Genome **Biology**

## MEETING REPORT

# Integrating systems biology data to yield functional genomics insights

Polly Fordyce[1]* and Nicholas Ingolia[2]

### Abstract

A report of the recent EMBO Conference 'From Functional Genomics to Systems Biology' held at the EMBL Advanced Training Centre, Heidelberg, Germany, 13-16 November 2010.

The emerging challenge in systems biology is the integration of large genomics and proteomics datasets to provide new biological insights. Powered by advances in DNA sequencing, high-resolution maps of genome-wide transcription factor and chromatin occupancy data have begun to shed light on the basic mechanisms that regulate gene expression and generate tissue-specific expression patterns. These datasets, combined with other large-scale physical and genetic interaction networks, are providing a better functional understanding of biological systems.

### Three-dimensional chromosome structure and interactions with regulatory proteins

Studies of three-dimensional chromosomal conformations revealed a complex architecture, with widespread interactions both within and between chromosomes and between chromosomes and other cellular structures. Job Dekker (University of Massachusetts Medical School, Worcester, USA) has expanded chromosome conformation capture techniques to allow both genome-wide identification of interactions and more detailed investigation of specific contacts between individual promoters and distant regulatory elements. These maps revealed distinct interaction-rich domains, representing active chromatin, and domains with few interactions, representing inactive chromatin, with many weak looping interactions between promoters and distant regulatory elements (over 2 Mb away). Rick Young (Massachusetts Institute of Technology, Cambridge, USA) proposed a

physical mechanism for establishing and maintaining these interactions, discovered from genome-wide co-localization data for various general transcription factors. In this model, interactions between transcription factors and mediator establish chromosome loops, and mediator then recruits cohesin to tether these loops in place.

To understand the function of chromosome architecture in *Caenorhabditis elegans*, Jason Lieb (University of North Carolina, Chapel Hill, USA) surveyed contacts between chromosome arms and the nuclear lamina. Large lamina-associated regions contained 'looped-out' segments with high concentrations of transcription machinery, suggesting that these regions could concentrate transcription factors and drive higher levels of expression. In another approach, Guillaume Filion (The Netherlands Cancer Institute, Amsterdam, The Netherlands) used principal component analysis to probe genome-wide localization maps for 53 chromatin proteins in *Drosophila* Kc cells. The data could be assigned to five major classes that seemed to illustrate functional chromatin types, including HP1- and polycomb-bound silenced chromatin, inactive chromatin lacking histone marks that comprised the majority of the genome, and two types of actively transcribed chromatin enriched for either housekeeping genes or tissue-specific genes.

### Role of specific *cis*-regulatory elements and their influence on regulating gene expression

Although high-throughput localization-based techniques are identifying distal regulatory elements at an unprecedented pace, it has been harder to link these *cis*-regulatory elements to their targets. To address this problem, Bing Ren (Ludwig Institute of Cancer Research and University of California, San Diego, USA) used chromatin immunoprecipitation sequencing in transgenic mice to create high-resolution maps of promoters, enhancers, and looping interactions, resulting in annotation of nearly half of the previously identified conserved non-coding regions. Although promoters were often active in multiple tissues, enhancer activity was tissue-specific, leading to a model in which multiple enhancers impinge on a single promoter, forming an enhancer-promoter unit that drives tissue-specific expression.

*Correspondence: polly@derisilab.ucsf.edu
[1]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94158, USA
Full list of author information is available at the end of the article

**BioMed** Central

Transcription factor binding is a dynamic process, and timing of binding events is critical for specifying appropriate developmental programs. Several speakers measured transcription factor occupancy profiles at multiple time-points during development to probe how changing occupancy leads to cellular differentiation. In studies of mesoderm specification in *Drosophila*, Eileen Furlong (European Molecular Biology Laboratory, Heidelberg, Germany) discovered that binding was significantly more dynamic than would be expected by either expression analysis or computational algorithms, revealing an additional level of complexity. Stuart Kim (Stanford University School of Medicine, Stanford, USA) and Susan Mango (Harvard University, Cambridge, USA) both harnessed the detailed knowledge of individual cell lineages in *C. elegans* to probe development. Kim created a 'digital gene expression atlas' with single-cell resolution, and used comparisons between these expression data and maps of transcription factor binding to identify the precise binding events that trigger differentiation. Mango focused specifically on pharyngeal development, finding that despite widespread expression, binding of forkhead box transcription factor PHA-4 was restricted to specific cells by association with the nuclear lamina. Binding occurred significantly before gene expression and was followed by extensive chromatin remodeling.

