonymous, 188,186 synonymous, 7,154 stopgain, and 303 stoploss variants, respectively. Of the variants with a MAF > 0.01, 97% have been previously identified in the 1000 genomes project. The rare variants (MAF < 0.01) include 58% of the observed splicing sites and 61% of the nonsynonymous variants. In the preliminary association analysis, we identified 36 SNVs with a p-value < 5×10-4, and are currently sequencing another 2000 prostate cancer cases and controls that will be included in the final analysis. Results from this analysis may shed light on the effect of rare variants on prostate cancer risk, as well as provide insight into the genetic basis for the observed racial/ethnic disparity in disease incidence.

404

Identification of Y chromosomes associated with risk for prostate cancer. *L.A. Cannon-Albright[1,2], J.M. Farnham[1], C.C. Teerlink[1], R.A. Stephenson[1,2,3].* 1) University of Utah School of Medicine, Salt Lake City, UT; 2) George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, UT; 3) Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT.

Published evidence suggests that genes present on the Y chromosome may be involved in increased risk for prostate cancer; however, the Y chromosome has received little attention in previous genetic studies of prostate cancer. An analysis of the computerized Utah Population Database (UPDB) genealogy of the Utah pioneers linked to statewide cancer data was used to test the hypothesis of a role of the Y chromosome in prostate cancer predisposition. Analysis was limited to the 1.25 million individuals in the UPDB genealogy who have at least 12 of their 14 immediate ancestors. Within the UPDB we identified all males with no father in the genealogy data (founders) and assigned them each a unique, sequential Y chromosome id (YID); each of their male descendants, and all of his male descendants, and so forth, were assigned this same YID, effectively identifying each independent Y chromosome in the UPDB. We identified 257,252 YIDs for which there were at least 2 males who shared each Y chromosome. The largest YID group included 2264 males. All YIDs were assumed to be distinct based on genealogy data. We estimated rates of prostate cancer for the population of 1.25 million individuals in the UPDB using birth year and birth state (Utah or not) cohorts and using the Utah Cancer SEER registry to identify cases. These rates were applied to all males in each YID group to estimate the number of prostate cancer cases expected. We tested these 1000 independent sets of males, each hypothesized to share the same Y chromosome, for a significant excess of prostate cancer. In these 1000 sets of between 167 and 2264 males sharing a Y chromosome we identified multiple Y chromosomes associated with a significant excess risk for prostate cancer ($p<0.05/1000$ correction for multiple testing). An example Y chromosome group of interest includes a total of 9750 total male descendants of the YID founder; there were a total of 65 prostate cancer cases among all the male descendants (45.6 expected; p=0.005); there were a total of 498 males sharing the Y chromosome of the Y group founder with 26 prostate cancer cases observed (9.5 expected, p=8E-6). Among the male descendants who did not share the Y chromosome of the founder there were 39 prostate cancer cases observed, with 36.1 expected (p=0.68). These results show strong support for a role of specific Y-chromosomes in prostate cancer predisposition and identify a powerful resource for identifying the genes or variants responsible.

405

Genome-wide Scan Identifies a Novel Locus Associated with Aggressive Prostate Cancer. *S.I. Berndt[1], Z. Wang[1,2], M. Yeager[1,2], W.R. Diver[3], S. Gapstur[3], V.L. Stevens[3], D. Albanes[1], S. Weinstein[1], J. Virtamo[4], J. Cornu[5], O. Cussenot[5], G. Cancel-Tauset[5], S. Lindström[6], P. Kraft[6], D. Hunter[6], L. Amundadottir[1], A. Black[1], J. Sampson[1], K. Jacobs[1], M. Tucker[1], S.J. Chanock[1].* 1) Division of Cancer Epidemiology & Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD; 2) Core Genotyping Facility, Advanced Technology Program, SAIC Frederick, Inc., NCI-Frederick, Frederick, MD; 3) Department of Epidemiology and Surveillance Research, American Cancer Society, Atlanta, GA; 4) Department of Health Promotion and Chronic Disease Prevention, National Public Health Institute, Helsinki, Finland; 5) CeRePP Hopital Tenon, Assistance Publique-Hôpitaux de Paris, Paris, France; 6) Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts.

