Genome Biology

Meeting report

Phenotypic variation meets systems biology

Teresa M Przytycka

Address: National Institutes of Health, National Library of Medicine, National Center of Biotechnology Information, Bethesda, MD 20894, USA.
Email: przytyck@ncbi.nlm.nih.gov

Abstract
A report of the 18th Annual Growth Factor and Signal Transduction Symposium, Ames, USA, 11-14 June 2009.

This year's annual growth factor and signal transduction symposium held at Iowa State University focused on systems-biology approaches to the study and modeling of complex biological processes. The topics discussed covered a wide spectrum of recent advances, including systems-level approaches to understanding transcriptional and posttranscriptional regulation and systems-level analysis and modeling of the responses of biological systems to perturbations. A few of the highlights of the meeting in these fields are reported here.

Systems approaches to transcriptional regulation

Transcriptional regulation can be broadly defined as the process in which transcription factors interact with each other, with DNA, and with other biomolecules to regulate gene expression. This complex regulatory process is the subject of intensive study utilizing a variety of experimental and computational techniques. Opening the meeting, David Hume (Roslin Institute, University of Edinburgh, UK) described his recent results on gene regulation in macrophages related to the RIKEN genome Network Project and FNTOM 4 (Functional Annotation of Mammalian Genome). Hume and his collaborators analyzed transcriptional control of the human monocytic cell line THP-1 throughout a time course of growth, arrest and differentiation. Using the deepCAGE technique, they measured the dynamics of genome-wide usage of transcription start sites. This analysis was followed with comparative genomics approaches to predict active regulatory motifs throughout time and to predict the key transcription factors driving differentiation and uncover their time-dependent activation. Some of the predicted factors were subsequently confirmed by knockdowns using small interfering RNAs. Challenging the concept of 'master regulators', Hume argued that cellular states are constrained by complex networks that involve a substantial number of both positive and negative regulations.

Transcriptional regulation was also the focus of Timothy Ravasi (University of California, San Diego, USA) who discussed systematic screens for protein-protein interactions among 1,988 human and 1,727 mouse transcription factors, along with quantitative assays of their expression across human and mouse tissues. He and his collaborators performed cross-species comparisons that revealed dozens of complexes with conserved patterns of interactions and tissue specificity. Ravasi argued that transcription factor 'hubs' tend to be expressed ubiquitously in many tissues, which is suggestive of a 'facilitator' role for these molecules.

Small non-coding RNAs impose an additional level of complexity on the regulation of gene expression. Rodrigo Gutierrez (Pontificia Universidad Catolica de Chile, Santiago, Chile) has used 454 DNA-sequencing technology supported by bioinformatics analysis to identify and analyze sRNAs that are regulated by nitrate treatments in *Arabidopsis*. This analysis revealed a novel nitrate-responsive microRNA-target regulatory module involved in a coordinated regulatory feedback loop induced by nitrate and repressed by nitrogen compounds produced by nitrate reduction.

Graziano Pesole (University of Bari, Italy) stressed that studies of gene expression should take into account factors related to gene structure and, in particular, to alternative splicing. He surveyed the full set of available transcript and expressed sequence tag (EST) data to identify alternative isoforms. The alternative splicing patterns are made accessible at the Alternative Splicing Prediction Database [http://t.caspur.it/ASPicDB].

Taking a different angle, a metabolic-centered study of gene regulation was described by Melissa Kemp (Georgia Tech and Emory University, Atlanta, USA), who focused on the control of intracellular protein thiols. She and her collaborators have implemented a computational model of the intracellular redox network to investigate hydrogen peroxide buffering within mammalian cells. She argued that the level of protein thiol modification is regulated by a highly connected redox enzyme network.

Genetic variation as a way of 'perturbing' biological systems

Another broad theme represented at the symposium was the study of genetic variations and their impact on gene

expression and other phenotypes. Several speakers described work that utilizes the natural genetic variation within and between species as genome-wide 'perturbations'. Using this approach, correlation of phenotypic and genotypic variation can provide cues about possible causal relationships between genetic mutations and diseases. Gene-expression profiling and the analysis of expression quantitative trait loci (eQTL) are being increasingly used to uncover disease-associated networks and basic information about gene regulation. Barbara Stranger (Harvard Medical School, Boston, USA) presented her pioneering work on the population genetics and genomics of human gene expression. She and her collaborators collected whole-genome expression data in lymphoblastoid cell lines of individuals represented in the HapMap consortium and performed association analysis of these expression data with genotypic variations for the human populations sampled in HapMap. This study provides a global view of prevailing patterns of human eQTLs. For example, Stranger reported that about half of the expressed genes are associated with genetic variations *in cis*, and over one-third of such associations are observed in two or more populations. Importantly, she observed statistically significant interactions among a large fraction of *cis*-eQTLs. Stranger also discussed her work on tissue specificity and conservation of eQTL relations between human cell types, including the detection of extensive cell-specific genetic effects.

