

MEETING REPORT

# Sequences to systems

Manolis Kellis<sup>1,2</sup> and John L Rinn<sup>1,3,4\*</sup>

## Abstract

A report of the seventh annual meeting on Systems Biology: Global Regulation of Gene Expression, 23-27 March 2010, Cold Spring Harbor, USA.

From a chemical standpoint, a cell can seem as a densely intertwined mess of molecules, but their emergent systems-level properties exhibit structure, coordination and order, leading to the orchestrated complexity we call life. As the seventh annual meeting on Systems Biology: Global Regulation of Gene Expression at the Cold Spring Harbor Laboratory showcased, systems biology aims to bridge the vast gap of knowledge between the pairwise molecular interactions of individual cellular components and the large-scale behavior of cellular circuits that result from their synergism. The talks given at the meeting exemplified a new definition of systems biology, where genomic technologies can revolutionize our understanding of classical problems by enabling molecular interactions to be linked to systems behavior in a mechanistic way.

A large number of definitions have been proposed for systems biology, including 'the study of a single pathway by someone with an attention-deficit disorder' humorously offered by Irv Bernstein and quoted in a keynote address by Steve Henikoff (Fred Hutchinson Cancer Research Center, Seattle, USA). Henikoff went on to demonstrate that Trithorax and Polycomb response elements that trigger cellular differentiation could be recognized by their patterns of histone turnover. He argued that histone variant H2A.Z is mutually exclusive with DNA methylation, drawing a novel connection between histones, their dynamics and the epigenetic inheritance of DNA methylation.

The importance of epigenetic information in mediating the stable inheritance of cellular states was underscored

in the context of human disease by the second keynote speaker, Kevin Struhl (Harvard Medical School, Boston, USA). By dissecting the transcriptional circuits responding to Src signaling that can yield stably transformed cells, Struhl uncovered a feedback loop mediated by the transcription factor NF $\kappa$ B that is sustainable for many generations in the absence of further signaling. Struhl went further to demonstrate that this pathway is targetable using FDA-approved drugs, opening up novel avenues of cancer therapeutics.

## Parsing the genome

The study of any system begins with a comprehensive list of parts and their interactions, and so a major theme of the meeting was the mapping of the building blocks of gene regulation: the control regions, the DNA words within them and their sentence structure.

On one hand, several talks focused on defining sites of transcriptional factor binding, histones and sites of DNA methylation at ever-increasing resolution and complexity. For example, two talks by Barbara Wold (California Institute of Technology) and Stephen Tapscott (Fred Hutchinson Cancer Research Center) presented data revealing thousands of binding sites for the master regulator of muscle differentiation MyoD, concluding that numerous additional functions and or combinatorics result in cell-fate decisions such as muscle differentiation. One of us (MK), presented a new algorithm specifically for revealing biologically meaningful combinations of chromatin marks, and their corresponding chromatin states, that reflect rich information about the underlying biological function of the associated genomic regions. The algorithms have distinguished more than 50 chromatin states, including 11 different classes of promoter regions, associated with different functional categories, and several classes of enhancer regions associated with different expression levels for downstream genes.

On the other hand, several talks focused on the DNA sequences predictive of regulatory activity. Eric Mendenhall (Massachusetts General Hospital, Boston, USA) presented remarkable data that distinct types of DNA sequences can recruit key chromatin factors such as Polycomb. Through multiple transgenic studies Mendenhall showed that he could insert DNA sequence motifs, even those 'that have never seen Polycomb before,' into

\*Correspondence: jrinn@bidmc.harvard.edu

<sup>1</sup>Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA

Full list of author information is available at the end of the article

the mouse genome that subsequently recruited Polycomb binding, suggesting a model of sequence-based recruitment of this complex. Eileen Furlong (European Molecular Biology Laboratory, Heidelberg, Germany) demonstrated the importance of combinatorial binding events in determining cell fate decisions, using a support vector machine to model transcription factor binding data and reveal the *cis*-regulatory motifs and motif combinations that define mesodermal fate in *Drosophila melanogaster*. A major conclusion was clear evidence that a single expression pattern could result from several different combinatorial codes of transcription factors and their dynamics. Arend Sidow (Stanford University, Stanford, USA) presented interesting insight into what determines nucleosome occupancy during the differentiation of distinct blood cell types both *in vitro* and *in vivo*. He presented clear evidence that *in vitro* nucleosomes associate with sequence-specific motifs, such as GC motifs, and exclude others, such as AT sites, whereas *in vivo* these sequence preferences are largely overridden. Moreover, Sidow demonstrated that nucleosome positioning is globally remodeled across different states of cellular differentiation.