Prostate cancer is the most common cancer in men, but the aggressiveness of the disease varies widely. Some prostate cancers are aggressive, leading to invasive or fatal disease. Others are indolent and unlikely to pose a serious health threat within a man's lifetime. Genome-wide association studies (GWAS) have identified over 70 loci associated with prostate cancer risk; however, none are specific for aggressive disease, which is critical for clinical risk prediction and prevention. We conducted a GWAS of 4,600 prostate cancer cases (including 4,545 cases with Gleason score) and 2,941 controls of European ancestry from the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial using the Illumina Omni2.5 chip (stage 1). To identify genetic loci associated with prostate cancer aggressiveness, we conducted a case-only analysis using Gleason score, which is pathological measure of tumor aggressiveness, in a linear regression model. The most promising loci with $P<0.001$ from this model were taken forward for replication in 6,572 cases (including 5,353 cases with Gleason score) and 6,391 controls of European ancestry from five studies using a custom Illumina iSelect chip (stage 2). A meta-analysis of stage 1 and stage 2 case-only results for Gleason score yielded one locus at 5q14.3, near *RASA1/CCNH*, that was associated with increased tumor aggressiveness at a genome-wide significance level ($\beta=0.11$, $P=6.90\times10^{-11}$). Case-control analyses stratified by disease aggressiveness further demonstrated that the locus was specific for aggressive disease. The 5q14.3 SNP was associated with an increased risk of aggressive prostate cancer (n=1170 cases/9,332 controls, Gleason ≥ 8; OR=1.27, $P=2.26\times10^{-6}$), but not associated with non-aggressive disease (n=5421 cases/9332 controls, Gleason ≤ 6; OR=0.97, P=0.22). Additional replication is underway. The SNP is intronic within the RAS p21 protein activator 1 (*RASA1*) gene, which suppresses RAS signaling and angiogenesis, and is 79 kb downstream from cyclin H (*CCNH*), which is involved in transcription regulation and cell cycle control. Both genes have high biological relevance for prostate cancer, particularly for invasive disease. In conclusion, this is the first study to identify and replicate a locus that is specific for aggressive prostate cancer and not indolent prostate cancer, representing an important advance in the understanding the etiology of clinically significant aggressive prostate cancer.

406

Frequent germline mutations in DNA repair genes in familial prostate cancer cases. *D. Leongamornlert[1], E. Saunders[1], T. Dadaev[1], M. Tymrakiewicz[1], C. Goh[1], S. Jugurnauth-Little[1], I. Kozarewa[2], K. Fenwick[2], I. Assiotis[2], D. Barrowdale[3], K. Govindasami[1], M. Guy[1], E. Sawyer[1], R. Wilkinson[1], A. Antoniou[3], R. Eeles[1, 4], Z. Kote-Jarai[1].* 1) Oncogenetics, The Institute of Cancer Research, Sutton, Surrey, United Kingdom; 2) Gene Function, The Institute of Cancer Research, London, United Kingdom; 3) University of Cambridge, Cambridge, United Kingdom; 4) The Royal Marsden Fundation Trust, Sutton and London, United Kingdom.

Prostate cancer (PrCa) is one of the most common diseases to affect men worldwide and the underlying aetiology is poorly understood compared with other complex diseases. Many common variants have been identified over the last few years which are associated with PrCa risk but very few rare, higher risk alleles have been found. To investigate if rare variants in 22 DNA repair genes have a role in PrCa predisposition, we analysed these genes in a UK collection of familial prostate cancer cases. Germline DNA samples were obtained from 191 PrCa cases which had 3 or more relatives affected with PrCa. These samples were deeply re-sequenced using Agilent SureSelect target capture and Illumina HiSeq systems. The exonic and intronic sequences of 22 tumour suppressor DNA repair genes were analysed for genomic variation using a pipeline consisting of BWA, GATK and Annovar. We identified 14 loss-of-function (LoF) mutations (7.3%) with the highest fraction contributed by *BRCA2* (4 mutations). Mutations were also found in *ATM, CHEK2, BRIP1, BRCA1, MUTYH, PALB2* and *PMS2*. In addition, 13 missense variants which are likely to modify protein function were also identified. LoF mutation carriers had significantly higher odds of nodal involvement, metastasis or death from PrCa OR 11.32 (95% CI: 2.27-69.56; *P*=0.00421). Using a modified segregation analysis approach, we estimate that LoF mutations in any of the studied genes confer a relative risk of PrCa of 4.64, (95% confidence interval 3.84-5.61). In summary we present evidence that mutations in DNA repair genes may have a significant role in familial PrCa predisposition and that carriers of LoF mutations in these genes are more likely to have aggressive disease. This may have clinical utility in the implementation of new screening and treatment strategies.

