Cell Reports
Resource

# A Survey of 6,300 Genomic Fragments for *cis*-Regulatory Activity in the Imaginal Discs of *Drosophila melanogaster*

Aurélie Jory,[1] Carlos Estella,[1,3] Matt W. Giorgianni,[1,4] Matthew Slattery,[1,5] Todd R. Laverty,[2] Gerald M. Rubin,[2] and Richard S. Mann[1,*]

[1]Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, 701 W. 168th Street, HHSC 1104, New York, NY 10032, USA

[2]Janelia Farms Research Campus, 19700 Helix Drive, Ashburn, VA 20147, USA

[3]Present address: Departamento de Biología Molecular, and Centro de Biología Molecular "Severo Ochoa," Universidad Autónoma de Madrid, Madrid, Spain

[4]Present address: R.M. Bock Laboratories, University of Wisconsin-Madison, 1525 Linden Drive, Madison, WI 53706, USA

[5]Present address: Institute for Genomics and Systems Biology, University of Chicago, 900 E. 57th St. KCBD 10115, Chicago, IL 60637, USA

*Correspondence: rsm10@columbia.edu

http://dx.doi.org/10.1016/j.celrep.2012.09.010

## SUMMARY

**Over 6,000 fragments from the genome of *Drosophila melanogaster* were analyzed for their ability to drive expression of GAL4 reporter genes in the third-instar larval imaginal discs. About 1,200 reporter genes drove expression in the eye, antenna, leg, wing, haltere, or genital imaginal discs. The patterns ranged from large regions to individual cells. About 75% of the active fragments drove expression in multiple discs; 20% were expressed in ventral, but not dorsal, discs (legs, genital, and antenna), whereas ~23% were expressed in dorsal but not ventral discs (wing, haltere, and eye). Several patterns, for example, within the leg chordotonal organ, appeared a surprisingly large number of times. Unbiased searches for DNA sequence motifs suggest candidate transcription factors that may regulate enhancers with shared activities. Together, these expression patterns provide a valuable resource to the community and offer a broad overview of how transcriptional regulatory information is distributed in the *Drosophila* genome.**

## INTRODUCTION

The control of gene transcription in eukaryotes is governed by DNA sequences known as *cis*-regulatory modules (CRMs), or enhancers (Bonn and Furlong, 2008; Istrail and Davidson, 2005). By binding combinations of transcription factors, CRMs integrate spatial and temporal inputs to produce a wide variety of gene expression patterns. In higher eukaryotes, CRMs can be many kilobases away from the start site of transcription, often making their identification a challenge. Large genes with complex regulation typically contain multiple CRMs (Bulger and Groudine, 2010). In many cases, individual CRMs are responsible for driving transcription in specific subsets of a gene's expression pattern. In addition, CRMs from the same gene may have overlapping spatial and temporal activities. CRMs such as these with partially redundant activities may be important for ensuring robust gene expression patterns, especially when animals are in stressful environmental conditions (Barolo, 2012; Frankel et al., 2010; Perry et al., 2010).

Much of our current understanding of CRM activities is derived from so-called enhancer bashing experiments, where DNA sequences are tested for transcriptional regulatory activity by using them to drive the expression of transgenic reporter genes. Such reporter genes typically use heterologous minimal promoter sequences and test CRM activities out of their native contexts. Despite several thousand reporter genes already described in the literature, to date no large scale, unbiased studies have been carried out to determine the frequency of CRMs within a eukaryotic genome and to ask general questions about the organization of the regulatory genome. To carry out such a study, we characterized the transcriptional regulatory activities driven by a set of 6,300 fragments derived from the genome of *Drosophila melanogaster* in the imaginal discs, epithelial structures that give rise to most of the adult fly. Together with the accompanying papers (Jenett et al., 2012; Manning et al., 2012; this issue), which describe the expression patterns driven by these same fragments in the adult and embryonic central nervous system (CNS), these studies present a relatively unbiased view of how transcriptional regulatory information is distributed in the *Drosophila* genome and significantly expand the number of characterized CRMs in *Drosophila* (http://redfly.ccr.buffalo.edu/index.php) (Gallo et al., 2011). Moreover, these reporter genes are all inserted into the same genomic position and all drive the expression of the yeast transcription factor GAL4 (Brand and Perrimon, 1993; Pfeiffer et al., 2008). Thus, in addition to providing a broad overview of CRM activities, the results from this screen provide researchers with a valuable set of tools that can be used to ectopically express or knock down any factor of interest in a large number of distinct spatial and temporal patterns. The complete expression patterns of these reporter genes are presented in an online database (http://www.janelia.org/gal4-gen1).

