[Fly 1:2, 123-124; March/April 2007]; ©2007 Landes Bioscience

Meeting Report

New Tools, Resources for Gene Regulatory Analysis in Drosophila

Marc S. Halfon[1,2]
David N. Arnosti[3,4]

[1]Department of Biochemistry; [2]Center of Excellence in Bioinformatics; State University of New York at Buffalo; Buffalo, New York USA

[3]Department of Biochemistry and Molecular Biology; [4]Program in Genetics; Michigan State University; East Lansing, Michigan USA

Correspondence to: Marc S. Halfon; Department of Biochemistry; Center of Excellence in Bioinformatics; State University of New York at Buffalo; 140 Farber Hall, 3435 Main Street; Buffalo, New York 14214-3000 USA; Tel.: 716.829.8126; Fax: 716.829.2725; Email: mshalfon@buffalo.edu/ David N. Arnosti; Department of Biochemistry and Molecular Biology; Michigan State University; East Lansing, Michigan 48824-1319 USA; Tel.: 517.432.5504; Fax: 517.353.9334; Email: arnosti@msu.edu

Original manuscript submitted: 04/09/07
Manuscript accepted: 04/16/07

Previously published online as a *Fly* E-publication:
http://www.landesbioscience.com/journals/Fly/abstract.php?id=4291

KEY WORDS

enhancer, cis-regulatory module, transcription, comparative genomics, bioinformatics, ChIP-chip

Characterization of cis-regulatory information in genomes poses some of the thorniest questions in biology today, with relevance to areas such as population genetics, evolution, and disease research. Progress reported at the 48th Annual Drosophila Research Conference provides insights on how developing technology and bioinformatics should accelerate our findings in this area, as well as highlighting bottlenecks that call for creative new approaches. Talks and posters related to these topics include the identification of cis-regulatory elements, as well as their functional characterization.

One motive for sequencing the *Drosophila pseudoobscura* genome in 2002 was to identify regulatory regions, which were expected to show a higher degree of conservation than other intergenic regions. Although such alignments have proved useful with select vertebrate genes, the apparent high level of constraint (indicating conserved functionality) on much of the *D. melanogaster* genome, together with the rapid turnover of binding sites within regulatory regions, meant that the pairwise alignments of intergenic regions was less informative than would be desired. Exploiting the greater phylogenetic breadth of the recently sequenced 12 Drosophila genomes, this issue is being revisited by Eisen and collaborators at Berkeley with a more detailed picture of regulatory regions of *even-skipped (eve)*. Chromatin immunoprecipitation/gene chip (ChIP-chip) analysis shows gap proteins such as Giant and Knirps binding to the enhancer regions, as expected. Focusing on closely related species of the *melanogaster* subgroup (*D. melanogaster, D. sechelia, D. simulans*), where alignments can be made with high confidence, it appears that the actual nucleotides within sites identified by DNAseI footprinting are conserved to an extent approaching that of protein-coding sequences, while overall conservation within the enhancer is much lower. More distant species show a much lower conservation of binding sites, however, suggesting enhancer structural reorganization over this time span.

Analysis of *eve* loci in more divergent Sepsid and Tephritid species indicates that functionally equivalent stripe enhancers can be identified by clusters of binding sites, although sequence conservation is too low to align to Drosophila sequences. Interestingly, an analysis of 280 Drosophila enhancers by the Halfon laboratory (University at Buffalo) suggests that dense binding site clustering is not a general property of all enhancers, thus approaches based on such clustering may be of limited utility.

Interestingly, Eisen showed that the pairwise alignments of the Tephritid *R. juglandis* and *C. capitata eve* regions revealed islands of conservation interrupted by less well conserved sequence, reminiscent of inter-vertebrate alignments, and suggesting that sequencing these flies with larger genomes (3–4 times that of Drosophila) might provide a useful tool for identifying regulatory sequences.

A workshop focusing on the dozen Drosophila genomes also highlighted other approaches to cis-regulatory analysis. Extending their work with yeast and vertebrate genomes, Kellis (MIT) established a genome-wide set of conserved motifs enriched in introns, UTR, and proximal promoter regions that represent likely transcription factor binding sites and targets for miRNAs. Cross-correlations with gene expression data links motifs associated with tissue-specific expression. Manak from NimbleGen summarized collaborative work with Ren and colleagues at UC San Diego that uses tiling arrays to find chromatin marks characteristic of enhancers (histone H3 lysine monomethylation) and promoters (H3 lysine trimethylation).

