October 1, 2002

Decoding cis-regulatory DNAs in the Drosophila genome

Michele Markstein, *University of Massachusetts - Amherst*
Michael Levine

Decoding *cis*-regulatory DNAs in the *Drosophila* genome

Michele Markstein* and Michael Levine†

Cis-regulatory DNAs control the timing and sites of gene expression during metazoan development. Changes in gene expression are responsible for the morphological diversification of metazoan body plans. However, traditional methods for the identification and characterization of *cis*-regulatory DNAs are tedious. During the past year, computational methods have been used to identify novel *cis*-DNAs within the entire *Drosophila* genome. These methods change the way that *cis*-DNAs will be analyzed in future studies and offer the promise of unraveling complex gene networks.

Addresses
*Department of Molecular Cell Biology, Division of Genetics & Development, 401 Barker Hall, University of California, Berkeley, California 94720, USA
*†Committee on Developmental Biology, University of Chicago, Cummings Life Sciences Center, 755, 920 East 58th Street, Chicago, Illinois 60637, USA

Current Opinion in Genetics & Development 2002, 12:601–606

0959-437X/02/$ – see front matter

Abbreviations

ChIP	chromosome immunoprecipitation
eve	*even-skipped*
hbr	*heartbroken*
sog	*short-gastrulation*
Su(H)	Suppressor of Hairless
zen	*zerknüllt*

Introduction

A clear revelation of the post-genome era is that organismal complexity does not correlate with gene number. The fruitfly, *Drosophila melanogaster*, contains <14,000 genes [1], whereas the nematode worm *Caenorhabditis elegans* — a considerably less complex animal — contains ~20,000 genes [2]. Complexity is more likely to be a manifestation of the total number of gene expression patterns that are produced during development. Perhaps *Drosophila* contains a greater number of *cis*-regulatory DNAs than *C. elegans*, and thereby exhibits more sophisticated morphologies and behaviors. The average fly gene might be regulated by three or four different *cis*-DNAs, and consequently, the fly genome could produce something like 50,000 distinct patterns of gene expression. If worms contain an average of just one or two *cis*-DNAs per gene, then the worm genome might produce half the total number of gene-expression patterns than flies, even though it contains 50% more genes. Here, we discuss recent efforts to employ computational methods to identify cis-regulatory DNAs within complex metazoan genomes.

Pre-genomics analysis of *cis*-regulatory DNAs

There are several classes of *cis*-regulatory DNAs, including enhancers [3], silencers [4], and insulators [5]. Enhancers represent the most thoroughly analyzed type of *cis*-regulatory DNA. The characterization of enhancers in transgenic worms, flies, sea urchins, ascidians, fish, frogs, chicks, and mice, suggests that a typical enhancer mediating cell-type specific expression is 300 bp to 1 kb in length and contains clustered binding sites for both transcriptional activators and repressors [6]. In flies, a typical enhancer such as the *even-skipped* (*eve*) stripe 2 enhancer [7] contains a total of 10 binding sites for three different sequence-specific transcription factors; two of the factors function synergistically to activate gene expression, whereas the third represses transcription in inappropriate cell types.

Emergence of computational methods: from *in silico* to *in vivo*

The detailed characterization of cell-specific enhancers in transgenic embryos is a laborious process. Consequently, comparatively few enhancers — something like 100 in all animals combined — have been examined [6]. Although several computational approaches have been developed to identify *cis*-regulatory motifs and regions *in silico*, until recently, very few of these predictions have been tested in multicellular animals [8,9•]. In the past year, bioinformatics methods have been used to identify authentic enhancers within the *Drosophila* genome [10•–13•]. Although these are early days for the computational identification of enhancers, the methods we briefly summarize in this review have permanently changed the way such *cis*-regulatory DNAs will be characterized in the future. Several of these methods are available as web-based tools at the URLs listed in Box I.

Computing Dorsal target enhancers

We begin with the dorsal–ventral patterning of the early *Drosophila* embryo. Localized patterns of gene expression depend on a sequence-specific transcription factor called Dorsal [14]. The Dorsal protein is distributed in a broad nuclear gradient, with peak levels in ventral regions and progressively lower levels in lateral and dorsal regions (Figure 1a). The Dorsal gradient differentially regulates as many as 25 different target genes in a concentration-dependent manner. Some of the target genes are activated only by high concentrations of the gradient. As a result, these genes exhibit localized expression in ventral regions that form the mesoderm. By contrast, other Dorsal target genes are regulated by low levels of the gradient, and are activated, or repressed, in lateral regions that form the neurogenic ectoderm.