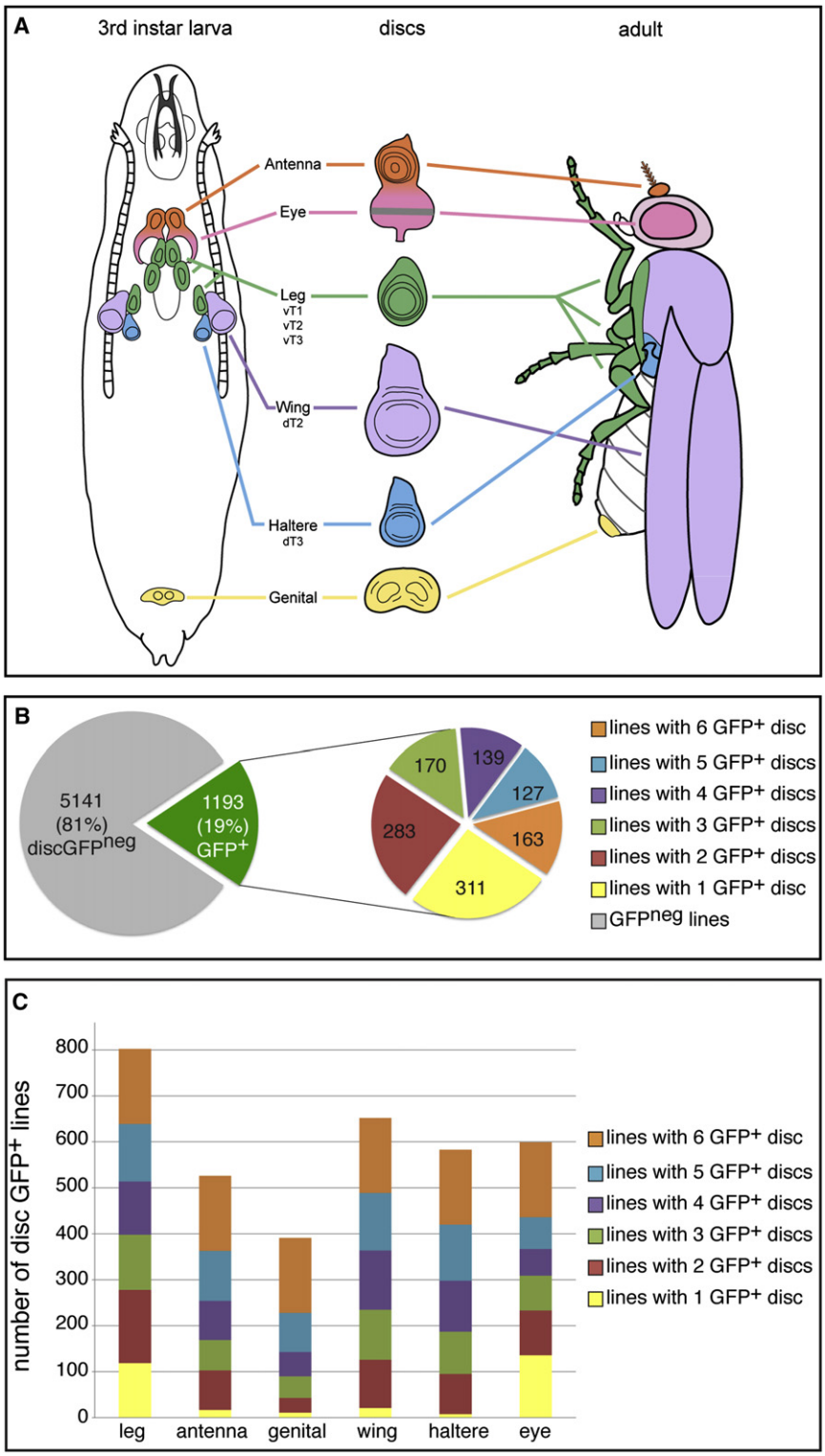


**Figure 1. Overall Distribution of Imaginal Disc GFP+ Reporter Genes**

(A) Schematics of the six types of larval imaginal discs that were screened for reporter gene expression.

(B) Nineteen percent of the 6,334 analyzed lines drove expression in one or multiple larval imaginal discs. Fifty percent of these disc GFP+ lines drove expression in one or two discs, whereas the other half drove expression in three or more discs.

(C) Very few lines drove expression only in the antennal, genital, wing, or haltere discs.

See also Figure S1.

## RESULTS AND DISCUSSION

To analyze the expression patterns driven by a large number of reporter genes at a comparable developmental stage, we examined six different third-instar (wandering stage) larval imaginal discs: three dorsal discs (eye, wing, and haltere) and three ventral discs (antenna, leg [one for each thoracic segment, T1 to T3], and genital). With the exception of the abdomen, these six imaginal discs give rise to most of the external structures of the adult fly (Figure 1A). In addition to the appendages, these discs also generate the main part of the body from which the appendages articulate, for example, the notum (derived from the dorsal region of the wing disc) and the ventral pleura (derived from the three leg discs; Mann and Morata, 2000; Morata, 2001) (Figure 1A). In addition to generating dorsal versus ventral structures of the adult fly, these six discs also differ with respect to their anterior-posterior (AP) coordinates. The most anterior disc in this set is the eye-antenna disc, which generates the head, eyes, and antennae; the first thoracic (T1) leg disc generates the T1 leg and pleura; the second thoracic (T2) segment of the adult is generated by the T2 leg (ventral) and wing (dorsal) discs; and the third thoracic (T3) segment of the adult segment is derived from the T3 leg (ventral) and haltere (dorsal) discs. Finally, the most posterior disc in this set is the genital disc, which gives rise to the analia and external reproductive structures of the adult. Consequently, genital discs are sexually dimorphic. In our characterization of reporter gene expression patterns, we analyzed both female and male genital discs, as well as all three leg discs (T1 to T3).

The genes chosen as the source of fragments for these reporter genes were suspected to have a role in nervous system development or function (Pfeiffer et al., 2008). For each gene, <1 to 5 kb genomic fragments that generally tiled the entire locus,

excluding exons, were cloned into GAL4-expressing reporter genes (Pfeiffer et al., 2008). Each reporter gene was inserted into a single genomic locus (*attP2* at 68A4) via phiC31-mediated transformation (Pfeiffer et al., 2008). We initially scanned the expression of nuclear GFP after each line was crossed to *UAS-nGFP*; larvae that showed evidence of expression in the discs were dissected and imaged for GFP and coimmunostained for at least two of the following three well-characterized disc markers, the transcription factors Distalless (Dll), Homothorax (Hth), or Engrailed (En). These markers are expressed in most of the imaginal discs at the late third-instar stage and helped with the characterization of the reporter gene patterns (Mann and Morata, 2000). Discs from at least five independent larvae were analyzed. For some experiments, DAPI was included to identify all nuclei.

## Gene Ontology Analysis

Altogether, we examined 6,334 reporter genes, generated from fragments from 892 genes. Of these 6,334 reporter genes about 19% (1,193) drove expression in one or more of the six imaginal discs that we analyzed (Figure 1B). The 1,193 lines that were positive were derived from 408 genes. Thus, although only 19% of the fragments were active in the discs, about 46% of the genes in this data set had at least one fragment that drove expression in the imaginal discs.

Relative to the 892 total genes in this data set, Gene Ontology (GO) analysis of the 408 positive genes revealed a statistically significant bias in favor of transcription factors and regulators of gene expression, as would be expected for regulators of development (Figure S1). Moreover, consistent with their expression in imaginal discs, several GO terms related to animal development, such as organ morphogenesis, organ development, and imaginal disc morphogenesis, were also significantly enriched in the positive gene set. In addition, GO terms related to nervous system function, including sensory organ and compound eye development, were also identified (Figure S1).

## Expression Biases between Imaginal Discs

In this section, we describe how the expression patterns driven by the 1,193 positive reporter genes were distributed among the various imaginal discs.