Thus, cis-regulatory regions are being identified through a combination of sequence conservation, overrepresentation of motifs linked to gene expression patterns, and most importantly, direct identification of proteins bound to genomic targets using ChIP methods. The identification of cis-regulatory elements is only the first step to unlocking genomic information, however. Studies of individual genes highlighted aspects of enhancer function that inform the more global studies noted above.

An exhaustive dissection of the *slp1,2* locus by Fujioka (Jaynes lab, Thomas Jefferson University) revealed patterning information for embryonic expression contained in multiple redundantly-acting elements located 5' of the *slp1* promoter and in the *slp1,2* intergenic region. While not the first demonstration of enhancer redundancy, this study underlines that transcriptional regulatory information can be distributed over widely spaced regions, thus evolutionary loss of individual binding sites might be compensated by changes in distal sequences. Interestingly, distal and proximal enhancers 5' of *slp1* show complex nonadditive relationships as well (L. Prazak, Gergen lab, Stony Brook).

A key step in traditional models of gene activation involves recruitment of RNA polymerase II and basal factors, and genome-wide ChIP surveys highlight binding of transcriptional machinery in promoter regions, which one might imagine serves as a proxy for active genes. Nibu (Weill Medical College, Cornell) reported on the Snail short-range transcriptional repressor acting in the blastoderm embryo. Using ChIP analysis of embryos, they find that the signal for the Dorsal activator is dramatically diminished when Snail is present at repressed target genes such as *rho, vnd,* and *sog,* suggesting that activator occupancy is limiting. In a similar study, Wang (Gergen lab, SUNY Stony Brook) used ChIP to measure repression of *slp1* by Runt and Ftz in embryos, and find that RNA polymerase is still present at the promoter. In this case not only was polII present, but it was CTD-phosphorylated, suggesting that transcription was initiated but elongation abrogated. McKay and colleagues (Mann lab, Columbia), used a novel ChIP method in which a *lacO* sequence is inserted in a sequence of interest. By successive ChIP's directed against tissue-specific expressed LacI (binding to the *lacO* sequence) and a DNA-binding protein of interest, binding to a regulatory region in a specific group of cells can be assessed. Similar to these other studies, they find pol II present at the *dll* promoter even in abdominal cells where the gene is repressed. Furthermore, they find that the *dll* enhancer remains in contact with the promoter whether or not the enhancer is active. Perhaps Lis's finding relating to polymerases at the inactive but "poised" *hsp70* gene is a common property of promoters.

Some of the differences between the structures of regulatory regions in related species might be functionally inconsequential, or might reflect extensively co-adapted systems, making it challenging to tease out more subtle aspects of cis-regulatory grammar. However, alterations in cis-regulatory information are also linked to quantitative traits differing between individuals, and more sophisticated analysis of the transcriptional code will likely profit from population studies. MacKay (North Carolina State) described inbred reference populations that her laboratory has generated differing significantly in behavior, starvation resistance, and sensory bristle number, among other traits. Forty of these lines are being deposited with the Bloomington Stock Center, and a proposal is pending to obtain 4X sequencing coverage of each genome. This resource promises to provide fertile ground for identifying regulatory changes that are sufficient to cause observable phenotypic changes.

FUTURE WORK

Most global studies of cis-regulatory regions, whether based on pure in silico genomic approaches or ChIP experiments, are focused on the identification of enhancer sequences. A deeper understanding of enhancer design, such as the mathematical modeling of synthetic enhancers to identify a cis-regulatory "grammar" (Arnosti, Michigan State), will permit quantitative prediction of their functional output, which is the feature that selection acts upon.

ChIP-chip studies and the availability of more genomes add welcome new tools to the Drosophila arsenal; however a bottleneck remains the lack of high-throughput ways for testing regulatory element function in vivo. Bioinformatic identification of cis-regulatory information by comparative genomics, as described by Kellis, or Brody, using the *cis*Decoder and EvoPrinter algorithms (Odenwald lab, NIH), requires functional assays for validation. Sharply increased from the previous year was the number of labs reporting the use of phage recombination systems to increase the efficiency of transformation, and, importantly, to provide common insertion sites to control for position effects (Markstein, Harvard). However, true high-throughput methods for transformation would significantly accelerate the rate of in vivo validation.