

MEETING REPORT

Broadening horizons: holistic viewpoints from the Biology of Genomes

Ambrose Carr and Dana Pe'er*

Abstract

A report on the Cold Spring Harbor Laboratory 27th annual meeting on the Biology of Genomes, held in Cold Spring Harbor, New York, USA, 6-10 May 2014.

Fine mapping in *cis* discovers substantial functional regulatory variation

'Biology of Genomes' is a broad annual meeting that showcases the year's progress in the many subfields of genomics. With so many talks, it is impossible to cover all of them. However, despite the variety of subject matter, several trends recurred throughout the conference.

One result that consistently resurfaced across multiple sessions is the ability of enlarged studies of tens of thousands of individuals to move beyond linkage to genes and finely map traits to individual regulatory polymorphisms. Jeffrey Barret (Wellcome Trust Sanger Institute, Hinxton, UK) was able to identify specific genes as contributors to disease by using data from 12,000 patients in the UK10K project. Hailiang Hailiang (MIT, USA) presented results from the inflammatory bowel disease (IBD) consortium that allowed the identification of specific variants that contribute to IBD, and Stephen Parker (NIH Bethesda, USA) presented the results of the ongoing FUSION study of type 2 diabetes, where a combination of RNA-seq and genotyping was used to identify muscle expression quantitative trait loci (eQTLs) associated with different physiological stages of type 2 diabetes.

With the enormous amount of data being generated, it is logical to ask: how much functional variation is found in coding versus noncoding regions of the genome? Alexander Gusev (Harvard University, USA) answered this question by annotating the genome using data from Encyclopedia of DNA Elements

(ENCODE) and Roadmap Epigenomics projects. He found substantial enrichment of psychophrenia QTLs in DNase hypersensitive sites, whereas there were essentially no QTLs in intron and intergenic sites. Together, these studies suggest that regulatory variation is responsible for a large amount of functional variation, and we are now capable of finding many concrete examples in *cis*, where chromatin state influences the cell or organism phenotype.

Even with these massive studies, there was still no mention of *trans*-QTL discoveries, suggesting that even use of tens of thousands of individuals leaves studies under-powered to locate distal effectors. Luke Jostins (Wellcome Trust Centre for Human Genetics, Oxford, UK) supported this hypothesis directly by reporting an inability to detect *trans*-eQTLs despite using a model specifically designed to detect them.

Transcriptional and translational regulation

Jacob Degner (European Molecular Biology Lab, Heidelberg, Germany) finely examined one method by which regulatory variation can have a functional impact on gene expression by using cap analysis of gene expression (CAGE), a method that precisely identifies transcriptional start sites (TSSs), on 80 cell lines at three time-points. By integrating genotype information, Degner defined promoter-shape QTLs, identifying two main classes of variants: those that cause the precise location of transcription initiation to vary at a higher rate, and those that shift the location of transcriptional initiation to a new position. It is not yet clear how these QTLs influence expression, but it suggests at least two mechanisms by which regulatory variation might regulate the 5' mRNA, which could have downstream effects on mRNA secondary structure, splicing and stability.

Beyond transcription, there are many downstream regulatory layers that are under genetic influence. Protein translation and protein expression are two such layers that received substantial attention. Alexis

* Correspondence: dpeer@biology.columbia.edu
Departments of Biological Sciences and Systems Biology, Columbia University, 1130 St Nicholas Avenue, New York, NY 10032, USA

Battle (Stanford University, USA) and Yoav Gilad (University of Chicago, USA) presented two sides of a collaboration that examines this question by comparing the incidence of eQTLs, 'ribo-QTLs' and 'protein-QTLs', the latter of which were measured by comparing genotype information with ribosome profiling and high-throughput 'stable isotope labeling with amino acids in cell culture' (SILAC) mass spectrometry. The results were heartening, showing that at least half of the functional variation that evokes an eQTL also associates with a protein-QTL. This indicates that RNA-based analyses are capturing a substantial portion of the cell phenotype. However, significantly, it shows the promise and importance of protein-based technologies that are capable of generating more-representative images of cell phenotypes.