The first question we asked concerned the disc specificity of reporter gene expression. We found that only about 26% of the lines were expressed in only one type of imaginal disc (Figure 1B). Notably, the frequency of disc-specific patterns was highest for the leg and eye discs and was very low or absent for the other discs (Figure 1C). Most disc GFP+ reporter genes (~75%) drove expression in at least two types of discs, and ~160 lines were expressed in all six imaginal discs (Figure 1B). These data suggest that most CRMs are not specific for individual disc types, consistent with the idea that different imaginal discs rely on many of the same signaling pathways and transcription factors for their development (Mann and Morata, 2000; Morata, 2001).

We next asked if the reporter genes had a tendency to be expressed in dorsal (wing, eye, and haltere) but not ventral (leg, antenna, and genital) discs or vice versa. When all GFP+ disc lines were considered, regardless of whether they were expressed in dorsal, ventral, or both types of discs, we found that there were a similar number of lines distributed among the ventral and dorsal discs (Figures 2A and 2B). For example, the number of lines expressed in leg discs but not in other ventral discs was 267, very similar to the number of lines (270) expressed in the eye disc but not other dorsal discs. Similarly, the number of lines expressed in the antenna disc but not in other ventral discs was 60, which is the same number of lines expressed in the wing disc but not other dorsal discs. These numbers reveal a remarkable amount of symmetry in the dorsal-ventral breakdown of reporter gene patterns.

This symmetry extends to ventral- and dorsal-specific expression patterns. There were 244 lines that were ventral-disc-specific (expressed in at least one ventral disc but no dorsal disc) and 277 lines that were dorsal-disc-specific (expressed in at least one dorsal disc but no ventral disc) (Figure 2C). Looking at the further breakdown of these two groups, nearly half of the "ventral, not dorsal" set were comprised of lines expressed only in the leg discs (119/244), whereas about half of the "dorsal, not ventral" set were comprised of lines expressed only in the eye disc (136/277) (Figures 2D and 2E). This bias in favor of leg and eye patterns was also seen among fragments that were expressed in at least one ventral and one dorsal disc (672 total; Figure 2C): 148/672 were expressed in the leg but not other ventral discs, and 134/672 were expressed in the eye but not other dorsal discs (Figures 2F and 2G). Together, these comparisons suggest that expression patterns in ventral or dorsal discs are equally represented within this data set. In addition, they reveal that leg and eye discs expressed the largest number of positive fragments, perhaps reflecting more complex transcriptional networks in these discs. It is also possible that the initial choice of neural-related genes as the starting point for this screen (Pfeiffer et al., 2008) may contribute to these biases, especially for the neuron-rich eye imaginal disc.

Another type of interdisc pattern that we monitored suggests regulation by the Hox genes, which establish morphological differences along the AP axis of the embryo and adult (Mann and Morata, 2000). The wing (T2) and haltere (T3), for example, are homologous appendages with distinct morphologies that differ only because of the activity of the Hox gene *Ultrabithorax* (*Ubx*) in the haltere and the absence of *Ubx* activity in the wing (Lewis, 1978). In the set of CRMs analyzed here, we found 557 lines expressed in a similar pattern in both the wing and haltere, 26 lines expressed in the haltere but not the wing and 96 lines expressed in the wing but not the haltere (Figure S2). Hox genes also establish differences between the pairs of legs in the different thoracic segments (Struhl, 1982). In the set of CRMs analyzed here, we found 783 lines with similar expression patterns in the leg discs from all three segments. In contrast, there were only 19 lines that were expressed in different patterns in the leg discs from T1, T2, or T3 (Figure S2). These numbers indicate that the number of Hox-regulated CRMs is a relatively small fraction of the total CRMs active in the discs. However, because these numbers only take into consideration on versus off regulation, they do not include CRMs whose activities or patterns are subtly modulated by Hox genes. Such modulatory functions are critical for how *Ubx* reduces the size of the haltere

