

## ASHG 2011: 1000 Genomes and TCGA progress, Ion Users Party!

Montreal October 14th, 2011. The data keep getting richer and deeper- Several presentations at the ICHG/ASHG meeting reviewed progress and provided a view of the surging scope of the large-scale genome projects that were launched a few years ago. For me and about 200 others, Ion Torrent's first User Group meeting and its most excellent party at the Intercontinental were a bonus to end the evening.

Gil McVean provided an update on genetic variation uncovered in the 1000 Genomes Project. He stated that the update was driven by integration of data and variant types, technologies, methodologies (statistical and quantitative tools), and validation. New variant and haplotypes integration files have been released by the consortium. On the 1000 Genomes website: "This October 2011 release represents an integrated set of variant calls and phased genotypes including SNPS, short INDELS and Deletions based on low coverage and exome sequencing data across 1092 individuals." (typo on website). The addition of exome data is clearly helping with variant discovery. Gil pointed out the huge increase in novel variant discovery going from the Pilot data (15.2M variants, 8.4M novel) to the Phase I, with 37.9M variants with 29.7M novel. The set now includes short INDELS and large deletions. The current genotype accuracy estimate is over 99% at chip heterozygous sites. Based on the data around rare variant discovery, he reported that sensitivity is around 96% to identify variants in new individuals.

In the first cancer session, Stacey Gabriel of the Broad Institute provided an overview of The Cancer Genome Atlas project. The goal was to sequence tumor/normal pairs from at least 25 forms of cancer. In addition to generating sequence data, the TCGA is compiling information across multiple data types, such as clinical diagnosis, treatment history, histologic diagnosis, etc, (see the TCGA website). The analysis is not being limited to exome sequencing (to be done on 100% of samples) or whole genome sequencing (only 10% of samples), but extends to approaches for obtaining SNP and CNV data (non DNaseq), methylation, mRNA and miRNA expression.

She revealed that 2,000 tumor samples with germline sequences are available. What are they learning? Some things are rather obvious: the somatic point mutation rate ranges over a thousand fold across different cancer types (i.e. lots in GBM, fewer in hematologic malignancies). Tissues exposed to carcinogens have the highest mutation rates, as reflected in variant data from melanoma and lung cancer samples. **In Partha Majumder's talk (NIBMG, India), he said that oral cancers are quite prevalent in Indian populations, thought to be due to chewing of the Areca nut (Betel nut).**

Daniel MacArthur (Wellcome Trust Sanger Institute) gave a interesting talk that highlighted problems in trusting current gene annotations when evaluating loss of function variants. He observed that loss-of-function variants have a much higher FP rate compared to other variant types.

He found two sources of error: reference sequence mistakes and annotation errors. His group will employ a new variation annotation tool, VAT, from Gerstein's lab at Yale.

The Ion Torrent User Group meeting opened with encouragement from Jonathan Rothberg to tackle the Grand Challenges set out earlier in the year. The throughput on the instrument from the 314, 316 and 318 chips are looking impressive. In the R&D lab, they are seeing output of 118MB, 650MB, and over 1G of AQ20 bases, respectively. The meeting featured several excellent talks from researchers using the platform for a variety of applications, including amplicon sequencing and miRNA studies. These researchers are also achieving significant output levels as methods are adapted and improved. Fair warning to the other platform/vendors: the user-based reports are extremely encouraging for this semiconductor-based sequencing technology.

WebLink: <http://www.realtimegenomics.com/Blog/ASHG-2011-1000-Genomes-and-TCGA-progress-Ion-Users-Party-1>