A presentation by Tali Raveh-Sadka (Weizmann Institute of Science, Rehovot, Israel) provided a reminder of the intrinsic role of underlying DNA sequence in determining expression levels. Raveh-Sadka and colleagues showed that poly(dA:dT) nucleotide tracts, which modulate nucleosome positioning, could change expression as much as altering transcription factor binding sites, establishing the importance of considering nucleosome positioning in gene expression.

Despite steps towards an increasingly high-resolution picture of the binding events that dictate patterns of gene expression, the symmetry-breaking process required to drive unidirectional transcription has remained mysterious. Jonathan Weissman (University of California, San Francisco, USA) described results obtained from deep sequencing the 3' end of nascent RNA transcripts to characterize the position of actively transcribing RNA polymerase molecules at single base-pair resolution. The data reveal that the histone deacetylation complex Rpd3s has a major role in biasing the direction of transcription. Lars Steinmetz (European Molecular Biology Laboratory, Heidelberg, Germany) presented an analysis of regulated non-coding transcripts in the budding yeast cell cycle, many of which result from bidirectional promoters. Many of these transcripts are antisense relative to a regulated protein-coding gene, and Steinmetz presented evidence that these overlapping transcripts can repress one another.

## Rewiring of transcriptional networks during evolution

Comparisons of transcription factor binding patterns between evolutionarily divergent species can reveal conserved regulatory mechanisms and shed light on how components are repurposed during evolution. Paul Flicek (European Bioinformatics Institute, Hinxton, UK) compared patterns of binding between multiple vertebrate species, whereas Alex Stark (Research Institute of Molecular Pathology, Vienna, Austria) focused on related *Drosophila* species. Surprisingly, binding patterns were found to be significantly more conserved between *Drosophila* species, even over comparable evolutionary distances; vertebrate binding patterns were largely species-specific, despite strong conservation of individual transcription factor sequence preferences. Flicek and colleagues found that transcription factor binding patterns from mice carrying a copy of human chromosome 21 recapitulated the human binding pattern, establishing that underlying sequence, and not just the cellular environment, has a crucial role in dictating binding.

Structural polymorphisms between individuals provide a particularly useful tool for identifying *cis*-regulatory elements because they can entirely eliminate a cluster of transcription factor binding sites. Bart Deplancke (École Polytechnique Fédérale de Lausanne, Switzerland) identified structural variants in *Drosophila* and connected them to expression quantitative trait loci, which are genetically defined polymorphisms in gene expression. This analysis effectively used natural variation to identify regulatory regions by examining the effects of their deletion.

## Post-transcriptional regulatory mechanisms

Experiments enabled by microarrays and deep sequencing are also providing global surveys and interesting mechanistic insights into post-transcriptional regulation. Patrick Cramer (University of Munich, Germany) presented an analysis of transcriptome dynamics in yeast that revealed coordinated changes in mRNA synthesis and decay following osmotic shock. Transcription alone could not account for changes in mRNA levels, emphasizing an important role for post-transcriptional control in stress responses, and it will be exciting to see how these changing mRNA levels are reflected in protein synthesis. Howard Chang (Stanford University School of Medicine, Stanford, USA) expanded on his genome-wide analysis of RNA secondary structure by showing how structures changed at different temperatures, and suggested that these changes in RNA folding could affect translation in order to mediate temperature-dependent gene expression.

## Reconstructing and analyzing biological regulatory networks

The biological function of a regulatory interaction depends not only on the precise interaction partners, but

also on its context in the regulatory network of the cell. Many networks include cross-regulation and feedback that can qualitatively change their overall behavior. Marian Walhout (University of Massachusetts Medical School) found that the transcription factors that induce microRNA expression were often themselves targeted for repression by the microRNAs they regulate, providing a pervasive form of negative feedback. In addition to microRNAs, gene regulatory networks involve diverse post-transcriptional and post-translational control mechanisms that affect the abundance and activity of transcription factors. Other presentations described the mapping of these mechanisms by identifying physical interactions using high-throughput co-purification and mass spectrometry. Mike Tyers (Wellcome Trust Center for Cell Biology, Edinburgh, UK) used this approach to find kinase and phosphatase targets, despite the technical difficulties in detecting these weak and transient associations. This study revealed new components of the well-studied target of rapamycin (TOR) kinase pathway linking nutrient sensing to protein synthesis and cellular growth. Mike Snyder (Stanford University School of Medicine) showed that mass spectrometry could be adapted to identify small molecule interactors as well, revealing unexpected sterol binding by several yeast kinases, two of which were regulated by sterol levels.