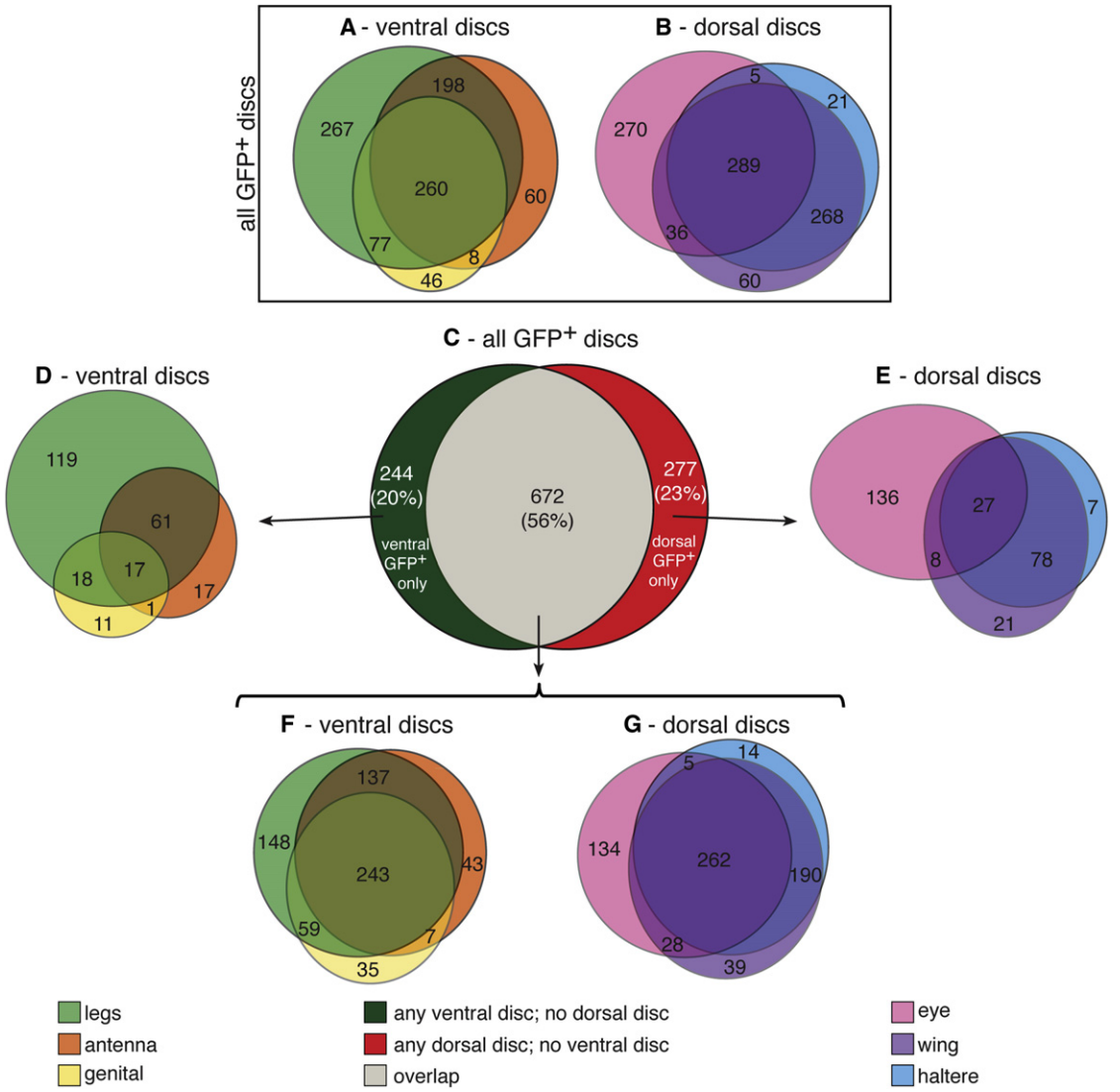


**Figure 2. Dorsal-Ventral Segregation of Imaginal Disc Patterns**

(A and B) Distribution of patterns among ventral (A) and dorsal (B) discs for all disc GFP+ lines.

(C–E) Forty-three percent of the disc GFP+ lines drove expression exclusively in ventral discs (20%; C and D) or dorsal discs (23%; C and E). Fifty-six percent of the lines had expression in at least one ventral and one dorsal disc.

(F and G) Distribution of patterns among ventral (F) and dorsal (G) discs for the 672 lines expressed in at least one ventral and one dorsal disc.

See also Figure S2.

disc relative to the wing disc (Crickmore and Mann, 2006, 2007, 2008; de Navas et al., 2006).

The somatic sex determination pathway establishes the morphological differences between the male and female genitalia by determining which splice isoform of the Zn finger transcription factor Doublesex ($Dsx^M$ or $Dsx^F$) is expressed (Estrada et al., 2003). We found 330 reporter gene patterns in common between the male and female genital discs. In contrast, there were only 52 lines expressed in the male, but not female, genital disc, and only nine lines expressed in the female, but not male, genital disc (Figure S2). Thus, as with Hox regulation, only a small subset of CRMs appear to be regulated in an on versus off manner in the male and female genital discs.

## Expression Biases within Imaginal Discs

We next examined the frequency of expression biases along the two main body axes (anterior-posterior [AP] and dorsal-ventral [DV]) within specific imaginal disc types. For the leg, antennae, wing, and haltere discs the percentage of lines that had such biases ranged from about ~7% to 13%, depending on the disc (Figure 3A). However, the type of bias differed depending on disc type: leg and antenna discs were more likely to have a bias along the DV axis, whereas haltere and wing discs were

