

Meeting report

Developing a systems-level understanding of gene expression

Olivier Elemento

Address: Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA. Email: elemento@molbio.princeton.edu

Published: 30 April 2007

Genome Biology 2007, 8:304 (doi:10.1186/gb-2007-8-4-304)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2007/8/4/304>

© 2007 BioMed Central Ltd

A report on the meeting 'Systems Biology: Global Regulation of Gene Expression' at the Cold Spring Harbor Laboratory, New York, USA, 28 March-1 April 2007.

This year's annual systems biology meeting at Cold Spring Harbor showcased a wide range of experimental, computational and theoretical approaches to studying the multiple facets of gene expression. The meeting also featured new technologies, several large-scale studies and studies on a diversity of model organisms - from *Escherichia coli* through *Arabidopsis* to humans.

Protein-DNA interactions

In his keynote speech, Uri Alon (Weizmann Institute, Rehovot, Israel) described how complex regulatory networks can be decomposed into simple and recurrent patterns, which he terms network motifs. Mathematical analysis predicts the dynamic functions of these motifs, and Alon's group has verified several of these predictions using highly accurate measurements of promoter activity *in vivo*. For example, he showed that the *Escherichia coli* L-arabinose-utilization system uses a motif called a feedforward loop for a delayed response to cyclic AMP stimulation and a rapid response to cAMP depletion.

Novel high-throughput techniques are being used to identify and characterize transcription factor-DNA interactions. Marian Walhout (University of Massachusetts, Amherst, USA) uses yeast-one hybrid technology to map the protein-DNA interactions in *Caenorhabditis elegans* neurons. She identified 94 transcription factors, which bind the promoters of around 40 neuronal genes. Together with previously published experiments, the overall protein-DNA interaction network so far uncovered by Walhout covers 20% of all *C. elegans* transcription factors across 250 promoters. Martha Bulyk (Harvard University, Cambridge, USA) is currently using her protein-binding microarray technology to systematically determine the DNA-binding

specificities of mouse transcription factors. To date, Bulyk and collaborators have determined the specificity of around 400 of the more than 500 mouse factors purified so far.

Identifying genomic regulatory elements

A major challenge of the post-genomic era is the identification of genomic regulatory elements, particularly those that act at a distance from their cognate genes. As many of these elements are under negative selection, conservation across multiple genomes provides a powerful way to detect them. Alex Visel (Lawrence Berkeley National Laboratory, San Francisco, USA) reported the identification of highly conserved regions in the human genome, and the testing of a large number of them for enhancer activity using a transgenic mouse assay. He and his group have tested more than 500 regions, 230 of which appear to be tissue-specific enhancers. They also created synthetic constructs in which they fused enhancers that drive expression in distinct tissues. Most surprisingly, the resulting expression patterns were always additive, and ectopic expression was never observed, suggesting that no interactions occurred between enhancers.

Because not all regulatory elements will be conserved, there is an obvious need for unbiased experimental approaches for their large-scale discovery. David Hawkins (University of California, San Diego, USA) described how histone-modification patterns (acetylation at lysine (K)18 or 27 on histone H3, and methylation at K4 on H3) reliably identify many of these distal enhancers, and described the use of this approach to map out several thousand enhancers in *Drosophila* S2 and wing imaginal disc cells.

Post-transcriptional regulation of gene expression

It is thought that at least 50% of human genes undergo alternative splicing. However, where and when alternative splicing occurs, its functional role, and the regulatory code that mediates splicing events are largely unknown. Using a microarray capable of monitoring approximately 7,000 alternative splicing events, Benjamin Blencowe (University

of Toronto, Canada) identified 150 exons that are preferentially skipped or included when the genes are expressed in the central nervous system. He also discovered several C/U-rich motifs in flanking introns and neighboring constitutive exons that may be involved in exon inclusion in the central nervous system.

MicroRNAs (miRNAs) are small RNAs that regulate gene expression. David Bartel (Massachusetts Institute of Technology, Cambridge, USA) presented an improved approach for predicting the target genes of miRNAs that does not rely on sequence conservation. Bartel's approach uses several newly discovered features of miRNAs. For example, he has observed that miRNA target sequences tend to be out of the path of the ribosome - that is, not in coding sequences and more than 15 nucleotides after the stop codon - and are more often found towards the beginning or the end of 3' untranslated regions. There is also a strong preference for high AU-content immediately around regions pairing with the 'seed' sequence (positions 1-8 of the miRNA). Pairing at the miRNA 3' end also appears to follow particular rules: Bartel showed that requiring strong pairing immediately 3' of the seed decreases prediction accuracy. Indeed, strong pairing often involves G-C bonds, which contradicts the preference for high AU-content around the seed.

Dan Hogan (Stanford University, Stanford, USA) has used chromatin-immunoprecipitation (ChIP) to pull down RNA targets for 40 RNA-binding proteins in yeast. The number of targets per protein is quite diverse, ranging from two (Nop13) to several thousands (Pab1). He identified the sequence motif for a dozen proteins, and hypothesized that the remainder bind secondary (or tertiary) structures, for example, short hairpins. He also raised the somewhat provocative hypothesis that many, if not most, mRNAs may be shuttled from the nucleus to subcellular foci by these RNA-binding proteins.

