

MEETING REPORT

Learning to swim in a sea of genomic data

Ekta Khurana^{1,2}

Abstract

A report on the 63rd American Society of Human Genetics (ASHG) meeting held in Boston, USA, 22–26 October 2013.

Meeting report

Advances in sequencing technology have provided a tremendous boost to human genetics research. Thus, it is not a surprise that attendance at the American Society of Human Genetics (ASHG) meeting has grown rapidly over the last few years. Aravinda Chakravarti (Johns Hopkins University, USA), recipient of the William Allan Award, noted that while about 600 people attended the meeting in 1977, this year the attendance was about ten times more. The immense growth in the field of human genetics was also exemplified by Jeff Murray (University of Iowa, USA) in his presidential address, where he contrasted how the discovery of millions of single nucleotide polymorphisms (SNPs) every day is routine now, whereas the discovery of a single SNP was publication-worthy in 1983.

The scope of the meeting was far beyond what can be covered in a short report, so I will focus on two major themes related to medical applicability of whole-genome sequencing data, including understanding the functional impact of variants and the importance of data sharing.

Going beyond sequence to function

Whole-genome sequencing has provided a sea of sequence variants, and we face the challenge of identifying the key disease variants among these. Understanding the functional impact of variants is crucial to eventually home in on the biological mechanisms leading to disease. Multiple sessions in the meeting focused on the functional interpretation of sequence variants. In the session titled ‘Variants, variants everywhere’, a multitude of integrative methods were presented to prioritize variants in disease

studies. Haiyuan Yu (Cornell University, USA) showed how a structurally resolved protein interactome can provide novel insights into disease variation. For example, the guilt-by-association principle that interacting proteins tend to be associated with the same disease does not apply to dominant mutations. A likely explanation for this observation is that dominant mutations tend to be gain-of-function and mostly cause the gain of different functions in interacting proteins, as opposed to recessive mutations that tend to be loss-of-function and disrupt the same interaction between two proteins.

We have known for a long time that noncoding genomic regions play a big role in human disease. However, the focus of most previous systematic approaches to uncover disease variants has been on coding regions. It was evident during the meeting that we have come a long way in understanding the noncoding part of the genome. I presented our approach called FunSeq, which integrates data from the 1000 Genomes and ENCODE consortia to prioritize noncoding variants in disease studies. Emmanouil T. Dermitzakis (University of Geneva, Switzerland) and Tuuli Lappalainen (Stanford University, USA) discussed finding causal regulatory variants by combining genomic and transcriptomic sequencing data. To identify causal variants, it is important to know the cell types in which they perturb regulatory processes. Nancy Cox (University of Chicago, USA) described transcriptome studies across multiple tissues and their application to common diseases on behalf of the Genotype-Tissue Expression (GTEx) consortium. Cox noted that, based on current results, the most physically proximal or overlapping gene is rarely the most strongly linked gene for trait-associated GTEx eQTLs. John Stamatoyannopoulos (University of Washington, USA) also presented DNase I hypersensitivity data from hundreds of cell/tissue types and developmental states. These talks highlighted the importance of integrating functional genomics studies with whole-genome sequencing data to uncover the effects of sequence variants.

When interpreting sequence variants in a functional context, it is important to be cautious as these results can have real-life implications. While discussing mutations

Correspondence: ekta.khurana@yale.edu

¹Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT 06520, USA

²Molecular Biophysics and Biochemistry Department, Yale University, New Haven, CT 06520, USA

that predispose to cancer, Nazneen Rahman (Institute of Cancer Research, UK) rightly reminded us that people make big decisions and have parts of their anatomy removed based on their genotype. In a different vein, Heng Li (Broad Institute, USA) brought to our attention the effects of errors in the reference sequence in next-generation sequencing (NGS) analysis. Many talks focused on technical details and new methods of variant calling from NGS data. Erik Garrison (Boston College, USA) presented a novel approach that uses a graph reference built from known variants to improve future variant calling. Additionally, Monkol Lek (Massachusetts General Hospital, USA) discussed the advantages and challenges of joint variant calling in a massive set of more than 50,000 exomes.