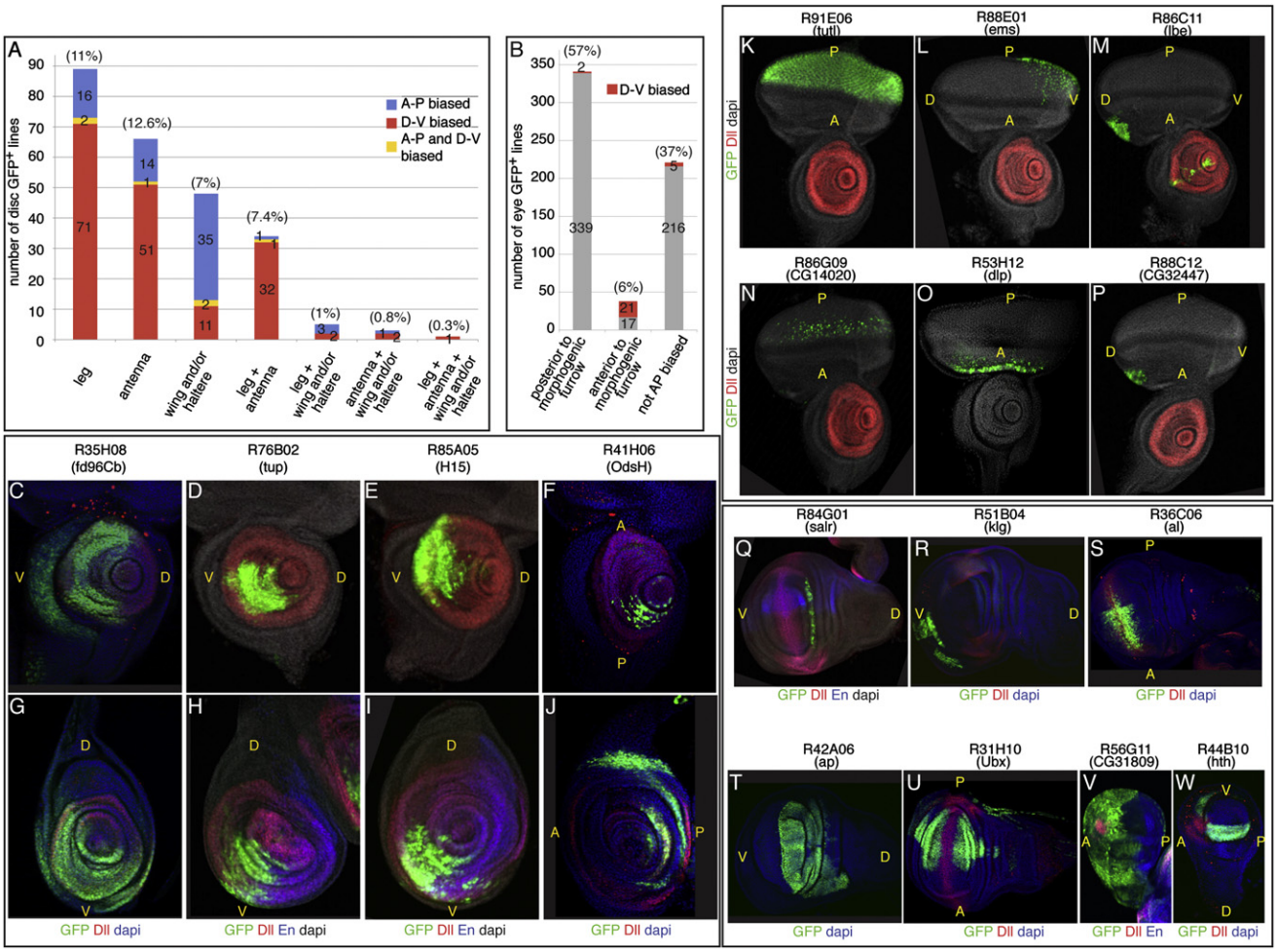


**Figure 3. Dorsal-Ventral and Anterior-Posterior Expression Biases within Discs**

(A) The percent above each bar represents the percent of lines showing biased expression, and the colors within each bar show the breakdown of these biases. The disc GFP+ lines that drove expression with a DV or AP bias in the leg and/or antenna discs preferentially exhibit a DV bias rather than an AP bias. In contrast, the lines that drove a biased expression pattern in the wing and/or haltere discs were more likely to have an AP bias.

(B) Fifty-seven percent of the lines that drove expression in the eye disc exhibit a biased expression posterior to the MF (examples in K, L, and N), whereas only 6% of the eye disc GFP+ lines drove GFP expression anterior to the MF (examples in M, O, and P). Most (21/38) of the lines that drove expression anterior to the MF were also biased along the DV axis, and many of these were expressed within the presumptive ocelli region of the eye disc (examples in M and P).

(C–J) Examples of antenna discs (C–F) and leg discs (G–J) from the same lines with biased expression patterns. Line and gene names are written above the panels.

(K–P) Examples of eye discs with biased expression patterns. Line and gene names are written above the panels.

(Q–W) Examples of wing or haltere discs with biased expression patterns. Line and gene names are written above the panels.

A, anterior; P, posterior; D, dorsal; V, ventral.

more likely to have a bias along the AP axis (Figure 3A). We also noticed that when a pattern was biased in a dorsal disc, it was very unlikely to be biased in a ventral disc: for example, only five reporter genes had biased expression along the same axis in both the leg and wing/haltere discs (Figure 3A).

At the third-instar larval stage, eye discs are unusual in that cells along the AP axis are at different developmental stages because of signals from the morphogenetic furrow (MF), which migrates in the disc from posterior to anterior during larval development (Kumar, 2011; Wolff and Ready, 1991). Posterior to the MF cells no longer divide and begin to differentiate into ommatidia, which include photoreceptors, cone, and pigment cells, whereas anterior to the MF the cells remain undifferentiated. Consistent with the neural bias for this screen, more than half (57%) of the reporter genes active in the eye disc were expressed specifically posterior, but not anterior, to the MF. In contrast, only 6% of lines expressed in the eye disc showed the opposite bias (expressed anterior to the MF, off posterior to the MF; Figure 3B). Interestingly, a very high percentage of the anterior, not posterior patterns (21/38), also had a DV bias (many expressed in the presumptive ocelli region (anterior/dorsal) of the eye disc; e.g., see Figures 3M and 3P), whereas