A total of seven different Dorsal target *cis*-regulatory DNAs were characterized through the traditional method of attaching random DNA fragments from the 5′ flanking regions of the target genes to a *lacZ* reporter gene. These *lacZ* fusion genes were individually integrated into the *Drosophila* genome using *P*-element gene transfer.

Box 1.

Web-based tools to identify *cis*-regulatory DNAs.

Identify binding site clusters with:

- Cister
  http://zlab.bu.edu/~mfrith/cister.shtml
- Fly Enhancer
  http://flyenhancer.org
  (includes sister sites for worm and plant)
- Cis-analyst
  http://www.fruitfly.org/cis-analyst/
- Target Explorer
  http://trantor.bioc.columbia.edu/search_for_BS/

Detect novel shared motifs with:

- Improbizer
  http://www.cse.ucsc.edu/~kent/improbizer/
- BioProspector
  http://bioprospector.stanford.edu/
- MEME
  http://meme.sdsc.edu/meme/website/intro.html

Embryos were collected from transgenic strains, and stained for *lacZ* gene expression to determine whether any of the DNA fragments were sufficient to recapitulate authentic aspects of the endogenous gene expression patterns. In this way, minimal tissue-specific enhancers were identified for two genes expressed in the mesoderm, *twist* and *snail* [15–17], and two genes expressed in the neurogenic ectoderm, *single-minded* and *rhomboid* [18,19]. Another three *cis*-regulatory DNAs, silencers, were likwise identified for the *tolloid*, *zerknüllt* (*zen*), and *decapentaplegic* genes [20–22]. These silencers keep expression off in ventral and lateral regions in response to high and low levels of the Dorsal gradient.

The characterization of these seven *cis*-regulatory DNAs required several years of effort from several laboratories. Markstein *et al.* [10•] recently developed a computational method for identifying clusters of Dorsal-binding sites in the *Drosophila* genome. This method led to the rapid identification of a new Dorsal target enhancer. The Dorsal protein binds DNA as an obligate dimer, and recognizes a spectrum of sites with dyad symmetry [23]. Dorsal-binding sites have been defined by SELEX assays as well as DNaseI and chemical footprinting assays. These studies provided concise consensus sequences representing 106 optimal, high-affinity binding sites [23]: GGGWWWWCCM and GGGWDWWWCCM (W = A or T, M = C or A, D = A or T or G). Four optimal Dorsal sites conforming to these consensus matrices have been observed within a 400 bp region of the 600 bp *zen* silencer sequence [23]. These high-affinity sites mediate the repression of *zen* even in dorso-lateral regions where there are vanishingly low levels of the Dorsal gradient. A survey of the entire *Drosophila* genome for clusters containing at least three optimal Dorsal-binding sites within 400 bp identified 15 novel clusters in addition to the *zen* cluster. As only four Dorsal clusters would be expected by chance alone — using the binomial distribution as an approximation — the occurrence of 16 clusters is consistent with the notion that positive evolutionary selection has maintained the functional integrity of the clusters.