Collectively these studies demonstrate the power of chromatin-level and DNA-level information for predicting regulatory activity, the importance of word combinations and mark combinations, and the interplay of static sequences and dynamic cell-type specific changes in defining the processes of differentiation and, more generally, gene regulation.

### **Unconventional interactions of the secondary type**

The late Ed Lewis famously said 'Progress will still need to be driven by the logic of genetics and by further increases in abstraction.' This sentiment was echoed at this year's meeting, most prominently in the unveiling of numerous unconventional protein-DNA and protein-RNA interactions.

A dramatic example was presented by Heng Zhu (Johns Hopkins Medical Center, Baltimore, USA), who has used human proteome arrays to map global DNA-binding interactions. His analysis identified a striking and atypical interaction between kinases and DNA. Specifically, Zhu identified a carboxy-terminal domain of the kinase ERK2 that binds to DNA with nanomolar affinity and through this interaction serves as a transcriptional repressor. He also presented initial evidence for more than 300 unconventional DNA-binding proteins, including unusual suspects such as RNA-binding proteins and mitochondrial proteins.

Howard Chang (Stanford University, Stanford, USA) presented a comprehensive study demonstrating modular binding of the large intergenic noncoding RNA (lincRNA) HOTAIR to chromatin-modifying complexes

such as PRC2, the histone demethylase LSD1, and COREST, a DNA-binding protein. Chang presented a model, based on functional assays, whereby HOTAIR binds all three of these unconventional RNA-binding proteins through a different specific structured domain and in this way can guide them specifically to their target genomic loci. Thus, lincRNAs could serve as adaptors to facilitate the formation of specific RNA-protein interactions. One of us (JR) presented evidence that such mediation is a global phenomenon that extends beyond HOTAIR and includes hundreds of lincRNAs interacting with chromatin-modifying complexes and other DNA-binding proteins involved in cancer. Overall, the genome could be replete with unorthodox RNA-binding interactions serving as RNA adaptors to facilitate the establishment and or maintenance of specific epigenetic states. Tom Gingeras (Cold Spring Harbor Laboratory, New York, USA) expanded the horizons of unorthodox RNA functions, presenting a rich diversity of noncoding RNA molecules that localize to distinct cellular compartments and are derived from unconventional processes, emphasizing the further need for abstraction in our understanding of the mammalian transcriptome and the potential regulatory roles of noncoding RNAs.

### **Close encounters of the three-dimensional type**

At this year's meeting the genome came to life in three dimensions, as a result of 'proximity ligation' techniques that define the three-dimensional neighbors of a given genomic sequence relative to all others in the genome. This reveals potentially important regulatory interactions mediated by DNA looping and compartmentalization.

Job Dekker (University of Massachusetts, Worcester, USA) described work using several variants of chromatin-conformation capture (3C) to reveal these interactions. Dekker and his colleagues have used carbon-copy 3C (5C) to resolve the three-dimensional structure of the beta-globin locus in both the activated and repressed states. He proposed a model by which a large genomic-linker region serves to open up or compact the locus upon activation or repression, respectively. Erez Lieberman-Aiden (Broad Institute, Cambridge, USA) has expanded the technique to take on the entire genome, and reported that megabases of DNA fold into a dense, highly organized, and knot-free structure known as a 'fractal globule', thus providing a glimpse of the three-dimensional genome.

Using a different proximity ligation approach called ChIA-PET, Yijun Ruan (Genome Institute of Singapore, Singapore) has revealed three-dimensional genomic architecture in the context of transcriptional regulation. Ruan specifically demonstrated tertiary DNA interactions mediated by transcription factors and RNA polymerase II. An interesting connection to this finding was presented

by Martin Hemberg (Children's Hospital, Harvard Medical School, Boston, USA) where he described that numerous enhancers, which often result from long-range DNA interactions, are associated with RNA polymerase II and are transcribed at low levels.