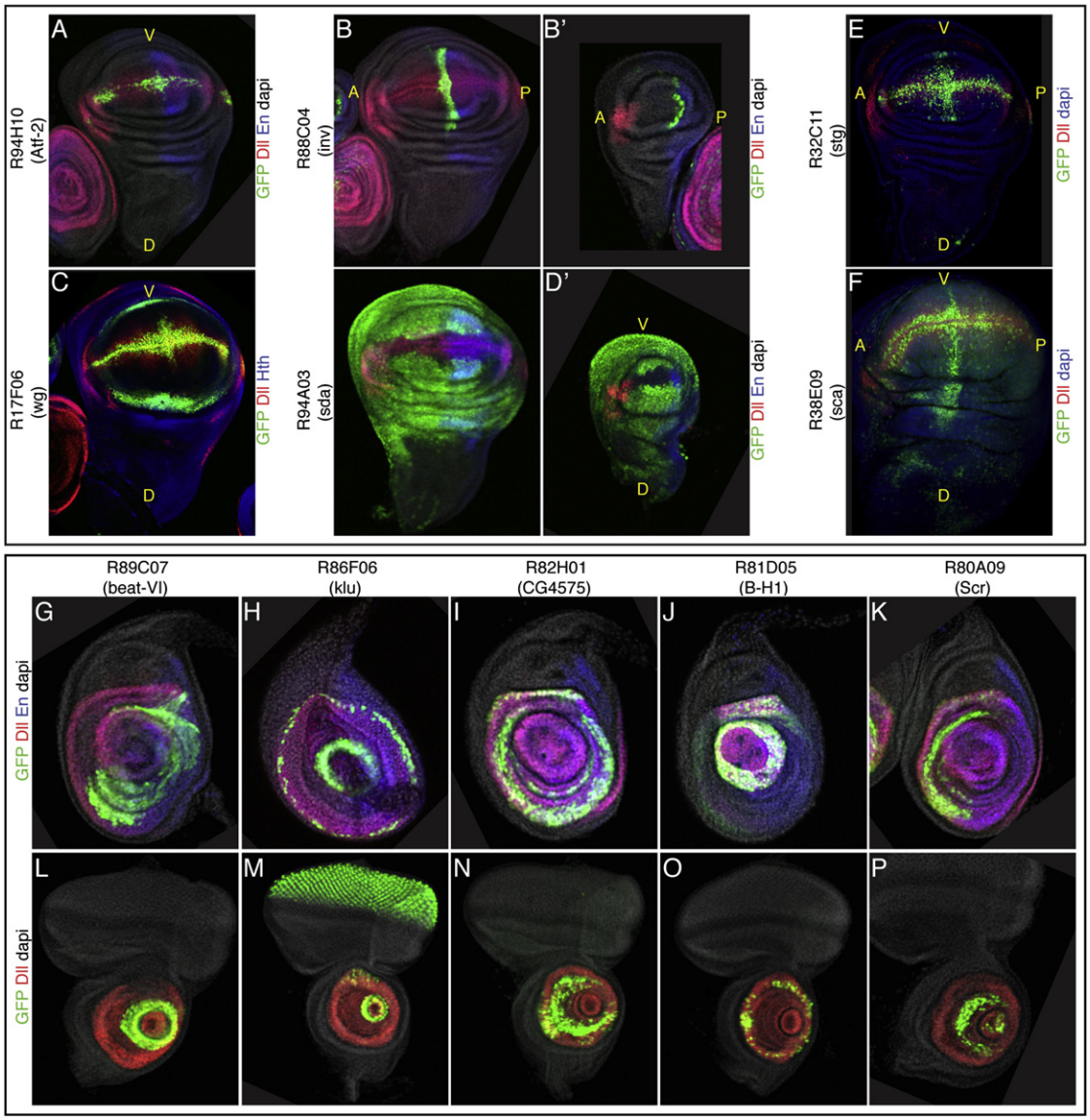


**Figure 4. Stripes, Antistripes, and Rings**

(A–D) Examples of lines that drove expression in a striped pattern at the DV or AP boundary of the wing and/or haltere discs. Several lines had an "antistripe" pattern at the DV boundary of the wing (D) and haltere (D') discs. Line and gene names are written to the left of the panels.

(E and F) Examples of two CRMs (from *scabrous* and *string*) that drove patterns that respect both the DV and AP boundaries. Line and gene names are written to the left of the panels.

(G–P) Examples of lines that drove expression in PD rings or partial rings in the leg (G–K) and antenna (L–P) discs. Line and gene names are written above the panels.

only 2/341 posterior, not anterior patterns, had a DV bias (Figure 3B). In summary, within this set of CRMs, the majority of eye disc patterns are in differentiated cells posterior to the MF. Although most of these CRMs drive expression in a repeated ommatidia-like pattern, we did not costain with markers to distinguish between neural and nonneural cell types.

Figures 3C–3W show examples of these types of biased patterns for ventral (antenna and leg) discs (Figures 3C–3J), the eye disc (Figures 3K–3P), and wing or haltere discs (Figures 3Q–3W).

## Jackpot Patterns

In this section, we highlight a small number of expression patterns that appeared at a surprisingly high frequency within our data set.

By the third-instar stage, the wing and haltere discs have well-defined DV and AP compartment boundaries (Blair, 2003; Mann and Morata, 2000). In our set of CRMs, we found many patterns that respect one of these boundaries. In the wing and/or haltere, there were 17 lines expressed along the DV boundary (Figures 4A and 4C). Interestingly, 11 lines had an inverse pattern: they

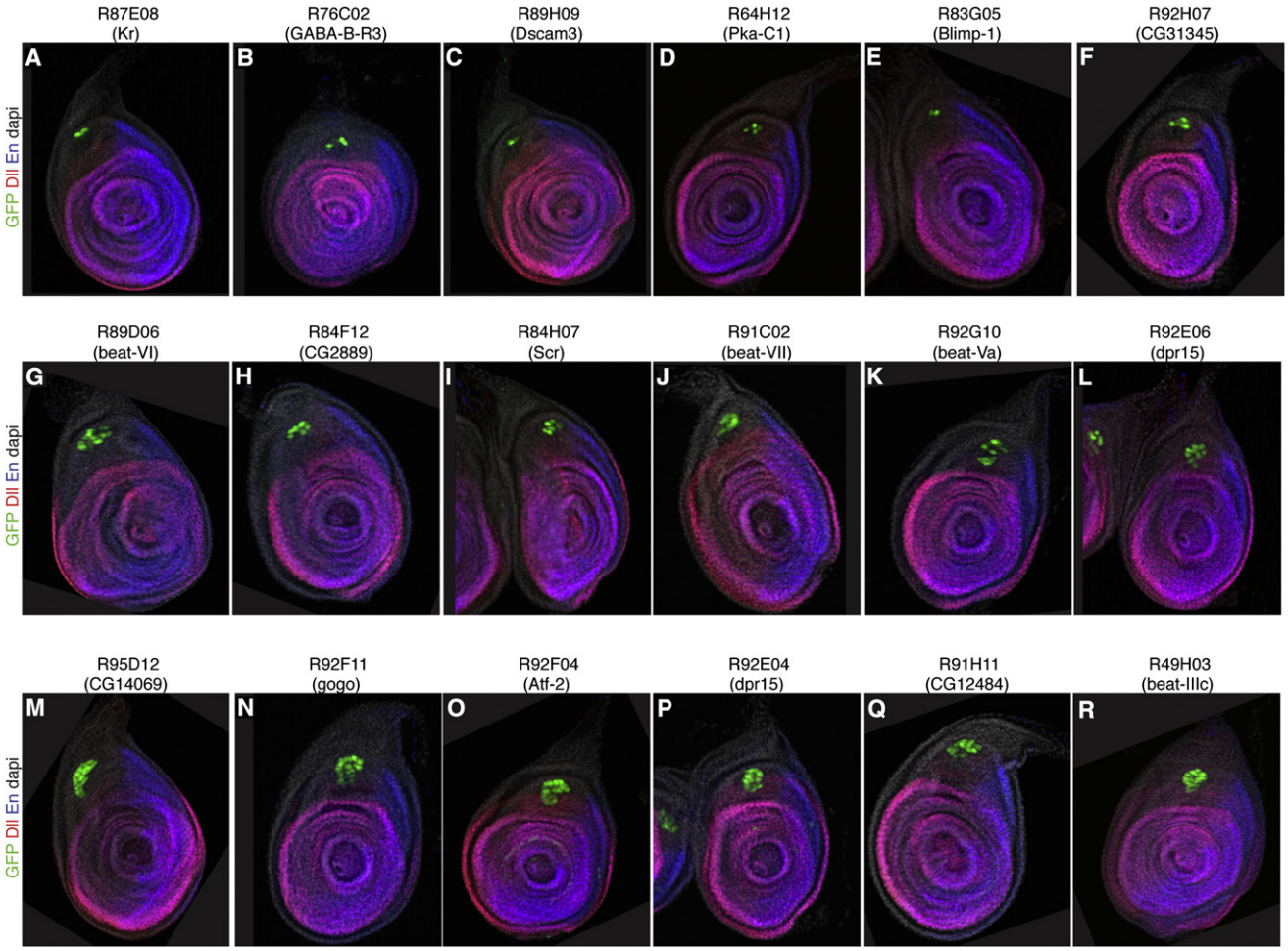


**Figure 5. Lines Expressed in the Femoral Chordotonal Organ of the Leg**

About 200 lines drove expression in a dorsal spot of the leg where the femoral chordotonal organ (FeCO) cells are located. Forty of these lines exclusively drove expression in the FeCO region of the leg disc. Line and gene names are written above the panels. Some lines drove expression in less than seven FeCO cells (A–F), others in 8–14 FeCO cells (G–L); and others in most of the ~30 FeCO cells (M–R).

were broadly expressed in the wing disc but were downregulated along the DV boundary (a so-called antistripe pattern; e.g., see Figure 4D). Interestingly, the breakdown of wing and/or haltere lines that respected the AP boundary was different: 15 lines were expressed in a stripe at or close to the AP boundary, but there were no lines that had an antistripe pattern at the AP boundary (Figure 4B). Notably, there were two lines (fragments from *string* and *scabrous*) that respected both the AP and DV boundaries (Figures 4E and 4F).