407

Pleiotropic effect of rare mutation in *HOXB13* on multiple cancers detected in a cohort of >100K individuals via imputation. *J. Witte[1], T. Hoffmann[1], L. Sakoda[2], L. Shen[2], E. Jorgenson[2], M. Asgari[2], D. Corley[2], L. Habel[2], L. Kushi[2], M. Kwan[2], C. Schaffer[2], S. Van Den Eeden[2], N. Risch[1].* 1) Epidemiology & Biostatistics, Univ California,San Francisco, San Francisco, CA; 2) Division of Research, Kaiser Permanente, Northern California, Oakland, CA.

A rare, high penetrance mutation for prostate cancer was recently identified in the *HOXB13* gene (G84E). We successfully imputed this mutation into the Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging, a large cohort of >100,000 individuals with genome-wide genotype data. The imputation reference panel included data from the 1000 Genomes Project supplemented with sequence data from 22 known G84E mutation carriers. To verify that the mutation was imputing accurately from the reference panel, we developed a classification and regression tree method based on pointwise linkage disequilibrium to identify a founder haplotype. We found that the *HOXB13* mutation (G84E) was associated not only with prostate cancer ($p=3.2\times10^{-13}$), but also with an increased risk of the following cancers in the cohort: breast, colon, melanoma, non-Hodgkin lymphoma, and endometrium (combined odds ratio=1.58, P-value=0.025; adjusted for multiple testing). In addition to providing evidence of a pleiotropic effect of the G84E *HOXB13* mutation on cancer, our observations show that with a sufficiently large reference sample one can successfully impute very rare variants into large cohorts, and confirm these with novel classification methods.

408

Detection of Large Rearrangements in *PMS2*. *D. Mancini-DiNardo[1], J.W. Landon[1], S. Rajamani[2], K. Moyes[1], C. Arnell[1], I. Dorweiler[1], K. Bowles[1], B. Leclair[1], B. Roa[1].* 1) Myriad Genetic Laboratories, Inc., Salt Lake City, UT; 2) Myriad Genetics, Inc, Salt Lake City, UT.

Heterozygous germline mutations in *PMS2* are associated with Hereditary Non-polyposis Colorectal Cancer (HNPCC or Lynch Syndrome). As one of the four primary mismatch repair genes associated with Lynch syndrome, the functional importance of *PMS2* has been clear, but its total contribution to Lynch syndrome was historically considered to be quite low. More recent studies suggest that the prevalence of *PMS2* mutations is comparable to *MSH6*, with as much as 15% of all Lynch syndrome attributable to *PMS2*. While inactivation of *PMS2* is caused largely by sequencing mutations, a significant proportion of all *PMS2* mutations have been reported to be large rearrangements (LRs). Until recently, detection of LRs in *PMS2* has been hindered by the presence of numerous pseudogenes that reside on chromosome 7, the most problematic of which is *PMS2CL*, lying 0.7Mb centromeric to *PMS2*. Due to extensive gene conversion events, this transcribed pseudogene bears striking homology to the 3' end of the functional *PMS2* gene, specifically in exons 9, and 11-15. By combining various methods for analyzing *PMS2*, our laboratory has been successful in ascribing the correct mutational status to both *PMS2* and *PMS2CL* in a given patient sample. Here we describe the results of our comprehensive analysis of *PMS2* which includes detection of LRs in regions that are complicated by the presence of the pseudogene. Our analysis of patients who have received full gene sequencing and LR analysis of *MLH1, MSH2, MSH6* and *PMS2,* has shown that deleterious and suspected deleterious mutations in *PMS2* comprise 14.3% of all mutations detected in these HNPCC-related genes (108/755). Among these *PMS2* mutation-positive patients, we determined that 25% (27/108) of these mutations are LRs compared with 75% (81/108) sequencing mutations. 29.6% of the LRs detected involve exons exclusively within the pseudogene regions. Though we have detected multiple duplications, all of the deleterious and suspected deleterious LRs we have found in *PMS2* thus far have been deletions. This analysis demonstrates that LRs represent a significant portion of the *PMS2* mutational spectrum, particularly within the exons shared by the pseudogene, *PMS2CL*. As mutations within *PMS2* alone are clinically actionable, it is of critical importance to distinguish between the mutational status *PMS2* and the pseudogene. Our laboratory utilizes a multi-faceted process to investigate the presence of mutations within this region in both *PMS2* and *PMS2CL*.

