Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Systems biology and the analysis of genetic variation

Editorial overview

Shamil R Sunyaev and Frederick P Roth

Current Opinion in Genetics & Development 2013, **23**:599–601

This review comes from a themed issue on **Genetics of system biology**

For a complete overview see the Issue

0959-437X/$ – see front matter, © 2013 Published by Elsevier Ltd.

http://dx.doi.org/10.1016/j.gde.2013.11.010

Shamil R Sunyaev
Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA
e-mail: ssunyaev@rics.bwh.harvard.edu

Dr. Shamil Sunyaev received his PhD from Moscow Insitute of Physics and Technology (MIPT) and completed postdoctoral training at European Molecular Biology Laboratory (EMBL). He is now an Associate Professor at Harvard Medical School and Research Geneticist at Brigham & Women's Hospital. He is also an Associate Member at Broad Institute of MIT and Harvard. He is interested in the computational analysis of genetic variation, including origin of spontaneous mutations, population genetics, prediction of the functional effect of allelic variants and development of statistical methods for genomics of Mendelian and complex traits.

Frederick P Roth[1,2,3]
[1] Donnelly Centre and Departments of Molecular Genetics and Computer Science, University of Toronto, Toronto, Ontario M5S 3E1, Canada
[2] Lunenfeld-Tanenbaum Research Institute, Mt. Sinai Hospital, Toronto, Ontario M5G 1X5, Canada
[3] Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA
e-mail: fritz.roth@utoronto.ca

Dr. Frederick (Fritz) Roth trained in physics and biology at UC Berkeley, in biophysics at Harvard Univ., and in industry at Millennium Pharmaceuticals. His research team–ten years at Harvard Medical School and three years at the Univ. of Toronto's Donnelly Centre and Mt. Sinai Hospital's Lunenfeld-Tanenbaum Research Institute–develops technology to map functional relationships amongst disease-relevant genes.

This issue of *Current Opinion in Genetics and Development* is devoted to Systems Biology approaches in the analysis of genetic variation. Studies of DNA variation within human and other populations are propelled by the strong interest in genetics of complex and Mendelian traits, population genetics and microevolution. Dramatic improvements in sequencing technology have led to accelerated accumulation of information about genetic variation. Interpretation of sequence variation has now become a bottleneck in our progress towards mapping and understanding of complex genetic disease. Systems biology approaches offer great promise in interpreting genetic variation from the functional standpoint. Furthermore, analysis of sequence variation holds a potential to inform systems biology by highlighting gene sets and pathways underlying organismal and molecular traits, and by revealing interactions and mechanics of evolution underlying functional modules. This issue covers both systems biology approaches to the analysis of genetic variation and new genetic analyses informative about biological systems.

In genetics of complex phenotypes, especially in genetic association studies of human common diseases, associating a disease with a locus is often only the beginning. Genome-wide association (GWA) studies may point to a statistically associated SNP, but this SNP may be linked to dozens of genes. It is not clear which SNP within this 'haplotype block' is causal, nor even which gene harbors the causal SNP(s). Genes may also be associated with disease through observation of a high-burden of mutations in cases relative to controls. It can be challenging both for common-variant and rare-variant disease associations to separate true- from false-positives. Three different reviews — Leiserson *et al.*, Carter *et al.*, and Atias *et al.* — describe network approaches to the analysis of human genetic disease. Collectively, they discuss methods by which network analysis can point us to the causal genes within disease-associated loci, and to identify causal paths from allele, to intermediate molecular phenotype, to disease. These approaches are relevant not only to GWA studies, but also to analysis of somatic mutations identified in tumour genome sequencing or rare variants found via exome or genome sequencing.

It is becoming a common theme that most of the variants discovered by genome-wide association studies are non-coding and of weak effect. Many of these variants likely affect transcriptional regulation. Stranger and Raj review the genetics of human variation underlying differences in gene expression between individuals (expression quantitative trait loci, or 'eQTLs'). For example, a sequence variant can alter the expression of the gene that harbors it (leading to observation of a 'cis-eQTL'). Among the

highlighted challenges are the need to carry out eQTL studies for diverse tissues and environmental perturbations, the need for more attention to 'trans-eQTLs', sequence variants potentially altering the expression of unlinked genes, to protein levels as genetic traits, and sequence variation that affects splicing.

The article by Trynka and Raychaudhuri reports on new approaches to interpret allelic variation involved in human common diseases through the impact on transcriptional regulation. Intersecting results of genome-wide association studies with results of genome-wide chromatin assays suggests that many of the variants underlying complex traits have regulatory roles in relevant cell types. This work highlights the importance of cell-type specific regulatory context and underscores the value of epigenomics.