The proximal-distal (PD) axis of the legs and antennae is prefigured by concentric ring or circular expression domains in the imaginal discs (Estella et al., 2012; Rauskolb and Irvine, 1999). In our screen, we found 98 CRMs that show ring-like patterns in either the leg or antenna discs (Figures 4G–4P), suggesting that they receive inputs that establish the PD axis. Twenty of these 98 lines were not expressed elsewhere in the imaginal discs.

Finally, we highlight two frequently observed patterns in the leg discs that are likely to be in sensory neurons. One of these appears to be within cells of the femoral chordotonal organ, a complex sensory structure that reports the angle of the joint between the femur and tibia segments in the adult leg (Field and Matheson, 1998). Altogether, 196 lines drove expression in presumptive chordotonal organ cells and, of these, 40 lines had no other expression in the imaginal discs (Figure 5). Interestingly, the number of positive cells ranged from <10 to ~30, suggesting that there is a significant amount of cell-type specificity within this group of cells.

About 70 lines were expressed in one to four cells close to the center of the leg imaginal disc (27 of these drove no other expression in the discs). Based on the shapes of their cell bodies and association with the leg nerve, we suggest that these are the pioneer neurons that initially establish the axon pathway from the distal tip of the leg disc to the ventral nerve cord (VNC) (Jan et al., 1985) (Figure 6). Although present in the distal-most domain of the leg disc, these cells do not express Dll (Figure 6).

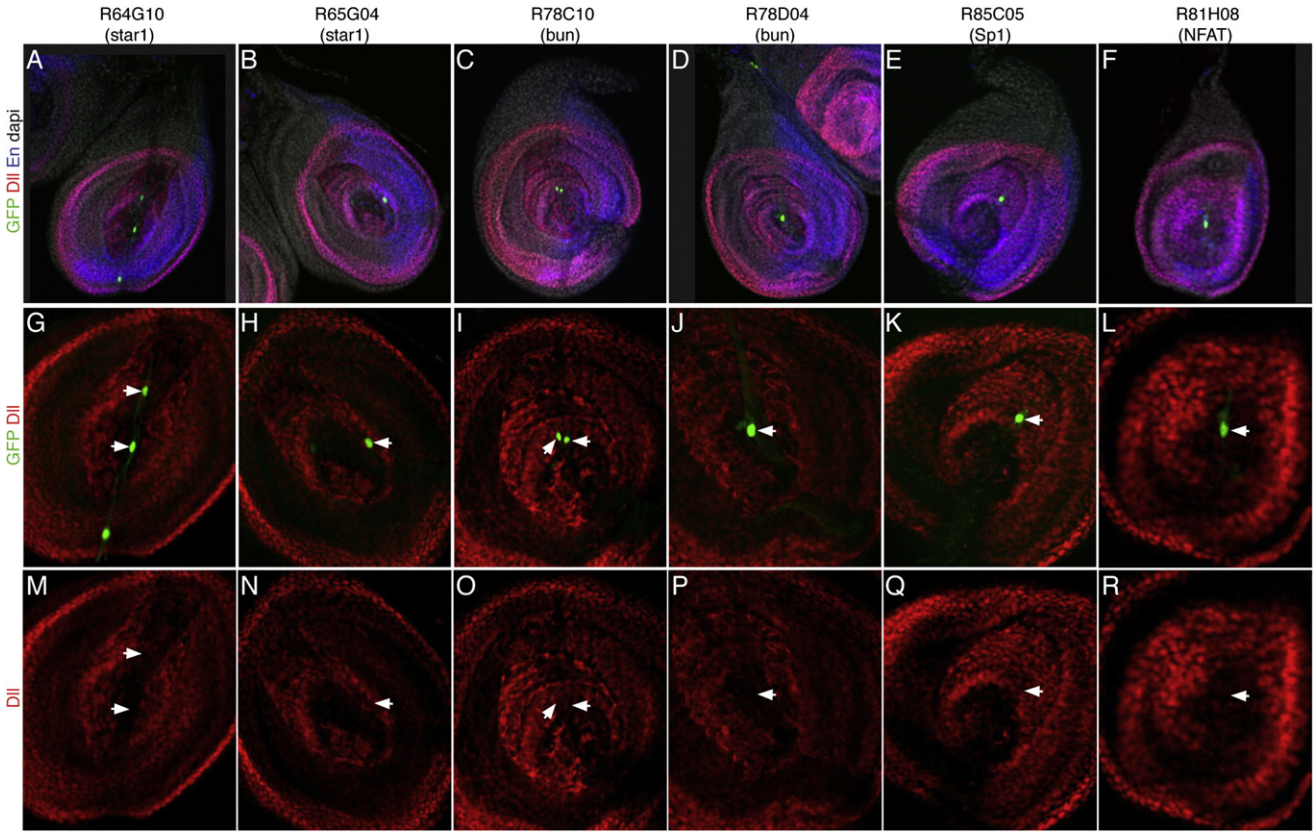


**Figure 6. Lines Expressed in Putative Leg Pioneer Sensory Neurons**

Example of lines driving GFP expression in one to three cells located on the nerve tracks of the leg discs (A–F) (G–R show higher magnification views of both the GFP and Dll channels). These cells do not express Dll (G–R).