One of the biggest challenges of the post-genomics era: effective data sharing

The week of the ASHG meeting overlapped with open access week, befitting one of the major meeting themes: data sharing. A message that was echoed by many speakers was the need to develop new ways of sharing data effectively. Sharing genetic variation data between clinicians and researchers will help advance our understanding of disease and lead to better patient care. Since the number of people getting their whole genomes sequenced is continuously increasing, Nathan Pearson (Ingenuity Systems, USA) presented the Empowered Genome Community, an initiative to help people share their genomes with each other and with scientists through an online genome interpretation application called Ingenuity Variant Analysis.

Mendelian diseases were the focus of many discussions involving data sharing. Debbie Nickerson (University of Washington School of Medicine, USA) noted that there are more than 7,500 conditions in the Online Mendelian Inheritance in Man (OMIM) database and approximately 3,500 remain unsolved. These include the very rare conditions where data sharing seems essential, but there is currently no unified database for patients with unsolved Mendelian diseases. Heidi Rehm (Partners Center for Personalized Genetic Medicine, USA) and Ada Hamosh (Johns Hopkins University School of Medicine, USA) discussed the 'genomic matchmaker' collaboration to enable cross-talk between clinical geneticists across the world by creating the broadest source of unsolved exomes and genomes that could be matched with other unsolved cases. Rehm said, 'There is a critical need in the community to enable groups to collaborate around building evidence for candidate genes for novel genetic disorders. Several groups have developed the capability to match phenotypes and candidate genes from unsolved cases in their systems; however, the success of a match is highly dependent on the wealth of cases from which one can draw.' In relation to this, many talks referred to Daniel MacArthur's

(Massachusetts General Hospital, USA) collection of exome data from more than 50,000 patients.

Another prominent discussion focused on different viewpoints about sharing secondary or incidental findings (results that are unrelated to the reason for the test or the research study but are relevant to the participant's future health) with patients and research study participants. Anna Middleton (Wellcome Trust Sanger Institute, UK) discussed international views on sharing incidental findings in research studies, and began her presentation by showing the picture of Sarah, a three year old girl who has an undiagnosed developmental disorder. After sequencing Sarah's genome, they found some pertinent mutations in genes implicated in developmental disorders, but they also had an incidental finding of a mutation in *BRCA1*. In an online survey conducted by Middleton and colleagues, the majority of people felt that information about life-threatening, untreatable conditions should be shared. It should be noted that for clinical settings, the American College of Medical Genetics and Genomics recommended earlier this year that a set of 57 genes should be evaluated and reported when a patient's exome or genome sequence is ordered.

Conclusions

Perhaps the biggest eureka moment of the meeting was when Jeanne Lawrence (University of Massachusetts Medical School, USA) convincingly showed that her lab was able to silence one entire copy of chromosome 21 in stem cells *in vitro*. Trisomy 21 or Down's syndrome is caused by an extra copy of chromosome 21. Using genome editing with zinc finger nucleases, Lawrence and colleagues inserted *XIST* (human X-inactivation gene) into chromosome 21 in stem cells with trisomy 21. They then showed using eight different methods that a single copy of the chromosome had indeed been silenced.

Thus, the tremendous advances in human genetics are clear. Moving forward into the future, a unifying message of the meeting was the need to integrate data from different resources. In line with this message, the final symposium consisted of Marc Vidal (Dana-Farber Cancer Institute, USA), Aviv Regev (Broad Institute, USA) and Gary Nolan (Stanford University, USA) discussing the marriage between genetics, systems biology and immunology. We heard how single-cell approaches can help tease apart cellular heterogeneity and enable a better understanding of cell biology. Advances in single-cell technologies are expected to greatly help us understand complex processes, such as cancer evolution and heterogeneity in immune response.

Abbreviations

NGS: Next-generation sequencing; SNP: Single nucleotide polymorphism.

Competing interests

The author declares she has no competing interests.

Acknowledgements

The author thanks Chris Tyler-Smith (Wellcome Trust Sanger Institute, UK), Jieming Chen and Suganthi Balasubramanian (Yale University, USA) for comments and suggestions.

Published: 6 December 2013

doi:10.1186/gb4144

Cite this article as: Khurana: Learning to swim in a sea of genomic data.
Genome Biology 2013 **14**:315.