

MEETING REPORT

Genome-wide insights into eukaryotic transcriptional control

James A Goodrich* and Jennifer F Kugel*

Abstract

A report of the Keystone Symposium on Dynamics of Eukaryotic Transcription during Development, Big Sky, Montana, USA, 7-12 April 2010.

The 2010 Keystone Symposium focusing on mechanisms of eukaryotic transcriptional regulation featured a strong emphasis on genomic approaches. Many presentations included data that coupled chromatin immunoprecipitation (ChIP) assays to deep sequencing (ChIP-seq) or tiling arrays (ChIP-chip) in order to map the locations of transcription factors and RNA polymerase II (Pol II) across the genomes of organisms from yeast to humans. In addition, biochemical techniques such as permanganate footprinting, nuclear run-on, and nuclease digestion were applied to cellular chromatin and coupled to deep sequencing. Here, we group the presentations that focused primarily on genomics into three general categories: promoter-proximal paused polymerases, chromatin and nucleosomes, and networks of transcriptional regulation.

Promoter-proximal paused polymerases

Historically, regulation of transcription was thought to occur primarily at the point of recruiting Pol II and its accessory factors to the promoters of genes. Now a battery of genomic studies is revealing that transcriptional regulation occurs at a post-initiation step at thousands of genes in both *Drosophila* and mammals. The signature of such genes is the presence of a promoter-proximal paused Pol II molecule - one that has initiated transcription and is poised for an activation signal in order to continue transcribing. Rick Young (Whitehead Institute and Massachusetts Institute of Technology, Cambridge, USA) in his keynote address provided a mechanism for how activation of paused polymerases

can occur. In human embryonic stem cells the transcription factor *c-Myc* functions as a 'pause release' factor at approximately one-third of genes. ChIP-seq data have shown that *c-Myc* associates exclusively with transcribed genes near their start sites, a signature unique among stem cell transcription factors. Moreover, chemical inhibition of *c-Myc*, or its knockdown, decreased the elongating form of Pol II (phosphorylated on serine 2) but not the initiated form of Pol II (phosphorylated on serine 5).

Karen Adelman (National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, USA) described her research showing that Pol II pausing followed by regulated release does not function solely as a transcriptional on/off switch. ChIP-chip studies showed that, in *Drosophila* cells, the pause-inducing factors negative elongation factor (NELF) and DRB sensitivity inducing factor (DSIF) occupy genes that are actively transcribed (genes with uniform Pol II distribution and with Ser2-phosphorylated Pol II). Thus, NELF is present at active genes and seems to be fine-tuning transcription. Computational analysis of NELF-dependent genes revealed an enrichment of GAGA sites, initiator (Inr), and TATA sequences, as well as a downstream motif centered at the +30 position that could function to control pausing. In a unique application of deep sequencing technology, Dave Gilmour (Pennsylvania State University, University Park, USA) presented data that revealed regions of melted DNA associated with a paused polymerase. *In vivo* permanganate footprinting, which detects single-stranded thymines in the DNA comprising a transcription bubble, was coupled to Pol II ChIP-seq to reveal polymerases paused at specific positions across the *Drosophila* genome. At least 10% of genes showed permanganate reactivity centered in the +20 to +60 region, which correlates nicely with Gilmour's *in vitro* biochemical data. Of these genes, 80% showed NELF occupancy and 50% had GAGA factor, consistent with these two factors controlling Pol II pausing.

The pioneer of paused polymerases, John Lis (Cornell University, Ithaca, USA), described a powerful genomic technique coupling nuclear run-on assays to deep sequencing (GRO-seq) to map the location, density, and

*Correspondence: jennifer.kugel@colorado.edu, james.goodrich@colorado.edu
Department of Chemistry and Biochemistry, University of Colorado, 215 UCB,
Boulder, CO 80309-0215, USA

orientation of transcriptionally engaged polymerases. GRO-seq revealed widespread divergent transcription at the promoters of mammalian genes containing CpG islands; however, at TATA-containing promoters, transcription was mostly unidirectional. GRO-seq was also applied to embryonic stem cells and mouse embryonic fibroblasts, and approximately 35% of genes had paused polymerases. In *Drosophila* cells, GRO-seq correlated reasonably well with ChIP-seq against Pol II, indicating that most of the Pol II detected around the start sites of genes is transcriptionally engaged.

Chromatin and nucleosomes

Eukaryotic transcription occurs in the context of chromatin. Critical to understanding transcriptional regulation is deciphering how nucleosomes are positioned, modified, and rearranged such that transcription factors and Pol II can access the DNA. Jon Widom (Northwestern University, Evanston, USA) described experiments that used deep sequencing to map nucleosome positions across the yeast genome with base-pair resolution. Taking advantage of a histone H4 Ser-to-Cys point mutation and hydroxyl radical chemistry, the exact position of DNA wrapped around each nucleosome was determined. This accurate map revealed information about the distances between nucleosomes, which has implications for understanding higher order chromatin structures, such as the 30 nm fiber.