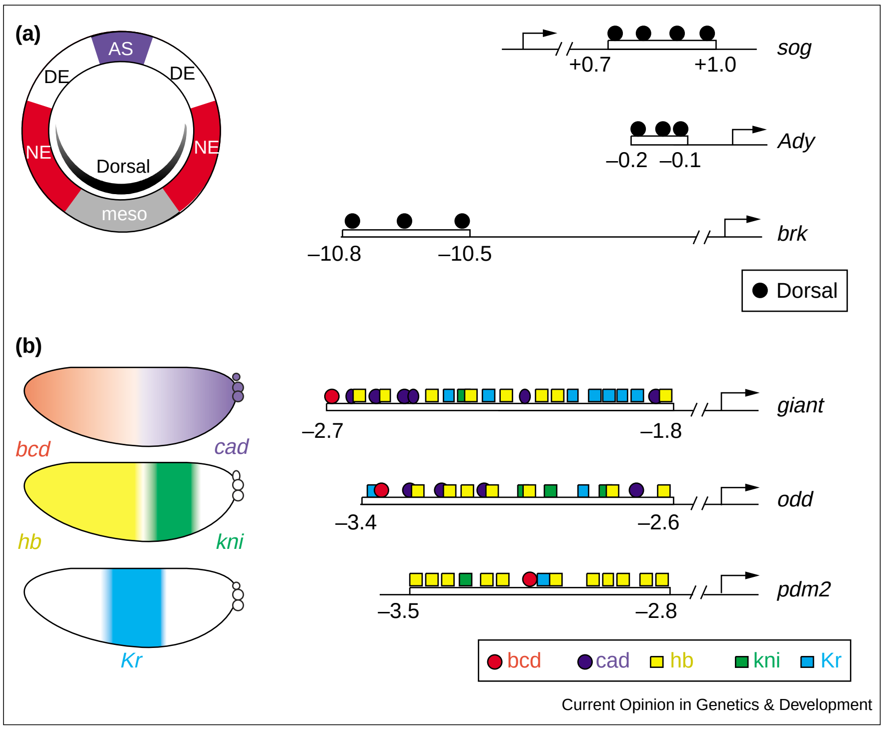
One of the novel clusters is located within the first intron of the *short-gastrulation* (*sog*) gene, which was identified as a potential target of the Dorsal gradient in genetic studies [24]. A 393 bp genomic DNA fragment that encompasses this cluster was placed 5′ of a *lacZ* reporter gene containing a minimal, 42-bp ‘naïve’ promoter (Figure 1a). The *sog–lacZ* fusion gene exhibits broad lateral stripes of expression in transgenic embryos that are similar to those observed for the endogenous gene. Another four of the novel clusters are associated with genes that exhibit early, localized expression across the dorsal–ventral axis. One of these clusters is located ~10.5 kb 5′ of the *brinker* gene, which is a known genetic target of the Dorsal gradient [25]. These new clusters are currently being tested for enhancer activities in transgenic embryos. It is possible that 5 of the 15 novel Dorsal binding clusters in the fly genome, one-third, correspond to authentic enhancers.

Computing *Suppressor of Hairless* target enhancers

A similar approach was used by Rebeiz *et al.* [11•] to examine the clustering of Suppressor of Hairless (Su[H]) binding sites in the *Drosophila* genome. Su(H) is a sequence-specific transcription factor that mediates Notch signaling in a variety of developmental processes in *Drosophila* [26]. A major function of Notch signaling and Su(H) is to inhibit neurogenesis. This is accomplished, at least in part, by the activation of a set of sequence-specific transcriptional repressors encoded by the *Enhancer of Split* gene complex. These Enhancer of split proteins bind to specific sites within the regulatory regions of *achaete–scute* genes, which are essential for the development of neurons within the central and peripheral nervous systems [26]. The Su(H) protein is localized within the nucleus, where it appears to function as a transcriptional repressor in the absence of Notch signaling. Upon signaling, the Notch receptor is proteolytically processed. The Notch intracytoplasmic domain (Notch^{IC}) is released from the cell surface and interacts with Su(H) within the nucleus [27]. The Su(H)–Notch^{IC} complex functions as a transcriptional activator and induces the expression of *Enhancer of split* genes, which were formerly repressed by Su(H) in the absence of Notch signaling [26].

Several Notch/Su(H) target genes have been characterized, and all of the regulatory regions that were identified contain clusters of Su(H)-binding sites conforming to the consensus sequence YGTGDGAA [11•]. A search of the *Drosophila* genome for statistically significant clustering of high-affinity Su(H) sites — defined by YGTGRGAA and CGTGDGAA — identified 46 novel clusters, ranging in size from 300 bp to 5 kb. On the basis of expression patterns of the associated genes, and the analysis of one of the clusters in transgenic flies, it would appear that at least

