550F

Frequency of gene usage and copy number variation within the rearranged Immunoglobin Heavy-Chain Variable locus based on immune repertoire sequencing. *M.J. Rieder[1], D. Williamson[1], A. Sherwood[1], R. Emerson[1], C. Desmarais[1], M. Chung[1], H. Robins[1,2], C. Carlson[1,2].* 1) Adaptive Biotechnologies, Seattle, WA; 2) Fred Hutchinson Cancer Research Center, Seattle, WA.

The human adaptive immune system is composed of both B and T cells that undergo somatic recombination at specific loci to create rearrangements of Variable (V), Diversity (D) and Joining (J) gene segments. For the B-cell immunoglobin receptor heavy-chain (IGH), the CDR3 regions are defined by the VDJ gene segments and nucleotide insertions/deletions at these junctions that create the vast sequence diversity of the IGH repertoire. Characterizing the germline DNA in these regions is impeded by the high sequence similarity between gene segments, mutation and copy-number variation (i.e. large insertions/deletions). Currently, there is a fundamental lack of information about the baseline IGH immune repertoire V gene usage and diversity within healthy human controls. To provide an estimate of this, we sequenced functionally recombined gene segments to infer the underlying gene structure. From a set of 132 healthy controls we sorted C19+/CD27+ B-cells from whole blood and amplified genomic DNA using a highly multiplexed PCR assay that targeted the rearranged IGH receptor locus. Following DNA sequencing and data processing to assign V, D and J gene families and names, we examined the usage frequency of IGHV gene segments across all individuals. We found that of the 98 V gene segments only 56 (57%) were used at a frequency > 0.1%, and ~10 showed little to no usage (present in <1% of individuals). This data also allowed us to identify two IGHV genes currently annotated as orphons (pseudogenes assigned to an alternate chromosomal location) that had unambiguous functional usage (IGHV4/OR15-8; IGHV3/OR16-09) and therefore must reside at the IGH locus on chromosome 14. Finally, by taking this functional approach we were able to screen all V gene segments for germline copy-number variation (e.g. large insertion/deletion events encompassing individual genes) by looking for an excess of deletion events or modal changes in gene usage. We confirmed that existence of 12 of 15 previously identified deleted IGHV gene segments. Strong deletion evidence was observed for an additional six IGHV genes (IGHV3-NL1, IGHV3-33, IGHV1-24, IGHV4-04, IGHV3-41, IGHV3-35) and ten with highly likely germline deletion events. These data suggest that functional immune profiling of rearranged immune receptors provides a more robust method of identifying individual structural variation and provides insight into the immune repertoire of healthy controls.

551W

Copy Number Variants near SLC2A9 are Associated with Hyperuricemia. *R.B. Scharpf[1], L. Mireles[2], E. Halper-Stromberg[3], A. Tin[2], A. Chakravarti[4], E. Boerwinkle[5], J. Coresh[2], W.H.L. Kao[2].* 1) Oncology, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Epidemiology, Johns Hopkins School of Public Health, Baltimore, MD; 3) Human Genetics, Johns Hopkins University School of Medicine, Baltimore, MD; 4) Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 5) IMM Center for Human Genetics, University of Texas School of Public Health, Houston, Texas.

Hyperuricemia is associated with multiple diseases, including gout, cardiovascular disease, and renal disease. Serum urate is highly heritable suggesting a strong genetic component, yet genome-wide association studies of single nucleotide polymorphisms (SNPs) and serum uric acid concentrations explain only a small fraction of the heritability. Whether common copy number variants (CNVs) contribute to uric acid levels is not known. Here, we use high-throughput genotyping arrays to assess DNA copy number on a genome-wide scale among 9,738 individuals of European ancestry who participated in the Atherosclerosis Risk in Communities (ARIC) study. Loss of DNA copy number at genomic coordinates 10,002,252-10,009,766bp on chromosome 4p16.1 is associated with a 5.39 percent increase of uric acid concentrations among women (95% CI: 4.32-6.47, p = 4.4e−24) and a 1.39 percent increase among men (95% CI: 0.30-2.49, p = 0.012). The CNV locus is approximately 200kb from several SNPs in the urate transporter SLC2A9 that have been previously associated with uric acid concentrations in ARIC, including SNP rs7675964. Among women with the same rs7675964 genotype, loss of copy number is associated with a 3.01 percentage increase (95% CI: 1.96-4.07) of uric acid concentrations (p = 4.84e-33). In addition to variation of DNA sequence, loss of DNA copy number may contribute to the genetic predisposition of serum uric acid concentrations, particularly among women, explaining some of the missing heritability from standard genome-wide analyses with SNPs.

