Genome Biology

MEETING REPORT

Genomics through the lens of next-generation sequencing

John A Capra[1], Lucia Carbone[2], Samantha J Riesenfeld[1] and Jeffrey D Wall[3]*

Abstract

A report on the 23rd annual meeting on 'The Biology of Genomes', 11-15 May 2010, Cold Spring Harbor, USA.

Recent advances in high-throughput sequencing technologies have greatly increased the scale and scope of genomics research, and this was evident throughout the recent Biology of Genomes meeting at the Cold Spring Harbor Laboratory. Here we describe some highlights of the meeting.

Functional and cancer genomics

In one of several talks that investigated the causes, dynamics and phenotypic effects of regulatory change, Mike Snyder (Stanford University, Stanford, USA) used chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) to examine the variability in transcription factor binding among individuals in yeast (*Saccharomyces cerevisiae*) and human. In both species, significant variation was observed, and the amount of binding was strongly correlated with gene expression. Much of the observed binding variation could be associated with specific single nucleotide polymorphisms (SNPs) and structural changes to the genome. On the basis of the patterns of variation, Snyder suggested that gene regulation may work like a government, with global regulators and local regulators all having strong effects, but with some focused on a more limited set of loci.

Snyder's talk introduced two themes that appeared throughout the meeting: the widespread adoption of high-throughput sequencing as an analysis strategy; and a focus on identifying and understanding regulatory elements. Axel Visel (Lawrence Berkeley National Laboratory, Berkeley, USA) continued these themes in a talk that highlighted the limitations of using comparative genomics to identify enhancers. By performing ChIP-seq with the enhancer-associated p300 protein on mouse forebrain and embryonic heart tissue, he and colleagues identified a large number of heart enhancers with very low evolutionary conservation compared to forebrain enhancers. This surprising result suggests that deep pathway conservation does not imply regulatory sequence conservation, that enhancer conservation is not predictive of function, and that there are global differences in enhancer conservation between tissues.

Li Ding (Washington University, St Louis, USA) described how sequencing samples from the same patients at different stages of the same cancer can help track changes that have occurred during cancer progression, and possibly lead to improved drug therapies. Thanks to an efficient pipeline, their genome-sequencing center can analyze tumor/pair samples in only 12 days. The analysis of about 150 cancer genomes using this pipeline has enabled comparisons of different cancer genomes from different points of view, including mutation rate, mutation spectrum, copy number variation and structural variation. The results showed by Elaine Mardis (Washington University, St Louis, USA) are an example of how powerful these tools are and what they are able to achieve. She presented the analysis of the relapse genome of an acute myelogenous leukemia patient and a comparison with the genome sequenced at initial presentation. Interestingly, this study was able to pinpoint relapse-specific mutations most likely involved in disease progression.

Complex trait mapping

One of the largest open issues in human genetics deals with the question of 'missing heritability': given the generally high estimates of heritability for many complex traits (such as genetic susceptibility to complex diseases), why have genome-wide association studies (GWAS) identified variants that explain only a small fraction of the heritable variation we know is out there? Several talks explored this question using a range of experimental and theoretical approaches. One hypothesis suggests that rare variants of large effect, which will generally be

*Correspondence: wallj@humgen.ucsf.edu
[3]Institute for Human Genetics, University of California San Francisco, 513 Parnassus Ave, San Francisco, CA 94143, USA
Full list of author information is available at the end of the article

missed by GWAS, are a crucial component of this missing variability. Richard Durbin (Wellcome Trust Sanger Institute, Cambridge, UK) and others described progress on the 1000 Genomes Project, which by next year will generate low-coverage (around 4x) whole-genome sequence data from more than 2,000 individuals. This dataset, in conjunction with new imputation algorithms for base-calling low-coverage data, will provide a near-complete catalog of rarer variants (for example, minor allele frequency ≥0.005) across the human genome, which in turn will facilitate efforts to identify rare variants affecting disease susceptibility.

Jeffrey Barrett (Wellcome Trust Sanger Institute, Cambridge, UK) addressed the subject of 'synthetic associations'. It has been proposed that many GWAS hits are not the result of common variants of modest effect, but rather are artifacts caused by linkage to multiple rare (but highly penetrant) variants. Barrett's talk outlined several compelling sources of evidence suggesting that these synthetic associations are likely to be quite rare. So, while rare variants may or may not explain the 'missing heritability' problem, they are not a probable cause of the associations already discovered by GWAS.

One alternative approach to understanding the genetic architecture of complex traits is to use a more tractable genetic system. Barak Cohen (Washington University, St Louis, USA) presented detailed genetic analyses of sporulation efficiency in the yeast *S. cerevisiae*. For this phenotype, just four SNPs (located in three transcription factor genes) combine to explain 87% of the total phenotypic variation, although very little of this (around 25%) could be ascribed to additive effects. Cohen also described a thermodynamic model that might explain the strong interactions (epistasis) observed among SNPs. This and other work raises the possibility that gene-gene interactions may be a large part of the answer to the missing heritability question.