Figure 1



Bioinformatics screens for *Drosophila* enhancers. **(a)** Summary of the Dorsal nuclear gradient and target enhancers. The diagram on the left represents a cross-section through a 2-h embryo. There are peak levels of nuclear Dorsal protein in ventral regions and lower levels in lateral regions. The nuclear gradient initiates the differentiation of the mesoderm (meso), neurogenic ectoderm (NE), dorsal epidermis (DE), and amnioserosa (AS) by regulating several target genes in a concentration-dependent fashion. Some of the Dorsal-binding clusters identified in the *Drosophila* genome are associated with genes that exhibit asymmetric patterns of expression across the dorsal ventral axis. The Dorsal binding cluster associated with the *sog* gene is located in the first intron of the transcription unit. This cluster mediates broad lateral stripes of gene expression in the neurogenic ectoderm. The *Ady* gene is expressed in the ventral mesoderm, and there is a cluster of optimal Dorsal sites located ~100 bp 5′ of the transcription start site. The *brk* gene is expressed in lateral stripes in the neurogenic ectoderm, and there is a cluster of Dorsal-binding sites located ~10.5 kb 5′ of the transcription start site. **(b)** Summary of maternal and gap protein gradients and segmentation enhancers. The diagrams on the left represent side views of early embryos. The maternal Bicoid (bcd) and Caudal (cad) proteins are distributed in opposing gradients, with Bicoid expressed in anterior regions. The gap repressor Hunchback (hb) is distributed in anterior regions, whereas Knirps (kni) is present in the presumptive abdomen (middle diagram). Finally, the Kruppel (Kr) repressor is localized in central regions of the embryo. Some of the binding clusters that were identified within the *Drosophila* genome are shown on the right. These clusters are associated with three segmentation genes: *giant*, *odd-skipped* (odd), and *pdm2*. A DNA fragment containing the cluster of cad, Kr, and hb sites in the 5′ regulatory region of *giant* was shown to direct a band of expression in posterior regions.

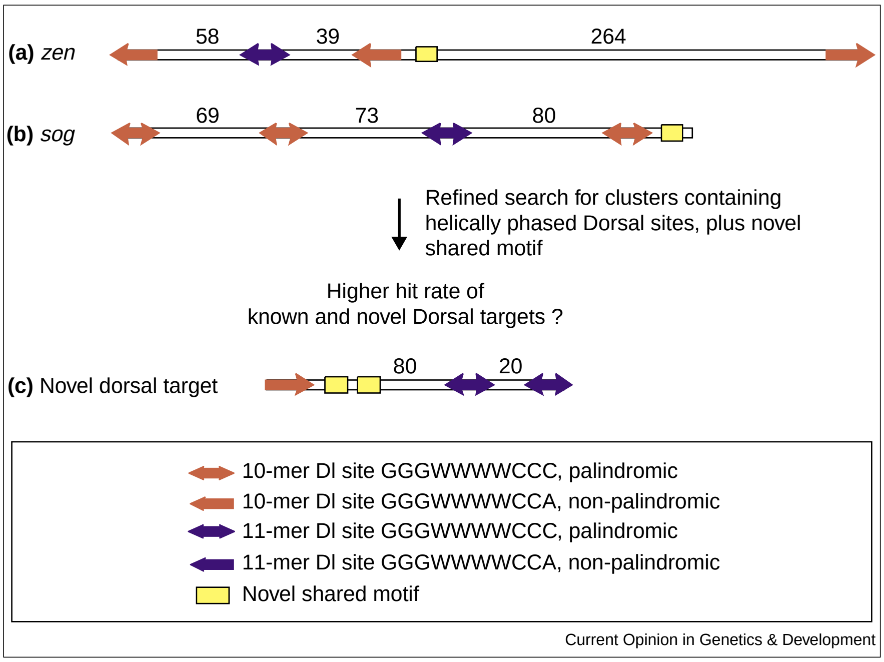
12 of the 46 clusters, ~25% correspond to Notch target enhancers. This ‘hit rate’ is similar to that described for Dorsal-binding clusters.

Computing segmentation enhancers

The preceding analyses document the efficacy of exploiting binding site matrices of single transcription factors to identify authentic *cis*-regulatory DNAs. Approximately one-fourth to one-third of the clusters identified by Dorsal or Su(H)-binding sites might correspond to either enhancers or silencers. However, as we discussed earlier, a typical enhancer is regulated by multiple factors [6]. It is reasonable to anticipate that the analysis of multiple factors should lead to a higher ‘hit rate’. That is, perhaps most of the clusters for multiple factors engaged in a common process correspond to authentic *cis*-regulatory DNAs. A seemingly ideal test case is provided by the segmentation process in the *Drosophila* embryo.