409

Functional analysis of the chr13q22.1 pancreatic cancer risk locus suggests allele-specific effects on DIS3 expression. *J. Hoskins[1], A. Ibrahim[1], H. Parikh[1], J. Kim[1], J. Jia[1], I. Collins[1], G. Petersen[2], L. Amundadottir[1].* 1) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Gaithersburg, MD; 2) Mayo Clinic, Rochester, MN.

Pancreatic cancer is the 10th most common cancer and 4th most common cause of cancer mortality in the United States. A genome wide association study revealed pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. The most significant single nucleotide polymorphism (SNP) identified in this GWAS, rs9543325, is located in chr13q22.1 within a gene desert. The nearest genes to rs9543325 are *KLF5, KLF12, PIBF1, DIS3* and *BORA*, which range from 265kb to 586kb away, respectively. Given the high linkage disequilibrium across this risk locus, there are many candidate functional variants to consider. Imputations of these functional variants did not improve the signal, but produced a set of highly correlated SNPs. In an effort to identify the functional variant(s) we performed eQTL analyses to test the association between the genotypes of these candidate SNPs and expression of nearby genes. Among 100 normal pancreatic tissue samples, *DIS3* showed the strongest association with a novel variant in the risk locus (*P*-value = 0.0004). Mutations in *DIS3* have previously been identified in acute myeloid leukemia and multiple myeloma, and its expression has been correlated with metastatic potential in colorectal cancer, suggesting this gene could be important in pancreatic cancer biology. Chromosome conformation capture (3C) was performed to test for physical interactions between the risk locus and nearby genes. This assay confirmed the three dimensional proximity of the risk locus and a region near the *DIS3* promoter. Finally, sub-regions of the risk locus were cloned upstream of a luciferase expression construct to assay for potential allele-specific enhancer/silencer activity. This assay revealed the sub-region containing the most significant *DIS3* eQTL acts as an allele-specific silencer. Together, these results suggest that a sub-region of the chr13q22.1 pancreatic cancer risk locus has allele-specific effects on *DIS3* expression, which may have implications in pancreatic cancer susceptibility.

410

Genome-wide association study of colorectal adenoma in the Nurses' Health Study and the Health Professionals Follow-up Study. *A.D. Joshi[1], A. Hazra[1,2], C. Chen[1], R.B. Hayes[3], P. Kraft[1], U. Peters[4], A.T. Chan[2,5].* 1) Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 3) Division of Epidemiology, New York University School of Medicine, New York, NY; 4) Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, WA; 5) Division of Gastroenterology and Hepatology, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA.