Genetic approaches have the potential to reveal functionally important interactions regardless of the molecular details. Aviv Regev (Broad Institute, Cambridge, USA) presented a systematic analysis of the immune response to Toll-like receptor stimulation by measuring how RNA interference (RNAi) perturbations affected ligand-induced gene expression profiles. This study revealed network-level features, such as feedback and cross-inhibition, as well as identifying new signaling proteins. Frank Holstege (University Medical Center, Utrecht, Netherlands) analyzed the steady-state expression profiles of yeast mutants and double mutants, similarly revealing complicated cross-talk between signaling proteins and transcription factors acting in the same pathway.

Unexpected double mutant phenotypes, known classically as epistasis, are a powerful indicator of functional interactions between genes. Comprehensive epistasis maps produced from systematic double mutant construction have identified complexes, pathways, and networks in budding yeast. Thomas Sandmann (German Cancer Research Center, Heidelberg, Germany) performed genetic interaction mapping in animal cells by using RNAi in place of mutation and using atypical effects of knocking down two genes in combination to define functional links between them. Trey Ideker (University of California, San Diego, USA) described a differential epistasis mapping in which changes in gene-gene interactions following genotoxic stress, relative to their untreated epistasis, revealed the central pathways involved in the DNA damage response.

One major goal of functional genomics is a complete, quantitative inventory of the contents of a living cell. Luis Serrano (Center for Genomic Regulation, Barcelona, Spain) presented a progress report on the comprehensive catalog of all small molecules, RNAs, and proteins in *Mycoplasma*, as well as quantitative characterization of the correspondence between mRNA and protein levels and the stoichiometry of proteins in complexes. This quantitative analysis of *Mycoplasma* illustrated that a single population of genetically identical cells was nonetheless heterogeneous, an idea that has attracted great interest in recent years. Lucas Pelkmans (ETH, Zürich, Switzerland) showed that it was possible to exploit this cell-to-cell variability in an RNAi screen to produce more nuanced phenotypic profiles. In their study of viral infection, Pelkmans and colleagues found an effect of cell density on susceptibility and used it to distinguish whether genes acted directly in infection or affected cell growth. Daphne Koller (Stanford University) presented a technique to extend the power of single-cell profiling to population measurements by deconvolving gene expression patterns of different cell types in heterogeneous tissues and thus determining cell-type-specific differences in regulatory networks. Experimental and analytical advances such as these will clearly be important in moving the functional characterization of the genome and the analysis of biological regulatory networks into living animals and plants.

## Disparate mechanisms producing circadian oscillators

Finally, two talks focused specifically on the mechanisms that drive robust oscillations in bacterial and mammalian circadian clocks, respectively. In earlier work, Erin O'Shea (Harvard University) and colleagues successfully reconstituted the three-component cyanobacterial clock *in vitro*, demonstrating that oscillations are produced solely by the ordered phosphorylation and dephosphorylation of the three components. Here, O'Shea presented evidence detailing the mechanism that links the phase of the clock to day/night cues: variations in the ratio of ATP to ADP concentrations during daily photosynthetic cycles break the inherent symmetry and bias the cycle in the correct direction. John Hogenesch (University of Pennsylvania School of Medicine, Philadelphia, USA) used an RNAi screen to perform a systems-level analysis of the mammalian circadian clock. He perturbed the system quantitatively by controlling the strength of RNAi knock-down and measured quantitative changes in the phase and period of the clock, rather than focusing on the small set of targets producing true aperiodicity.

As evidenced by this meeting, panoramic snapshots of cellular behaviors can enhance our understanding of gene expression and gene regulatory networks. Future insights will likely be gained both from resolving the heterogeneity within these population-level measurements and watching the dynamic changes in these networks as cells respond to their environment.

**Author details**
[1]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94158, USA. [2]Department of Embryology, Carnegie Institution for Science, Baltimore, MD 21218, USA.