

MEETING REPORT

Dispatches from the functional phase of genome biology

Luke Jostins*

Abstract

A report on the 25th annual meeting on The Biology of Genomes, Cold Spring Harbor, USA, 8-12 May 2012.

Keywords High-throughput genomics, genome regulation, non-humans, complex traits.

The annual Biology of Genomes conference at Cold Spring Harbor has long been an important conference for those with an interest in the science of genomes. It has acted as a focal point for many large collaborative projects, including the HapMap and 1000 Genomes projects, both of which held (and, in the case of the 1000 Genomes Project, continue to hold) pre-meetings in the days prior to the conference. Its small size, intense program and broad remit ensure that much ground is covered within the general subject of genomics. Whether at the talks, during the pre-meetings or at the bar, it is a good place to assess the state of play of genomics, and hazard guesses at where the field is heading.

This report does not purport to be a complete survey of the topics discussed at the conference, and in particular I will not discuss the many interesting talks on evolutionary and population genetics. Instead, I wish to describe a set of talks that highlight two important themes that emerged from this conference. The first is how high-throughput experimental techniques are giving greater resolution to the dynamics of genome regulation, and the second is the rapid progress being made in the genetics of non-human complex traits using genome-wide methods.

The nitty-gritty of genome regulation

For a number of years the Biology of Genomes conference has focused on understanding genome regulation. Last year a large number of talks were given on cataloguing

and classifying DNA regulatory elements, with the Encyclopedia of DNA Elements (ENCODE) project presenting initial results of its large scan of non-coding functional elements. This year, ENCODE made another strong showing, with Ewan Birney (European Bioinformatics Institute, UK) updating us on the progress of the project as it approaches publication, with additional talks from Mark Gerstein (Yale University, USA) and Michael Hoffman (University of Washington, USA) describing methods for constructing regulatory networks and functional partitions using ENCODE data.

Beyond cataloguing regulatory elements, many talks presented new experimental studies that dig deeper into the nitty-gritty dynamics of genome regulation. These high-throughput studies looked at the relationship between different genetic and epigenetic marks, and started putting together detailed models of what drives gene regulation in health and disease.

Daniel Gaffney (Wellcome Trust Sanger Institute, UK) presented an analysis of the positioning of nucleosomes (the basic unit of eukaryotic chromatin) in seven lymphoblastoid cell lines, specifically looking for cases where nucleosomes are placed in systematic locations across cells, rather than being randomly placed. Gaffney reported that consistent positioning is the rule rather than the exception, with both sequence context and presence of bound proteins regulating nucleosome location. Another talk by Robi Mitra (Washington University, USA) emphasized the regulatory importance of nucleosomes, showing that the positioning of nucleosomes over promoters is a strong predictor of variation in gene expression in yeast. Mitra noted that transcriptional noise might be adaptive in certain circumstances, and showed evidence that cells preferentially place nucleosomes to drive expression variance in response to ethanol-induced stress.

Other talks reported attempts to dissect out the drivers of expression in model systems. Barak Cohen (Washington University School of Medicine, USA) reported the use of a synthetic promoter system (described at last year's conference in the context of non-additive *cis*-regulation) to study the effect of sequence changes in the promoter of rhodopsin on gene expression. Cohen

*Correspondence: lj4@sanger.ac.uk
Statistical and Computational Genetics, Wellcome Trust Sanger Institute,
Cambridge CB10 1HH, UK

showed that a minority of variance in expression was driven by changes in the affinity of the bound transcription factor, and demonstrated an important role of competitive binding within the promoter. Jason Gertz (HudsonAlpha Institute, USA) presented work tracking down the causes of estrogen receptor binding in two cancer cell lines, showing very different behavior between binding sites that are shared and those that are cell-line specific. Binding sites shared between cell lines tended to be a result of passive high-affinity binding to motif sequences, whereas cell-line-specific binding fell mostly in highly regulated open chromatin, and was positively regulated by transcription factor co-occurrence and negatively regulated by methylation.

Helena Kilpinen (University of Geneva, Switzerland) presented initial results from a pilot study that may turn out to be an indication of how future studies of epigenetic variation will look. Kilpinen analyzed next-generation sequencing assays of a dozen markers in two parent-child trios, including transcription, methylation, RNA polymerase binding and the binding of a number of transcription factors. By looking for allelic differences in epigenetic marks, and tracing these across generations and between unrelated individuals, it was possible to make inferences about which marks were sequence-context related, which were inherited across generations, and which were individual specific, as well as determining causal relationships between genetic and epigenetic effects. Once scaled up to dozens of trios this approach could potentially give a very complete image of normal epigenetic dynamics and variation in healthy humans.

