Genome **Medicine**

## MEETING REPORT

# Beyond the genome and into the clinic

David J Segal*

### Abstract

A report of BioMed Central's third annual Beyond the Genome conference, held at Harvard Medical School, Boston, September 27-29, 2012.

**Keywords** Bioinformatics, cancer, clinical diagnostics, epigenomics, ethics, human disease, human genomics, next-generation sequencing, pre-clinical models, rare variants.

Genetic differences among humans range from very common (minor allele frequency (MAF) nearly 0.5) to the very rare (MAF <0.001). In the pre-genomic era, the hunt for disease-causing variants was restricted to rare alleles of high penetrance, and required a careful analysis of the mode of inheritance. With the advent of high-throughput technologies, it became possible to perform genome-wide association studies on tens of thousands of people and to examine hundreds of thousands of common SNPs. It was therefore notable that a pervasive opinion at this year's meeting, articulated by keynote speaker Richard Gibbs (Baylor College of Medicine, USA) among others, was that we are clearly beyond the genome-wide association study, and that common SNPs, if they have any value for health, were not worth talking about on this occasion. The search is on once again for rare variants, only this time we wield the ever-increasing power of modern genomics.

## Bioinformatics for the masses (masses of data and masses of users)

Finding rare variants is not easy. The first day focused on the formidable bioinformatics challenges. Although the current capabilities to find SNPs in short-read sequences are robust, accurate calling of compound SNPs and the phasing of variants on chromosomes remains elusive. An important challenge today is identifying multinucleotide polymorphisms, such as small insertions and deletions

*Correspondence: djsegal@ucdavis.edu
Genome Center and Department of Biochemistry and Molecular Medicine, University of California, 451 Health Sciences Drive, Davis, CA 95616, USA

**BioMed** Central

(indels), because only part of an indel-spanning read will map perfectly to the reference genome (Gabor Marth, Boston College, USA). A related challenge is repeated sequences: the 'dark matter' that makes up most of the human genome and the Achilles heel of short-read sequencing. One solution is to include long, although currently error-prone, reads from platforms such as PacBio in such a way that the combined data reduce the errors and span the gaps (Mike Schatz, Cold Spring Harbor Laboratories, USA).

The challenge for tomorrow is robust identification of copy number variants, important for medicine but currently difficult to characterize with precision using automated, unsupervised methods. Such methods are the goal of the modern tool developer. To get informatics tools out of the hands of data analysts and into the hands of data generators will require the tools to be easy, robust and reproducible. The session ended with a bioinformatic challenge for the audience. Conference attendees were challenged to identify a famous quote that had been encoded as DNA sequence and inserted into an unknown genome, which itself required *de novo* assembly from a set of short reads. As a testament to what can be accomplished with today's skills and tools, the prize was won in under an hour.

## What to do about the many rare variants

Clearly some rare variants have a large effect on disease. The problem is there are too many rare variants. The number of SNPs with MAF 0.001 to 0.5 has plateaued at around 36 million, but the number of SNPs found with an MAF <0.001 is soaring. Speakers on the second day grappled with the uncertainty and implications of finding such variants on a medical, ethical and legal level. The observation that about 500 missense mutations and 100 loss-of-function mutations exist in apparently healthy people raises important ethical questions about returning this information to patients, especially in the frequent case that such alleles are incidental to the patient's original health concern. An important concept to emerge was that better predictions of a variant's health consequence could be made by incorporating additional knowledge or 'priors', such as family history, known protein and gene interaction networks, gene expression data, and comparison with genomes of healthy individuals (Ben

Raphael, Brown University, USA; Josh Stuart, University of California Santa Cruz, USA; Lynn Jorde, University of Utah, USA; Daniel MacArthur, Massachusetts General Hospital, USA; John Carpten, TGen, USA).

Some portion of variant calls will be wrong because of wrong annotations, wrong priors (including the use of healthy controls that were not actually healthy), errors in sequencing, and errors due to multiple testing (Isaac Kohane, Harvard Medical School and Children's Hospital, USA; James Lupski, Baylor College of Medicine, USA). Several strategies were put forward to guard against such errors in the clinic. Leslie Biesecker (National Human Genome Research Institute, USA) suggested a provocative paradigm shift from studying the genetics of people with known disease to studying the diseases of people with known genetics. Sharon Plon (Baylor College of Medicine, USA) and others cited the use of genetics review panels, not simply bioinformatics pipelines, to decide what information rises to the rigor of clinical disclosure. Some clinical testing paradigms avoid the time, cost, errors and ethics of large-scale multiple testing by focusing on only small panels of 'actionable' variants (estimated to encompass as few as 47 to 57 clinically actionable genes; Elizabeth Worthey, Medical College of Wisconsin, USA; Yost Shawn, University of California San Diego, USA). Database annotations are improving (newer ones are improving faster than the older, more established ones), and an interesting suggestion was made that variants in databases should include fields for 'user feedback'. Gibbs closed the session by forecasting that genomics in the next 10 years would see the most application in Mendelian disorders, cancer and technology development, and less in studying complex diseases or healthy adults.

## Beyond Mendel: other uses of the genome to improve public health

The theme of the third day was the use of genomics to study more dynamic processes. Most cells of the body contain the same genome, but they assume different functional roles because of the genes they transcribe. Gene transcription in each cell is controlled by the epigenome, and each cell type contains a different epigenome. Two talks discussed what information transcription factors use to determine which genes they regulate (Frank Pugh, Penn State University, USA), and how the three-dimensional architecture of the genome determines which genes can interact (Job Dekker University of Massachusetts Medical School, USA). Other talks described changes in DNA methylation patterns that were dependent on factors such as age, tissue, disease states such as cancer, and the location within the three-dimensional architecture of the genome (Peter Laird,

University of Southern California, USA; Vardhman Rakyan, Barts and The London School of Medicine and Dentistry, UK). In many cases, the effects of these changes on gene expression were unclear. However, in some cases the changes had unambiguous consequences for gene expression and could therefore be predictive of, for example, cancer phenotypes.

Cancer is unlike most other diseases in that the genome itself is dynamic and changes during disease progression. Genomics can provide a window on the clonality of cancer lineages and the heterogeneity within tumors (Samuel Aparicio, BC Cancer Research Center, Canada; Arend Sidow, Stanford University, USA). These intensive genomics and systems biology efforts are also being applied in the treatment of cancer. Genome-wide methods were used to identify the right target, choose the right drug, and will ultimately be used to find the right patients (those with compatible genetics). A critical component of these efforts was the assay system. In developing inhibitors to the oncoprotein NRAS in melanoma, Lynda Chin (MD Anderson Cancer Center, USA) took the perspective that cells in a real tumor behave quite differently from cells in a culture dish or xenograft. The most informative pre-clinical assays need to recapitulate the many facets of cancer in a person, such as a mouse model in which tumors arise endogenously from an inducible *NRas* gene. On the opposite end of the spectrum, keynote speaker Stuart Schreiber (Broad Institute of Harvard and MIT, USA) used cancer cell lines, but a lot of them. Nine hundred cell lines that had been carefully characterized genetically and transcriptionally were assayed to identify factors controlling their susceptibility or resistance to 480 chemical compounds at 16 different concentrations. A subset of interesting cases was further screened against a library of 30,000 compounds. From these studies emerged a wealth of provocative hypotheses. Although many advances were reported throughout the conference, these latter talks were perhaps the most inspiring in that they showed how genomic information could be used not just to understand or diagnose disease but to actually improve upon the treatments of today, truly moving us towards the goals that lie beyond the genome.