Evolutionary genomics

Next-generation sequencing technologies now enable researchers not affiliated with genome centers to conduct their own genome-sequencing projects. Peter Donnelly (Oxford University, UK) described some preliminary findings from the PanMap project, a collaborative effort to sequence and analyze the genomes of ten Western chimpanzees (*Pan troglodytes verus*). Donnelly and colleagues were especially interested in the evolution of recombination rates, and the recent fusion between chimpanzee chromosomes 2a and 2b can be used as a 'natural experiment' to estimate if and how chromosome position influences recombination rate. The results suggest that recombination rates are more affected by chromosomal position than they are by local sequence context.

Another exciting genome project was described by Svante Pääbo (Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany). As a follow-up to the recent publication of the Neanderthal genome, Pääbo and colleagues have now generated roughly 2x coverage of the whole genome from an unclassified hominin fossil found in Denisova Cave in southern Siberia. Preliminary results suggest that the Denisova fossil is more closely related to Neanderthals than to modern humans, though the divergence between Neanderthal and the Denisova fossil is larger than the divergence between any two extant modern humans. Further studies will be needed to clarify the precise evolutionary relationships between this fossil and other hominin groups.

New directions

Several talks gave exciting glimpses of how next-generation sequencing technology can enable novel, high-throughput experimental analyses. Rob Mitra (Washington University School of Medicine, St Louis, USA) introduced a promising new technology for investigating the regulatory networks of development across a cell lineage. By attaching a transposase to a transcription factor, Mitra forces the insertion of a transposon 'calling card' near the site of DNA-binding events. These transposons, and thus binding sites, can then be identified by next-generation sequencing. As these transposons survive cell divisions, the binding history of a transcription factor can be traced though a cell lineage. The method has been applied successfully in yeast and tests are in progress in vertebrates. If successful, this approach would yield a powerful tool for decoding how regulation drives tissue-specific development.

Although the meeting focused primarily on humans and model organisms, considerable attention was given to the large quantities of diverse data being generated by the sequencing of microbial communities through projects such as the Global Ocean Sampling Expedition and the Human Microbiome Project (HMP). Katherine Pollard (Gladstone Institutes, University of California, San Francisco, USA) argued that traditional approaches to genomic analysis must be significantly adapted to take advantage of the new kinds of information in metagenomic data, which is produced by shotgun sequencing the DNA extracted from environmental samples. She showed that phylogenies inferred from metagenomic sequence reads allow new ways of defining species, such as Operational Taxonomic Units (OTUs), of discovering novel OTUs, of defining and comparing microbial community diversity, and of estimating microbial ranges in geographic and niche spaces.

In regard to the human microbiome, Jennifer Wortman (University of Maryland School of Medicine, Baltimore, USA) emphasized the HMP's goal of discovering

potential correlations between changes in microbial community composition and the health of the human host. Referencing studies of the vaginal and gut microbiomes, she showed that different types of communities require bioinformatic tools with different levels of resolution and specialization. By sequencing the complete genomes of several closely related microbes collected from coastal ocean populations, B Jesse Shapiro (Massachusetts Institute of Technology, Cambridge, USA) took a step towards understanding microbial speciation with his presentation of a well-supported sympatric model of speciation in which populations are ecologically differentiated by a set of niche-specific genes.

Finally, James Taylor (Emory University, Atlanta, USA) offered an integrated vision of how we might aim to do science in the age of next-generation sequencing. He emphasized two fundamental directions: first, increasing access to the ability to perform large-scale computational analyses; and second, and perhaps more important, ensuring that such analyses are done in a way that supports and encourages the integrity of scientific investigation. Taylor demonstrated by an analysis of a mitochondrial genome resequencing experiment that commercial cloud computing platforms can be used in conjunction with Galaxy, a web-based genome analysis tool, to facilitate large-scale analyses that potentially involve multiple software programs, while maintaining the transparent provenance of the data and parameters. The resulting record of every step of the workflow guarantees that an analysis is reproducible and can be clearly communicated. Taylor also noted that in order to completely guarantee reproducibility, the original data themselves must be stored permanently, which raises another challenge for this new scientific paradigm.

New experimental technologies are giving individual labs the opportunity to conduct large-scale genomic studies that were unimaginable just a few years ago. However, the data generated on this scale present new challenges in interpretation, analysis and data management. Given the quality of the science presented at this meeting, we are confident that the community will find creative and collaborative solutions for these issues.

Author details
[1]Gladstone Institute of Cardiovascular Disease, University of California San Francisco, 1650 Owens Street, San Francisco, CA 94158, USA. [2]Childrens Hospital of Oakland Research Institute, 5700 Martin Luther King Jr Way, Oakland, CA 94609, USA. [3]Institute for Human Genetics, University of California San Francisco, 513 Parnassus Ave, San Francisco, CA 94143, USA.

Published: 25 June 2010

doi:10.1186/gb-2010-11-6-306
Cite this article as: Capra JA, *et al*.: Genomics through the lens of next-generation sequencing. *Genome Biology* 2010, **11**:306.