Opposing gradients of two maternal homeodomain proteins, Bicoid and Caudal, lead to the localized expression of several gap genes, which encode zinc finger repressors, including Hunchback, Krüppel, and Knirps (e.g. see [28]). Segmentation stripes of gene expression, such as the localized expression of *eve stripe 2*, are established through the interplay of the maternal Bicoid and Caudal activators, and the gap repressors (Figure 1b). The *eve stripe 2* enhancer contains five binding sites for the Bicoid activator, as well as six binding sites for gap repressors [7]. Berman *et al.* [12•] created position weight matrices for the binding sites of Bicoid, Caudal, Hunchback, Krüppel, and Knirps using a broad spectrum of binding sites compiled from 19 segmentation enhancers previously described in the literature. These matrices were used to identify clusters of Bicoid, Caudal, and gap-binding sites throughout the non-coding portion of the fly genome. A search for clusters containing 15 sites within 700 bp identified half of the 19 enhancers upon which the search was based and an additional 22 novel clusters. These clusters were pooled with 6 other novel clusters identified by an equally stringent screen for clusters containing 4 rather than 5 of the transcription factors. Of these pooled 28 novel clusters ~10 are associated with genes showing localized expression along the anterior–posterior axis, such as *odd* and *pdm2* (Figure 1b). One of the 10 clusters was shown to mediate the posterior expression pattern of the segmentation gene, *giant* [12•,29,30]. So, something like one-third of the novel clusters might correspond to authentic segmentation enhancers. This value is similar to the hit-rate observed for individual Dorsal- and Su(H)-binding site clusters. Bicoid, Caudal, and the gap proteins

Figure 2



Additional features might provide more refined bioinformatics searches. The diagrams in **(a,b)** summarize the distribution of Dorsal-binding sites in the *zen* silencer sequence and the *sog* intronic enhancer. Just a subset of the optimal *Dorsal* recognition sequences are present in the *sog* enhancer. Every site is palindromic and three of the four sites contain four rather than five central W residues. Neighboring sites are separated by similar distances. It is conceivable, but not known, that these special features of the *sog* enhancer are important for its function, and might provide a foundation for more sophisticated computational searches. In addition, it should be possible to compare coordinately regulated *cis*-regulatory DNAs, such as the *zen* silencer and *sog* enhancer, to identify additional conserved sequence motifs. A hypothetical motif is indicated by the yellow box in the diagrams. All of this information, the phasing and type of Dorsal-binding sites, as well as additional conserved motifs might permit a higher hit-rate in the search for novel Dorsal target *cis*-regulatory DNAs (summarized in **[c]**).

bind DNA as monomers and recognize degenerate sequences with low-binding affinity (e.g. [31]). Perhaps the clustering of multiple monomeric factors provides no more specificity than the clustering of a single dimeric protein such as Dorsal. One possible way to improve studies based on multiple sequences with low specificity (or information content) may be to use Boolean operators (i.e. 'and', 'or', 'not') to require specific combinations of sites, thereby increasing stringency.

Computing heart enhancers

An effort to use specific combinations of multiple binding sites (using the Boolean operator 'and') was recently reported by Halfon *et al.* [13•]. These authors focused their analysis on the well-characterized heart enhancer from the 3′ regulatory region of the *eve* gene [32–34]. This enhancer is regulated by a variety of transcription factors, including dTcf, Mad, and Pointed, which are mediators of *Wingless*, *Transforming growth factor-β* (*TGF-β*), and *Sevenless* (*Sev*) receptor tyrosine kinase signaling pathways, respectively [11•]. The *eve* heart enhancer also contains binding sites for the mesoderm determinant, Twist, as well as Tinman, a homeodomain transcription factor essential for heart differentiation in flies and mice [35]. The *Drosophila* genome was examined for clusters containing at least two instances of each of the binding sites — Mad, Pointed, Twist, and Tinman — and one instance of the Tcf site, in an effort to identify additional heart-specific enhancers. A total of 33 novel regions were identified. One is located within the first intron of the *heartbroken* (*hbr*) gene, which is specifically expressed in the heart [36]. A DNA fragment containing the clustered sites was shown to direct heart-restricted expression in transgenic embryos. Two of the remaining clusters appear to be associated with heart-specific genes. However, when examined by transgenic analysis, these clusters did not recapitulate the endogenous profiles of the associated genes. This lower than expected hit rate may be due to the limited information content of some of the binding site matrices used in the study (e.g. Mad-binding sites). This suggests that the use of Boolean operators may not be able to overcome the noise created by poorly defined sites and highlights the need for well-defined, information-rich binding site matrices.