Steve Henikoff (Fred Hutchinson Cancer Research Center, Seattle, USA) described a new technique to measure nucleosome replacement across the *Drosophila* genome to obtain insight into nucleosome dynamics. His group sequenced nucleosomal DNA isolated by biotin tagging and affinity purifying histone H3.3, the replication-independent histone variant that can replace replication-coupled H3. They identified sites of nucleosome turnover, and comparison of these data with the genome-wide occupancy of Polycomb and Trithorax complexes revealed that nucleosomes turn over at both silent and active chromatin, albeit at different rates. Combining elements of chromatin structure and dynamics, Shirley Liu (Dana-Farber Cancer Institute, Boston, USA and Harvard School of Public Health, Cambridge, USA) described ChIP-seq experiments that monitored the occupancy of nucleosomes containing di- or tri-methylated lysine 4 on histone H3 (H3K4me2/3) in a prostate cancer cell line before and after stimulation of the androgen receptor. She described a computational approach to identify sites at which an H3K4me2-modified nucleosome was evicted whereas adjacent nucleosomes were retained. Importantly, the sites of nucleosome eviction successfully predicted transcription factor binding sites for NK3 homeobox 1 (NKX3.1) and the helix-turn-helix transcription factor Oct1, which

were subsequently verified by ChIP linked to quantitative PCR.

Networks of transcriptional regulation

With genome-wide technologies becoming more accessible and increasingly high throughput, it is becoming possible to analyze how multiple factors together contribute to transcriptional regulation across a genome. For example, Frank Pugh (Pennsylvania State University) described the use of ChIP-chip to identify occupancies of 200 factors at each of the 6,000 genes in yeast. Approximately 600 genes had 75 or more proteins bound. When genes regulated by Spt-Ada-Gcn5 acetyltransferase (SAGA) were compared with transcription factor IID (TFIID)-regulated genes, the SAGA genes had a larger repertoire of bound factors, and these were also present at higher levels. Pugh also described ChIP-exo, which incorporates exonuclease digestion into a ChIP-seq assay to map transcription factor binding sites with base-pair resolution. Approximately 95% of Reb1 occupancy sites identified with this technique were within 1 bp of a Reb1 cognate sequence.

Bas van Steensel (Netherlands Cancer Institute, Amsterdam, the Netherlands) defined different types of chromatin domains on the basis of differential occupancies of 53 proteins across the *Drosophila* genome using DNA adenine methyltransferase identification (DamID). Hidden Markov modeling of the 53-dimensional dataset revealed five chromatin domains defined by unique combinations of proteins. The most surprising domain occupied about 55% of the genome, showed little transcriptional activity but no known repressive chromatin modifications, and is poorly understood. Interestingly, preferences emerged for which chromatin domain is adjacent to other chromatin domains.

To analyze and ultimately understand regulatory processes in a global manner, Frank Holstege (University Medical Center Utrecht, the Netherlands) described the use of high-throughput microarray experiments to define mRNA profiles after knockout of signaling pathway components and transcription machinery in yeast, with approximately 1,000 knockouts analyzed so far. Analyses of gene expression from cells with double-deletions of kinases and phosphatases have revealed three phenotypes of expression: complete redundancy, quantitative redundancy, and incongruent redundancy. The latter category was unexpected yet is relatively common and can arise from partial redundancy between two factors coupled to unidirectional inhibition of one factor by the other.

Rather than focusing on multiple transcription factors, work presented by Michael Wilson (Cambridge Research Institute, Cancer Research UK, Cambridge, UK) focused on multiple genomes. With the goal of exploring the evolution of gene regulation, he used ChIP-seq against

two transcription factors (CCAAT enhancer binding protein alpha (CEBPA) and hepatocyte nuclear factor 4 alpha (HNF4A)) and determined their locations across five vertebrate genomes. The data revealed significant differences in factor occupancy between species. Ultra-conserved binding events that aligned in all five species occurred only rarely. When an expected binding event was missing in one species, occupancy by that factor within 10 kb was observed only about half the time. From this work, the plasticity of regulatory interactions during divergence among vertebrates is emerging.

Lastly, the network of transcriptional control in response to estrogen stimulation was described by Lee Kraus (Cornell University). His laboratory used GRO-seq to determine how the map of transcriptionally engaged Pol II molecules changes in human breast cancer cells during a time course of estrogen treatment. The sensitivity of this approach meant that they could identify about tenfold more protein-coding genes rapidly

regulated by estrogen than did microarrays. Moreover, the data revealed changes in Pol I and Pol III transcripts, divergent transcription, antisense transcription, microRNA transcription, and transcription from non-genic regions. From these data a new picture of signal-induced transcriptional regulation emerges that was not obtainable using previous genomic technologies.

In summary, this meeting was filled with applications of genomic technologies that revealed new insight into mechanisms of transcriptional regulation, involving paused polymerases, chromatin structure, and the interplay between many protein factors.

Published: 7 June 2010

doi:10.1186/gb-2010-11-6-305

Cite this article as: Goodrich JA, Kugel JF: Genome-wide insights into eukaryotic transcriptional control. *Genome Biology* 2010, **11**:305.