Biophysical approaches for studying gene expression

It has been shown that the LacI repressor finds its operator 100 times faster than expected from simple three-dimensional diffusion models. According to the facilitated diffusion model ('1D+3D' model), transcription factors alternate between diffusing along the DNA (one-dimensional) and jumping from one site to another (three-dimensional), until they find their target. Leonid Mirny (Massachusetts Institute of Technology, Cambridge, USA) reported that the experimentally measured affinity of transcription factors for nonspecific DNA is too high for the 1D+3D model to work. However, if the model is extended by requiring genes for transcription factors to be in spatial proximity to their targets in DNA, it yields estimates of search time that are compatible with measurement. Mirny then showed that in bacterial genomes, transcription factors are often located

close to their targets (LacI is located right next to its operator in *E. coli*), and hypothesized that the fast search times may be an important factor in shaping these genomes.

Nir Friedman from Sunny Xie's group at Harvard University (Cambridge, USA) described a system for measuring levels of the protein β -galactosidase within a single cell at single-molecule resolution. Single cells trapped in microfluidic chambers are treated with a fluorogenic substrate for the enzyme, and each expressed copy of the enzyme creates a large number of fluorescent molecules as the readout. Friedman showed that proteins are produced in random bursts, with an exponentially distributed number of molecules per burst. He also described an analytical model for reconciling real-time measurements of protein levels in single cells with population-wide distributions of protein levels.

A thermodynamics model for predicting gene-expression patterns from sequence, taking into account the concentrations of transcription factors and their known sequence affinities was presented by Eran Segal (Weizmann Institute, Rehovot, Israel). His approach also explicitly takes into account competition between factors for the same DNA sequences and includes contributions from weak binding sites. When applied to the segmentation gene network in *Drosophila*, his approach recovered the correct expression patterns for 80% of *cis*-regulatory modules.

Spatio-temporal patterns of gene expression in multi-cellular organisms

In multicellular organisms, spatial aspects of gene expression are often studied by expressing green fluorescent protein (GFP) under the control of endogenous promoters. Uwe Ohler (Duke University, Durham, USA) described a computational approach for extracting gene-expression information from confocal images of such experiments, with emphasis on *Arabidopsis* roots. His approach involves mapping root images onto prototypical root templates using image-distortion algorithms followed by the measurement of organ-specific GFP intensities. Work in progress includes scaling his approach to a 'root array' in development, where up to 5,200 roots with distinct promoter-GFP fusions can be studied in parallel.

Denis Dupuy (Harvard University, Cambridge, USA) is systematically characterizing spatio-temporal gene-expression patterns in *C. elegans* (an effort he calls the 'Localizome'). He has generated about 2,000 *C. elegans* strains, each expressing GFP under the control of an endogenous promoter. Worm cultures from each strain are analyzed using a novel type of flow cytometer capable of measuring worm sizes (different worm sizes correspond to different development stages) and generating profiles of fluorescence intensity along the worm body axis. This high-throughput analysis generates spatio-temporal profiles of gene expression. In a preliminary analysis,

Dupuy showed that genes with similar profiles tend to be functionally related.

Epigenetic modifications and gene regulation

Gordon Robertson (University of British Columbia, Vancouver, Canada) described how his group used ChIP and Solexa DNA sequencing to map chromatin modifications (lysine trimethylation of H3 at different positions) in human leukemia cells. Solexa machines can currently sequence 4-9 million 27-bp-long fragments per lane (Robertson's machine has eight independent lanes), with around 60% of the reads mapping to unique places in the human genome. His results confirm that trimethylation on H3 K4 correlates with transcription initiation, whereas trimethylation on H3 K27 correlates with transcript elongation. He also identified multiple large domains of H3 K9 trimethylation on chromosome 19q, one of which covers a dense cluster of 32 genes for KRAB-ZNF transcriptional repressors.

Using a high-resolution tiling array covering the four human Hox gene complexes, Howard Chang (Stanford University, Stanford, USA) discovered more than 200 noncoding RNAs expressed in diverse human tissues. He presented strong evidence that HOTAIR, a noncoding RNA encoded in the *HOXC* locus, acts as a *trans*-repressor of the *HOXD* locus by establishing a silent chromatin domain.

Using yeast tiling arrays, Oliver Rando (University of Massachusetts, Amherst, USA) measured the turnover of histone H3 at a single nucleosome resolution, in G1-arrested cells (to avoid DNA duplication interfering with chromatin states). He found that nucleosomes located at transcription start sites exhibit higher histone turnover rates than nucleosomes at coding sequences. This is surprising, as it was believed that most nucleosome disruption was caused by the passage of RNA polymerase over coding regions. Rando has also found that high histone turnover occurs at the boundaries of chromatin domains, possibly acting to prevent their spread.

This meeting made it clear that new and improved technologies (such as sequencing, microfluidics, high-density tiling arrays and microscopy) are fueling the rapid expansion of a systems-level understanding of gene expression. These technologies are revealing the importance of noncoding RNAs and their role in regulating gene expression, as well as the extent of post-transcriptional regulation. Epigenetic modifications, the dynamic nature of chromatin and its role in regulating gene expression are also becoming better understood. Scientists are now applying experimental and computational techniques originally developed for the genomes of unicellular model organisms to complex multicellular ones, including humans. With speakers drawn from the most innovative groups in the field, the Cold Spring Harbor meeting continues to be one of the major annual scientific rendezvous for systems biologists.

Acknowledgements

I thank Chang Chan, Alison Hottes, Manuel Llinás, Tiffany Vora and Saeed Tavazoie for insightful comments and suggestions.