More motifs, higher hit rate?

In all of the cases we have discussed, dorsal–ventral patterning, Notch signaling, segmentation, and heart morphogenesis, the bioinformatics methods led to both false negatives and false positives. For example, many of the genes that are known to be regulated by the Dorsal gradient were not identified on the basis of optimal binding clusters because they are regulated by low-affinity recognition sequences that possess extensive degeneracy [10•]. A major goal of future efforts will be to increase the hit-rate and eliminate false negatives. Toward this end we anticipate the need to develop computational methods to identify shared sequence motifs among coordinately regulated enhancers. The use of such motifs has the potential to refine subsequent computational searches for new enhancers that mediate related patterns of gene expression.

Whereas several wet-lab methods are available for identifying coordinately expressed genes and enhancers in (e.g. cDNA microarray assays, chromatin immunoprecipitation [ChIP], automated large-scale *in situ* hybridizations), relatively few attempts have been made to identify and test novel motifs shared among co-expressed genes in multicellular animals [37•,38]. A recent study comparing *Drosophila* segmentation enhancers for shared motifs accurately identified both known and novel motifs. The known motifs correspond to binding sites for matenal and gap transcription factors [39]. The novel shared motifs may therefore be relevant to enhancer function but functional tests for these motifs have not yet been reported. The use of such motifs has the potential to refine subsequent computational searches for new enhancers that mediate related patterns of gene expression. Some of the computational methods that can identify shared motifs are available as web-based tools at the URLs listed in Box 1.

An initial attempt to test computational predictions of novel binding sites was reported by Halfron *et al.* [13•]. The *eve* enhancer was compared with the novel clusters identified as potential heart enhancers in the *Drosophila* genome. A conserved sequence motif was identified that is located in both the *eve* and *Hbr* heart enhancers. This motif is related to the binding site for the Oct-1 transcription factor. Mutations in this site cause an otherwise normal *eve* heart enhancer to direct an expanded pattern of expression. Whole-genome searches that include this motif, along with more stringently defined Tcf, Mad, Pointed, Twist, and Tin binding sites should lead to a higher hit-rate for new heart enhancers.

Conclusions

The preceding studies demonstrate that bioinformatics methods can be used to identify novel enhancers. This will forever change the way that *cis*-regulatory DNAs are characterized in complex metazoan genomes. However, these studies have not yet revealed a '*cis*-regulatory code', whereby gene-expression patterns can be inferred from simple sequence analysis. Better predictions may be achieved through the identification of additional conserved features of *cis*-regulatory DNAs, beyond the simple clustering of binding sites. For example, the interferon 'enhanceosome' contains binding sites that are separated by fixed distances, which facilitate protein–protein interactions [40]. The *sog* intronic enhancer contains four evenly spaced Dorsal binding sites that are separated by distances of ~80 bp — about one turn of the nucleosome (Figure 2). Helical phasing, nucleosome phasing, and the stereochemistry of binding, provide a 'grammar' that should increase the hit-rate in future computational searches for *cis*-regulatory DNAs.

Acknowledgements

We thank Yutaka Nibu for help with the figures and Kate Senger for helpful discussions. This work was funded by a grant from the NIH (GM46638).