Mutations and deletions

Clustered regularly interspaced short palindromic repeats (CRISPRs) were one of the big stories of the past year, bringing major improvements in the ability of scientists inexpensively to achieve complete gene knockout disruption. Neville Sanjana (Broad Institute of Harvard and MIT, USA) presented a genome-scale CRISPR screen, over every gene in the genome, developed in Feng Zhang's group (Broad Institute, USA). They screened melanoma cell lines treated with vemurafenib, a highly effective late-stage melanoma drug against which most patients eventually develop resistance. In their screen, they identified many genes known to contribute to vemurafenib resistance, as well as some new candidates, highlighting the potential of CRISPR screens in drug discovery.

Several other groups asked questions about what can be learned from examining natural mutations. Michael Stratton (The Wellcome Trust Sanger Institute, Hinxton, UK) catalogued all the different mutation types present in the *TP53* gene across numerous tissues and cell lines and compared the mutation types with the effects of known mutagens. He found that over 50% of the mutations could be attributed to known mutagens such as UV light or aflatoxin and was able to identify the percentage of mutations that each mutagen was responsible for. However, much of the mutation load could not be explained by his panel, suggesting that there are novel mechanisms of DNA mutation that have not yet been investigated.

Minyoung Wyman (Columbia University, USA) leveraged mutation rates in an innovative way by comparing the rate at which germ cells accrue two different sets of mutations: mutations resulting from mitosis that require cell division, and non-mitotic mutations. As germ cells do not divide between birth and puberty, the two mutation types have different rates, and the historical

difference in these rates allowed Wyman to estimate the average generation time throughout human history: a slow increase from 25 to 29 years.

Computation and technology

Perhaps the most consistent message throughout the conference was that, as data increase in magnitude, their utility is limited by the sophistication of our computational approaches. There were several excellent presentations that highlighted this fact, perhaps none more strongly than that of Joseph Pickrell (New York Genome Center, USA), who showed that, by using factor analysis to integrate multiple correlated variables, much more informative eQTLs can be discovered and used to distinguish causative linkages between symptoms and comorbid ones caused by the same functional variant. Dana Pe'er (Columbia University, USA) showed how integrating multiple data types improves statistical power enough to detect novel driver mutations that are missed by any data type alone.

Matthew Stephens (University of Chicago, USA) reexamined the foundations of a ubiquitous statistic: the false discovery rate (FDR). Stephens showed that the bimodal null model using common FDR algorithms has the effect of depressing the corrected *P*-values output by the procedure. Therefore, we are likely over-estimating our experiments' false discovery rates. Finally, Oliver Stegle (European Bioinformatics Institute, Hinxton, UK) examined the effect of the cell cycle on gene expression in single-cell sequencing. Unlike bulk expression analyses, which average over a population of cells, single-cell sequencing is very sensitive to the cell cycle, and Stegle demonstrated how an algorithm that corrects for cell-cycle stage reveals hidden information on the stage of differentiation in embryonic stem cells. Together, these talks demonstrate the importance of modeling the sources of variance in biological experiments and the power of integrative analyses to uncover biologically meaningful phenomena obscured under any single assay.

Conclusions

Holistically, Biology of Genomes 2014 provided a broad view into many cutting-edge projects. Promising results were reported from many corners of genomes: for example, large studies leveraged their improved power to identify specific regulatory SNPs; functional variation that affects expression corresponded to matched variation in protein expression in more than half of the cases; some early results of genome-wide deletion screens successfully identified genes involved in drug resistance; and new computational approaches were able to integrate multiple assays profitably to uncover a whole greater than its parts.

Abbreviations

CAGE: Cap analysis of gene expression; CRISPR: Clustered regularly interspaced short palindromic repeat; eQTL: Expression quantitative trait locus; FDR: False discovery rate; IBD: Inflammatory bowel disease; SILAC: Stable isotope labeling with amino acids in cell culture; TSS: Transcriptional start site.

Competing interests

The authors declare that they have no competing interests.

Published online: 25 July 2014

doi:10.1186/s13059-014-0416-0

Cite this article as: Carr and Pe'er: Broadening horizons: holistic viewpoints from the Biology of Genomes. *Genome Biology* 2014 15:416.