Mapping complex traits in non-human species

The success of genome-wide association studies in mapping a large number of loci for complex traits in humans has been well reported in previous Biology of Genomes conferences. Studies of this type were also represented this year, with Mark McCarthy (University of Oxford, UK) reporting initial results of whole-genome and whole-exome sequencing in type 2 diabetes, and myself describing the discovery of over 70 new loci for inflammatory bowel disease. However, an interesting theme of the Biology of Genomes conference was the plethora of new successes in the mapping of complex traits in non-human species using genome-wide approaches.

Two talks at the conference described trait studies in domestic animals. Elaine Ostrander (National Human Genome Research Institute, USA) talked about a study of canine squamous cell carcinoma of the digit, a disease particularly prevalent in particular breeds, including Giant Schnauzers, Briards and black poodles. The breed structure of dogs comes in particularly helpful here, as the very long haplotypes within breeds allowed the detection of the oncogene in a poodle-only dataset using

only a 50,000 variant map, and the association could then be fine-mapped to a few tens of kilobases by looking at haplotypes across different susceptible breeds. Michael Shapiro (University of Utah, USA) described work on mapping traits in domestic pigeons. His group has collected samples from 2,000 birds, and performed next-generation sequencing in 41 samples from a diverse set of breeds. One success story of this project was the mapping of a single coding mutation underlying the crest formation, a trait that has been studied for over a hundred years in breeding experiments. By looking at the haplogroup structure around the mutation across different breeds, they could show that the mutation arose only once, and was incorporated into other breeds by hybridization and introgression.

In both these cases, the researchers benefited from the phenotypic diversity and well-described breed structure of their species, as well as the previous non-molecular characterizations of their traits of interest. They also both made good use of the networks of breeders and owners to put together impressive sample sets. Other talks about complex traits in non-human species leveraged different advantages of working with their specific species, in most cases allowing experiments that likely would not have been possible otherwise.

In some cases these studies utilized existing resources to allow otherwise expensive studies to become feasible. For instance, Amelia Baud (University of Oxford, UK) described a sequencing and phenotyping study using a rat breeding resource called the NIH heterogeneous stock. This resource consists of the descendants of 8 founder strains, crossed together for 50 generations. By sequencing the 8 founders, Baud was able to impute entire genomes into 1,407 phenotyped descendants using genotyping chips, allowing a large sequence-level association study on a range of phenotypes for relatively little cost.

Other studies used properties of natural species variation. Magnus Nordborg (Gregor Mendel Institute, Austria) utilized the natural tendency of *Arabidopsis thaliana* to form inbred lines to study the relationship between environment, genotype and gene expression. By growing clones of the same plant in different temperatures, and using genome, methylation and RNA sequencing, Nordborg demonstrated that only a small proportion of variance in gene expression is driven purely by environment, though much is driven by the interaction between environment and genotype or epigenomic marks. Similarly, Felicity Jones (Stanford University, USA) used the well-documented property of sticklebacks to evolve repeatedly from saltwater to freshwater forms to map the genetics of freshwater tolerance. The study looked for regions of the genome that tended to divide between freshwater and saltwater species, rather than along

geographic lines, thereby identifying over 80 regions that seemed to be involved in the freshwater transition. Jones was then able to look into what functional variants may underlie these changes, and showed that (consistent with human complex traits) only a minority of selected loci were driven by missense mutations in protein-coding genes.

Other studies were notable for extracting important results from quirks of experimental design. On the human/non-human boundary, Ran Blekhman (Cornell University, USA) presented a clever presentation on the interaction between human genetics and human microflora. Using the Human Microbiome Project sequencing data, Blekhman was able to extract data from human contamination in the microbial samples, and use these to call variants in the human hosts. Comparing these host variants to the microbial strains present identified a number of loci that correlated with the presence of certain bacterial species. Interestingly, these variants appear to be enriched for associations in inflammatory bowel disease, a disease in which the microbiome is believed to play an important role.

Conclusions

Meeting reports in *Genome Biology* have described the Biology of Genome conferences for over 8 years. Back in

2004, Mark Stapleton described how that year's conference represented a transition in genomics from a structural phase (genome sequencing) to a functional phase. In many ways, we are now deep into the functional phase, with the emergence of successful genome-wide association studies of human traits, and with extensive and detailed catalogs of coding and non-coding functional elements. However, this year's conference demonstrated that turning these results into a complete understanding of biological function, both in humans and non-humans, requires detailed and thoughtful approaches, as well as a continued escalation of data production. We are only just starting to put the biology back into the genome.

Abbreviation

ENCODE, Encyclopedia of DNA Elements.

Competing interests

The author declares that they have no competing interests.

Published: 27 June 2012

doi:10.1186/gb-2012-13-6-316

Cite this article as: Jostins L: Dispatches from the functional phase of genome biology. *Genome Biology* 2012, 13:316.