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF *et al.*: The genome sequence of ***Drosophila melanogaster***. *Science* 2000, 287:2185-2195.
2. Waterston R, Sulston J: The genome of ***Caenorhabditis elegans.*** *Proc Natl Acad Sci USA* 1995, 92:10836-10840.
3. Banerji J, Rusconi S, Schaffner W: Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 1981, 27:299-308.
4. Brand AH, Breeden L, Abraham J, Sternglanz R, Nasmyth K: Characterization of a "silencer" in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell* 1985, 41:41-48.
5. Bell AC, West AG, Felsenfeld G: Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* 2001, 291:447-450.
6. Davidson EH: *Genomic Regulatory Systems: Development and Evolution*. San Diego: Academic Press; 2001.
7. Small S, Blair A, Levine M: Regulation of even-skipped stripe 2 in the ***Drosophila*** embryo. *EMBO J* 1992, 11:4047-4057.
8. Frith M, Hansen U, Weng Z: Detection of cis-element clusters in higher eukaryotic DNA. *Bioinformatics* 2001, 17: 878-889.
9. • Pennachhio LA, Rubin EM: Genomic strategies to identify mammalian regulatory sequences. *Nat Rev Genet* 2001, 2:100-109.

Review of classic wet-lab and genomic approaches (including phylogenetic footprinting, position weight matrices and inter-species comparisons) to identify *cis*-regulatory DNAs in mammals.

10. • Markstein M, Markstein P, Markstein V, Levine M: Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the Drosophila embryo. *Proc Natl Acad Sci USA* 2002, 99:763-768.

This study examines clusters of Dorsal binding sites in the *Drosophila* genome and reports the identification of a *sog* intronic enhancer using the programme 'Fly Enhancer', available at http://flyenhancer.org

11. • Rebeiz M, Reeves NL, Posakony JW: SCORE: a computational approach to the identification of cis-regulatory modules and target genes in whole-genome sequence data. *Proc Natl Acad Sci USA* 2002, 99:9888-9893.

Binding clusters for the Su(H) protein are shown to be associated with Notch target genes in the *Drosophila* genome. This study uses the 'SCORE' programme which detects clusters of binding sites and selects those which are most statistically significant, based on Monte Carlo simulations. SCORE is not currently available online, but Perl scripts are available upon request to the authors.

12. • Berman BP, Nibu Y, Pfeiffer BD, Tomancak P, Celniker SE, Levine M, Rubin GM, Eisen MB: Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the ***Drosophila*** genome. *Proc Natl Acad Sci USA* 2002, 99:757-762.

Clusters of multiple maternal and gap proteins are shown to be associated with segmentation genes in the *Drosophila* genome. One such cluster corresponds to an authentic enhancer in the giant 5′ regulatory region. This study used the programme '*cis*-analyst' to identify and plot the clusters, and it is available on the web at http://www.fruitfly.org/cis-analyst

13. • Halfon MS, Grad Y, Church GM, Michelson AM: Computation-based discovery of related transcriptional regulatory modules and motifs using an experimentally validated combinatorial model. *Genome Res* 2002, 12:1019-1028.

The *eve* heart enhancer is used as a model to identify an intronic enhancer in the *Hbr* gene. Comparison of the *eve* and *Hbr* enhancers identifies a new conserved motif that mediates repression within the *eve* heart enhancer. To find new motifs the study used the programme 'cooccur_scan.pl' available at http://arep.med.harvard.edu/Halfon_Grad_etal/supplemental.html

14. Stathopoulos A, Levine M: Dorsal gradient networks in the Drosophila embryo. *Dev Biol* 2002, 246:57-67.
15. Thisse C, Perrin-Schmitt F, Stoetzel C, Thisse B: Sequence-specific transactivation of the Drosophila twist gene by the dorsal gene product. *Cell* 1991, 65:1191-1201.
16. Jiang J, Kosman D, Ip YT, Levine M: The dorsal morphogen gradient regulates the mesoderm determinant twist in early Drosophila embryos. *Genes Dev* 1991, 5:1881-1891.
17. Ip YT, Park RE, Kosman D, Yazdanbakhsh K, Levine M: dorsal-twist interactions establish snail expression in the presumptive

mesoderm of the ***Drosophila*** embryo. *Genes Dev* 1992, 6:1518-1530.

18. Ip YT, Park RE, Kosman D, Bier E, Levine M: The dorsal gradient morphogen regulates stripes of rhomboid expression in the presumptive neuroectoderm of the ***Drosophila*** embryo. *Genes Dev* 1992, 6:1728-1739.

19. Kasai Y, Stahl S, Crews S: Specification of the Drosophila CNS midline cell lineage: direct control of single-minded transcription by dorsal/ventral patterning genes. *Gene Expr* 1998, 7:171-189.