Collectively these studies have added a new dimension to systems biology and provide novel insights into the three-dimensional architecture of the human genome.

### Systematic evolution

The evolutionary dimension of gene regulation was addressed in regard to the selective pressures and plasticity of regulatory regions and elements. Ollie Rando (University of Massachusetts, Worcester, USA) presented an evolutionary analysis of nucleosome positioning in 12 ascomycete species. Perhaps most notably, the nucleosome positioning corresponded to specific gene sets, such as mitochondrial genes or splicing genes, and differed in physiologically relevant ways in different species.

Continuing the theme of evolution and binding localization, Mike Snyder (Stanford University, Stanford, USA) presented an unprecedented dataset comprising the binding sites for several key transcription factors across ten different human individuals. Snyder showed that genetic variation between individuals had resulted in the loss and gain of binding sites for key transcriptional players such as RNA polymerase II and the transcription factor NFκB. Thus, individual variation results in unique binding patterns of key factors, which should provide potential insights for personalized medicine and disease susceptibility.

Alexander Stark (IMP, Vienna, Austria) presented comparative work on the DNA binding of transcriptional regulators of mesoderm development in six *Drosophila* species at a range of evolutionary distances from each other. He showed that not only were the locations of the peaks representing binding-site occupancy conserved, but also the intensity of the peaks, suggesting that lower-occupancy sites are also functional and are selectively maintained. He presented evidence that these sites were subsequently occupied at later developmental stages, suggesting that lower occupancy may be associated with priming for subsequent activation.

### Cell circuits and dynamics

The ultimate goal in systems biology is to define cellular circuits and their resulting cellular outputs. This year's meeting reported unprecedented progress in mapping cell circuits in response to specific stimuli, perturbing them and monitoring their cellular responses. A comprehensive example was presented by Aviv Regev (Broad Institute, Cambridge, USA), who described work with a model mammalian cell system - mouse dendritic cells - which trigger distinct inflammatory or antiviral responses

depending on the input stimuli. Regev's strategy was to monitor the complete transcriptional response to each stimulus, identify the key regulators (transcription factors, chromatin-modifying complexes and RNA-binding proteins) in each pathway, and then individually perturb the actions of hundreds of these regulators via stable gene knockdown and monitor the resulting transcriptional changes. This not only identified numerous new regulators across these critical cellular pathways, but also revealed the cooption of cell-cycle circuits in the antiviral response. This approach to systems biology could also be applied to numerous other cell-based models.

Underlying seemingly stable cell circuits are multiple additional layers of modulation, which include both stochastic single-cell variability in transcript levels and regulated processes of post-transcriptional inhibition and splicing. Alexander Van Oudenaarden (Massachusetts Institute of Technology, Cambridge, USA) demonstrated that cell-fate decisions in *Caenorhabditis elegans* intestinal specification arise from an underlying process of variable transcription levels. These stochastic processes resulted in threshold-dependent ON or OFF decisions of master regulators that guide cellular differentiation. David Bartel (Whitehead Institute, Cambridge, USA) described how the dramatic effects of post-transcriptional regulation by microRNAs (miRNAs) can result from even the slightest fine-tuning of transcript levels and translation efficiency. In a potential reversal of the dominant dogma in miRNA regulation, he also presented evidence from mammalian cells that for two miRNAs at least, the major effect of targeting appears to be transcript degradation rather than translational inhibition. Chris Burge (Massachusetts Institute of Technology, Cambridge, USA) presented studies that revealed a remarkable complexity in mRNA transcript splicing and how these isoforms are altered in cancer.

Collectively, this year's meeting revealed new insight into genomic sequences, binding sites and structures that has moved systematic study of gene regulation to new dimensions of both understanding and further abstraction in systems biology.

#### Author details

<sup>1</sup>Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA. <sup>2</sup>Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. <sup>3</sup>Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA 02215, USA. <sup>4</sup>Department of Pathology, Harvard Medical School, Boston, MA 02115, USA.

Published: 25 May 2010

doi:10.1186/gb-2010-11-5-303

Cite this article as: Kellis M, Rinn JL: Sequences to systems. *Genome Biology* 2010, 11:303.