The vast majority of colorectal cancers (CRC) arise from adenomatous polyp (adenoma) precursors. Although several genome-wide association studies (GWAS) have identified germline susceptibility loci for CRC, there are relatively little data characterizing genetic risk of adenomas. We therefore conducted a GWAS of advanced adenomas among participants enrolled in the Nurses' Health Study (NHS) and the Health Professionals Follow-up Study (HPFS) cohorts. Advanced adenoma cases (N = 513 in NHS, 314 in HPFS) had adenomatous polyps diagnosed on endoscopy with >1cm size and/or tubulovillous, villous, or high-grade dysplasia histological type. Controls (N = 578 in NHS, 351 in HPFS) had no adenoma diagnosis on colonoscopy, and were matched to cases on age, reason for endoscopy and time of most recent endoscopy. Genotyping was done using Illumina OmniExpress array and imputed to CEU reference panel (HapMap II release 24). Study-specific effect estimates were adjusted for age at blood draw and the top 3 eigenvectors and combined using an inverse-variance weighted meta-analysis. We also conducted an *in silico* meta-analysis GWAS of a total of 2,298 adenoma cases (all histological types and sizes) and 10,392 controls of European ancestry, from GWAS studies nested within NHS and HPFS, that were originally designed to study other outcomes, including type 2 diabetes, coronary heart disease, kidney stones, breast cancer, endometrial cancer, mammographic density, as well as controls from GWAS studies of CRC. The 1p31.1 locus, rs17520771 was associated with an odds ratio of 0.56 (0.46, 0.69; $p= 3.2\times10^{-8}$) for advanced adenoma. In addition, we found that CRC-susceptibility SNPs previously identified in GWAS were also associated with risk of advanced adenomas (rs4939827, SMAD7, p= 0.00014) and all adenomas (rs6983267, 8q24, p= 0.00036). However, after corrections for multiple comparisons, we did not identify significant associations between genotyped/ imputed SNPs and overall adenoma risk. In this case-control study nested within two large, population-based cohorts, we observed a possible association between the 1q31.1 locus and risk of advanced adenoma. Moreover, in an *in silico* analysis of GWAS data collected within other studies nested in these cohorts, we observed associations between established CRC susceptibility SNPs and adenoma risk. These data support the need for further studies examining genetic predisposition to different stages of colorectal tumorigenesis.

411

Large numbers of individuals required to classify and define risk for a rare VUS in known cancer risk genes. *B.H. Shirts[1], A.G. Jacobsen[1], G.P. Jarvik[2,3], B.L. Browning[2].* 1) Laboratory Medicine, University of Washington, Seattle, WA; 2) Medical Genetics, University of Washington, Seattle, WA; 3) Genome Sciences, University of Washington, Seattle WA.

Background: Up to half of unique genetic variants in genomic evaluations of familial cancer risk are rare variants of uncertain significance (VUS). Population studies indicate that as genomics enters the clinic additional rare variants will continuously be identified and that classification of rare VUS will be an ongoing issue. **Methods:** We modified standard power calculations to explore sample sizes necessary to classify and estimate relative risk for breast cancer and colon cancer variants responsible for varying levels of relative cancer risk. We required 80% power and tolerated a 10% false positive rate, since most variants clinically tested will be in known genes with high pretest probability of being deleterious based on in-silico results. We specifically evaluated variant population frequencies of 0.1%, 0.01% and 0.001%. Population-based studies were assumed to have equal numbers of cases and controls, and family-based studies were assumed to have equal proportions of first and second-degree relatives of carriers of the rare variant in question. We examined relative risk of breast cancer ranging from from 12 to 3 and relative risk of colon cancer ranging from 20 to 5. **Results:** Population-based samples necessary for classification of rare breast cancer variants ranged from 1,044 to 997,183 depending on relative risk and population frequency. Samples necessary for classification of rare colon cancer variants ranged from 661 to 425,633. The larger samples were necessary for rarer and less penetrant variants. Family based relative risk methods were robust to changes in variant population frequency and required at between 10 and 150 individuals, depending on penetrance. Even variants with moderately rare frequency (0.01%) and six-fold breast cancer risk required large population based samples or family based samples (27,387 and 31 respectively), samples larger than those used most large research studies. **Conclusion:** While the classification of the pathogenicity and associated risk of disease from variants is clinically important, we demonstrate that it may be unlikely that most rare missense variants will be classifiable in the near future and that accurate relative risk estimates may never be available for very rare variants. This knowledge may alter strategies for communicating information about variants of uncertain significance to patients.