20. Kirov N, Zhelnin L, Shah J, Rushlow C: Conversion of a silencer into an enhancer: evidence for a co-repressor in dorsal-mediated repression in ***Drosophila***. *EMBO J* 1993, 12:3193-3199.

21. Doyle HJ, Kraut R, Levine M: Spatial regulation of zerknullt: a dorsal-ventral patterning gene in Drosophila. *Genes Dev* 1989, 3:1518-1533.

22. Huang JD, Schwyter DH, Shirokawa JM, Courey AJ: The interplay between multiple enhancer and silencer elements defines the pattern of decapentaplegic expression. *Genes Dev* 1993, 7:694-704.

23. Ip YT, Kraut R, Levine M, Rushlow CA: The dorsal morphogen is a sequence-specific DNA-binding protein that interacts with a long-range repression element in Drosophila. *Cell* 1991, 64:439-446.

24. Francois V, Solloway M, O'Neill JW, Emery J, Bier E: Dorsal-ventral patterning of the Drosophila embryo depends on a putative negative growth factor encoded by the short gastrulation gene. *Genes Dev* 1994, 8:2602-2616.

25. Jazwinska A, Rushlow C, Roth S: The role of brinker in mediating the graded response to Dpp in early Drosophila embryos. *Development* 1999, 126:3323-3334.

26. Barolo S, Posakony JW: Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. *Genes Dev* 16:1167-1181.

27. Struhl G, Adachi A: Nuclear access and action of notch ***in vivo.*** *Cell* 1998, 93:649-660.

28. Wimmer EA, Carleton A, Harjes P, Turner T, Desplan C: Bicoid-independent formation of thoracic segments in Drosophila. *Science* 2001, 287:2476-2479.

29. Eldon ED, Pirrotta V: Interactions of the Drosophila gap gene giant with maternal and zygotic pattern-forming genes. *Development* 1991, 111:367-378.

30. Kraut R, Levine M: Spatial regulation of the gap gene giant during ***Drosophila*** development. *Development* 1991, 111:601-609.

31. Hoey T, Levine M: Divergent homeo box proteins recognize similar DNA sequences in ***Drosophila***. *Nature* 1988, 332:858-861.

32. Knirr S, Frasch M: Molecular integration of inductive and mesoderm-intrinsic inputs governs even-skipped enhancer activity in a subset of pericardial and dorsal muscle progenitors. *Dev Biol* 2001, 238:13-26.

33. Su MT, Fujioka M, Goto T, Bodmer R: The ***Drosophila*** homeobox genes ***zfh-1*** and ***even-skipped*** are required for cardiac-specific differentiation of a numb-dependent lineage decision. *Development* 1999, 126:3241-3251.

34. Halfon MS, Carmena A, Gisselbrecht S, Sackerson CM, Jimenez F, Baylies MK, Michelson AM: Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. *Cell* 2002, 103:63-74.

35. Frasch M: Intersecting signalling and transcriptional pathways in ***Drosophila*** heart specification. *Semin Cell Dev Biol* 1999, 10:61-71.

36. Michelson AM, Gisselbrecht S, Buff E, Skeath JB: Heartbroken is a specific downstream mediator of FGF receptor signalling in ***Drosophila***. *Development* 1998, 125:4379-4389.

37. Shannon MF, Rao S: Of Chips and ChIPs. *Science* 2002,
• 296:666-669.

Review of current advances in the dissection of transcriptional networks in yeast, worms, flies, and mice using cDNA microarrays, chromatin ChIP assays and bioinformatics.

38. Simin K, Scuderi A, Reamey J, Dunn D, Weiss R, Metherall J, Letsou A: Profiling patterned transcripts in ***Drosophila*** embryos. *Genome Res* 2002, 12:1040-1047.

39. Papatsenko DA, Makeev VJ, Lifanov AP, Regnier M, Nazina AG, Desplan C: Extraction of functional binding sites from unique regulatory regions: the ***Drosophila*** early developmental enhancers. *Genome Res* 2002, 12:470-481.

40. Merika M, Thanos D: Enhanceosomes. *Curr Opin Genet Dev* 2001, 11:205-208.