eQTL analysis was also the central point of a study of genome diversity and gene expression in maize reported by Antoni Rafalski (Pioneer Hi-Bred International, Johnston, USA). This work surveyed, using the technique of array-comparative genome hybridization (aCGH), the presence of structural variation, including copy number variation (CNV) in maize genomes. Rafalski found that structural variations such as CNVs are common features in the maize genome and that some of these variations are quite large and comprise several genes. He also described a complementary eQTL study in inbred populations derived from a cross between B73 and Mol17.

Genetic variation can lead not only to variation in transcript levels but also to other phenotypic variation, such as variation in metabolite and protein levels or other quantitative traits. Lauren McIntyre (University of Florida, Gainesville, USA) discussed computational methods for analyzing transcript level and linking these transcript-level analyses to phenotypic variation in *Drosophila melanogaster*. In contrast, Dan Kliebenstein (University of California, Davis, USA) described research aiming to connect natural variations in gene expression in *Arabidopsis* to their phenotypic consequences, focusing on the plant metabolome and, more specifically, on aliphatic and indolic glucosinolate pathways. He and his collaborators discovered that major regulators of gene expression variation in these pathways are two biosynthetic enzymes rather than transcription factors. Kliebenstein also made the interesting observation that, in his study, the majority of false-positive associations are not due to general population structure, but appear to be caused by selective sweeps.

Computing intracellular interactions

There were also several interesting computationally oriented presentations. Mona Singh (Princeton University, Princeton, USA) proposed the language of network schemas for describing recurring patterns of specific types of proteins (a protein type can be, for example, 'kinase') and their interactions. Such network schemas describe proteins and specify the topology of interactions among them and can be seen as an extension of the familiar network motifs. Singh outlined the computational approach for identifying network schemas that are recurrent and over-represented, and showed the results of an application of this methodology to the physical interaction network in *Saccharomyces cerevisiae*. The program NetGrep [http://genomics.princeton.edu/singhlab/netgrep] developed by Singh's group to perform this type of analysis is available online.

Whereas Singh explores the properties of protein-interaction networks, Maricel Kann (University of Maryland, Baltimore, USA) focuses on the prediction of protein interactions. Specifically, she discussed new insights into methods of uncovering protein-protein interaction using the coevolution principle. While coevolution of interacting proteins has been confirmed by several computational approaches, it is not clear which biological dependency contributes most significantly to uncovered coevolution signal. Her results indicated that both binding and non-binding positions contribute to the overall coevolution signal. However, she showed that, when controlling for the number of residues, binding sites provide a stronger signal than a randomly selected set of non-binding positions.

Gene knockouts perturb the system in a more drastic way than genetic polymorphism. Such knockouts allow identification of essential genes - that is, genes indispensable for cell survival in optimal conditions. I (TMP) presented my group's recent results related to uncovering the relation between gene essentiality and network topology. We showed that the over-representation of essential proteins among high-degree nodes should be attributed to their involvement in 'essential complex biological modules', a group of densely connected proteins with shared biological function that are enriched in essential proteins.

Hamid Bolouri (California Institute of Technology, Pasadena, USA) presented a reverse-engineering method developed by his group. Reverse engineering is the process of uncovering the structure of a system by reasoning from observations of its behavior, such as gene expression level.

Bolouri presented the results of ongoing studies in collaboration with Ellen Rothenberg (California Institute of Technology) to apply reverse engineering to the regulatory network that underlines T-cell development in mice. His interactive tool, BioTapestry [http://www.biotapestry.org], for building, visualizing and simulating genetic regulatory networks is available online.

Another computational topic represented at the symposium was a simulation of dynamic activity of small biological networks. Chao Tang (University of California, San Francisco, USA) uses this approach to study cell responses to stimuli. He has simulated all possible three-node signaling network topologies and found that only two major core topologies are capable of achieving robust adaptation. Stefan Hoops (Virginia Bioinformatics Institute, USA) discussed recent tools for simulation of a biological network activity by a hybrid ODE/stochastic approach implemented within their network simulation and analysis software COPASI, while Yiannis Kaznessis (Biotechnology Institute, University of Minnesota, St. Paul, USA) presented the synthetic biology modeling software suite SybBioSS [http://synbioss.sourceforge.net], available online.

In summary, while the symposium covered a broad range of topics, several common themes echoed in many presentations. These recurring views indicated the shift of the systems biology field from a static description of large biological systems towards more dynamics and context-dependent analysis.

Published: 3 August 2009
doi:10.1186/gb-2009-10-8-313