552T

Comprehensive comparison of copy number variations detection using Illumina Omni 2.5M and Affymetrix CytoScan® arrays. *C. TANG[1,2], E. WONG[1], H. GUI[1], S. CHERNY[1,4], P. SHAM[1,2,4,5], P. TAM[3,5], M. GARCIA-BARCELÓ[3,5].* 1) Department of Psychiatry; 2) Centre for Genomic Sciences; 3) Department of Surgery; 4) State Key Laboratory of Brain and Cognitive Sciences; 5) Centre for Reproduction, Development and Growth, the University of Hong Kong, Pokfulam, Hong Kong.

Structural variation has been recognized as a genetic risk factor contributing to human diseases, and in particular, congenital disorders. Smaller scale copy number variations (CNVs) have also been linked to a number of neurodevelopmental phenotypes, including intellectual disability as well as autism spectrum disorders. The precise detection of CNVs is therefore necessary for understanding disease pathogenesis. Recently, the new generation of SNP-based arrays, Affymetrix CytoScan® and Illumina Omni 2.5M offer an unique opportunity for improved discovery of CNVs with their special design. We explored the performance of these new platforms by genotyping in duplicate on each platform, 4 samples from patients diagnosed with a congenital disease. Performance of the CNV calling was assessed on the basis of sensitivity and specificity, both within and across platforms using various CNV detection software. Similar to previous generations of SNP-based genotyping arrays, the concordance of CNVs was found to be moderate and dependent on the calling software. In general, Cytoscan offered higher sensitivity whereas more specific calls were achieved using Omni. To conclude, multiple CNV calling methods should be employed for reliable CNV calling.

553F

Characterisation of the *RNU2* CNV, a bulky neighbour for *BRCA1*. *C. Tessereau[1, 2], N. Monnet[1], M. Imbert[1], M. Buisson[1], L. Barjhoux[1], C. Cuenin[7], C. Schluth-Bolard[3, 4], D. Sanlaville[3, 4], Z. Herceg[7], E. Conseiller[2], M. Ceppi[2], L. Duret[5], OM. Sinilnikova[1, 6], S. Mazoyer[1].* 1) Genetics of Breast Cancer, Cancer Research Center of Lyon, CNRS UMR5286/Inserm U1052/ Université Lyon 1, Lyon, France; 2) Genomic Vision, Bagneux, France; 3) Service de Génétique, Laboratoire de Cytogénétique Constitutionnelle, Centre de Biologie et de Pathologie Est, Hospices Civils de Lyon, France; 4) INSERM U1028, CNRS UMR5292, Université Claude Bernard Lyon 1, Equipe TIGER, 69000 Lyon, France; 5) Laboratoire de Biométrie et Biologie Evolutive, Université de Lyon, Université Lyon 1, CNRS, INRIA, UMR5558, Villeurbanne, France; 6) Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Hospices Civils de Lyon, Centre Léon Bérard, Lyon, France; 7) Epigenetics Team, Molecular Carcinogenesis and Biomarkers Group, International Agency for Research on Cancer, F-69008, Lyon, France.

The question of the implication of multiallelic CNVs in complex traits remains largely open as most of them cannot be genotyped by array technology. In this work, we focused on the *RNU2* locus, a variable number of tandem repeats that contains the gene coding for the snRNA U2, an essential element of the splicing machinery. *RNU2* was shown many years ago to reside close to the breast cancer susceptibility gene *BRCA1* but is still missing from the latest human genome assembly and cannot therefore be investigated by recent genomic approaches. Using unassembled contigs, we precisely located *RNU2* within the chromosome 17 reference assembly, 124 kb telomeric of *BRCA1*. By FISH analyses on combed DNA (Molecular Combing), we determined more precisely the exact allelic number of repeats than with the previously used Pulse Field Gel Electrophoresis technique and found a range of 6-82 and a level of heterozygosity of 98% in 41 individuals. We used the 1,000 Genome Project data for analysing the variability of this macrosatellite by mapping sequence with unlocalized human genomic contigs and confirmed its high degree of polymorphism suggesting that depth-of-coverage calculation is a very useful tool for accurate multi-allelic CNV characterization. We found 24 frequent SNPs within the *RNU2* basic unit, and the genotype data for 1,106 individuals confirmed previous results showing a concerted evolution of this CNV. Thanks to our precise location, we were able to confirm that the *RNU2* array is within the *BRCA1* linkage disequilibrium block, which allowed us to study the *RNU2* array transmission over a large number of generations. A surprising resulting observation is that a highly variable locus can nevertheless be highly stable. Given the high level of polymorphism of this locus, we also measured the expression of U2 snRNA by qRT-PCR in 16 individuals carrying 36 to 110 repeats. Although the U2 level varied up to 6.6 times between individuals, it is not linked to the *RNU2* CNV copy number. By pyrosequencing, we found a higher level of methylation of the *RNU2* gene enhancer sequences in individuals with the highest copy number, suggesting that methylation could be involved in dosage compensation. These findings extend our knowledge of a recently neglected CNV that could be valuable for evaluating the potential role of structural variations in disease due to its location next to a major cancer susceptibility gene.