It is not always the case that both parental alleles are expressed equally. This can arise from cis-eQTLs, for example, variant alleles leading to changes in promoter strength, coding changes leading to nonsense-mediated decay. It can also arise due to chromatin effects that are programmed, for example, X-chromosome inactivation. Savova *et al.* review the phenomenon of autosomal monoallelic expression (MAE), which causes some (but not all) loci to experience inactivation of one or the other parental allele (causing an alternating mosaic of expression of the two alleles across cells within a tissue). We are only beginning to understand the mechanisms and selective effects of this intriguing phenomenon.

There is also progress in understanding the functional impact of coding variation. Sahni *et al.* discuss the impact of amino acid allelic variants on protein–protein, protein–DNA and protein–RNA interactions. They suggest that the new field of 'edgetics' (the genetics of biological links or 'edges' between nodes in a graph) should go beyond exploring the protein-centric impact of alleles, for example, on protein folding and stability, to understand the impact of allelic changes on specific interactions. Variants impacting specific interactions can provide unique clues towards understanding of the molecular basis of phenotypic variation and highlight pathways involved in Mendelian and complex phenotypes.

In spite of successes of computational approaches and *in vitro* experiments, studies in model organisms commonly provide the most convincing proof of functional significance of sequence variants *in vivo*.Fowler and Dunham discuss the value and limitations of model organisms to functionally characterize sequence variation, focusing on *Saccharomyces cerevisiae*. Examples include the prediction from yeast studies of the role of mismatch repair genes in human colon cancer. Yeast can also be used to study human variation, for example by reconstituting human variation within corresponding positions of orthologous yeast proteins. More directly, where exogenous expression of a wild-type human gene complements a mutation within a yeast gene, this phenomenon can exploited to assess the function of candidate disease-causing variation if the variant human gene fails to complement.

Several articles in this issue focus on principles of allelic architecture of complex traits, such as the distribution of allelic effect sizes, number of loci involved in complex traits and interactions between individual loci.

Falke *et al.* summarize recent findings in the effort to identify the molecular basis of complex traits in plants. Their review points out that major-effect alleles in the same 'candidate genes' frequently appear in different plant gene-mapping studies. However, they describe methods that analyze mutations of all effect sizes. Recent experiments show that, in spite of the existence of large effect alleles, the multitude of smaller effect alleles are important for complex trait variation and must be investigated to fully explain the molecular basis of phenotypic variation.

Genetic mapping in a simpler yeast system is informative about the distribution of allelic effect sizes. Fay discusses recent progress in the search for molecular mechanisms underlying quantitative trait variation in yeast. He notes that current studies primarily identify variants in protein coding regions, although a substantial fraction of DNA variants are not SNPs. Interestingly, multiple alleles involved in the same quantitative trait may be observed in a locus and these alleles may be linked. Analysis of gene expression suggests that most of variants involved in quantitative traits induce pleiotropic changes on expression of genes unrelated to the trait.

The review by Fu *et al.* further discusses sources of genetic complexities in both model organisms and humans. He covers the results of recent population sequencing studies in our own species. These studies are informative about the potential models of allelic architecture of complex traits, although many questions remain unanswered. This review argues that new technologies pave the way towards more mechanistic understanding of complex trait variation.

Nourmohammad *et al.* focus on molecular traits such as gene expression, and also on other molecular phenotypes such as binding affinity. They describe recent theoretical developments which highlight universal principles underlying evolution and maintenance of these traits and which determine their allelic architecture. This review discusses properties of population variation of molecular traits that are independent of specific functional details.

How many loci are involved in complex phenotypes? New technology revitalizes an old idea to find genes that respond to artificial selection using hitchhiking (HH)

mapping. Model systems, such as Drosophila can be subjected to selection and the response to selection can be detected by DNA sequencing as a region of reduced diversity due to hitchhiking with a selected allele. Nuzhdin and Turner discuss recent literature on HH-mapping. They suggest that the initial conclusions, that thousands of loci are involved in complex traits responsible for the response to selection, are questionable. They find the results to be consistent with dozens rather than thousands of loci under selection, and suggest new ways to fully realize the power of HH-mapping.

The analysis of allelic variation that is arguably the most relevant to Systems Biology is the study of genetic interactions (epistasis, in Fisher's sense of the word). Intramolecular interactions are one important class of epistatic interactions. A fundamental question is whether the interactions are primarily pair-wise, or more complex higher order interactions play a substantial role. Weinreich *et al.*, analyze current examples of exhaustive testing of finite allelic combinations. They suggest a new method to evaluate the effects of higher order interactions and apply it to the data available from literature.

We hope that together, the articles in this issue serve to paint a current picture of the concepts, accomplishments and promises of informing and applying systems biology in genetic studies.

Funding

SRS was supported by National Institutes of Health (R01 GM078598; R01 DK095721; R01 MH101244; U54 CA143874.; U01 HG006500). FPR was supported by the National Institutes of Health (HG004233, HG001715 and HL107440); an Ontario Research Fund — Research Excellence Award; a Canadian Institute for Advanced Research Fellowship; the Canada Excellence Research Chairs Program; and by the Avon Foundation.