### De Novo Discovery of Candidate Binding Sites

CRMs function by binding combinations of transcription factors. Previous attempts to discover transcription factor binding sites in sets of DNA fragments have largely focused on sequences identified by chromatin immunoprecipitation (ChIP) experiments (Bailey and Machanick, 2012). Having parsed the imaginal disc CRMs into groups that share similar expression patterns, we asked to what extent can we use these sets of DNA sequences to discover putative binding sites for transcription factors that may regulate these CRMs. We carried out motif discovery on four sets of CRMs: the putative leg chordotonal organ lines (40 CRMs), putative pioneer sensory neurons in the distal leg disc (27 CRMs), antenna and/or leg disc ring patterns (20 CRMs), and lines expressed in the putative ocelli (dorsal-anterior) region of the eye imaginal disc (25 CRMs). In all cases we discovered statistically significant patterns, some of which matched binding sites for known transcription factors (Figure 7). The identification of putative binding sites for these factors is consistent with previous work on these cell types. For example, binding sites for Sp1, a ventral selector protein (Estella and Mann, 2010), were discovered in the leg sensory cells, binding sites for Nubbin (Nub), which is known to be expressed in leg segmentation (Rauskolb and Irvine, 1999), were discovered in our leg/antennal ring set, and binding sites for the Six family protein Sine oculis (So) and Sister of odd and bowl (Sob), both known to play a critical role in eye or ocelli development (Bras-Pereira et al., 2006; Brockmann et al., 2011), were discovered in the ocelli CRM set (Figure 7). Occasionally, groups of CRMs representing broader categories were also enriched for biologically relevant motifs. For example, binding sites for Sp1 were enriched in discs with ventral-specific expression relative to discs with dorsal-specific expression (data not shown). Together, this analysis suggests that a significant fraction of CRMs within each of these groups may be coregulated by some of the same transcription factors. Moreover, these results set the stage for future molecular dissections of these CRMs by analyzing a potential role for these putative binding sites in CRM activity.

## CONCLUSIONS

In summary, this analysis provides an overview of the transcriptional regulatory landscape that exists in the *Drosophila* genome. The GAL4 lines and CRMs described here will provide both valuable tools for many future experiments and provide the starting materials for asking more mechanistic questions about how each of these patterns is generated during development. For example, characterized CRMs can be used to drive the

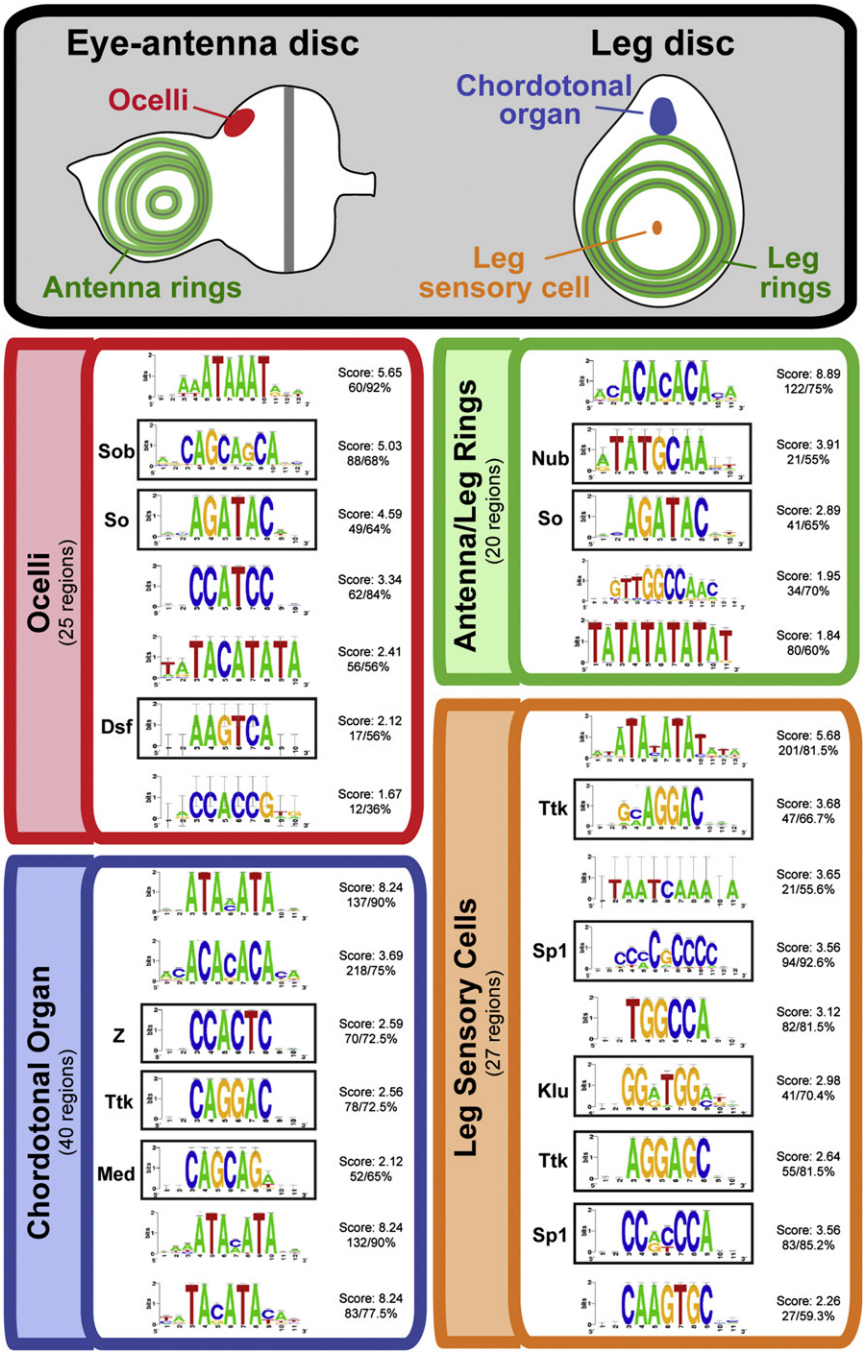


**Figure 7. DNA Motif and Putative Binding Site Discovery**

Significant motifs identified using RSAT peak-motifs (Thomas-Chollier et al., 2012). Annotated regions of enhancer expression are indicated in the eye-antenna and leg schematics in the top panel (red, ocelli; blue, chordotonal organ; green, antenna/leg rings; orange, leg sensory cell). Motif logos represent position-specific scoring matrices (see Experimental Procedures). Motifs in the red box were identified in DNA regions driving ocelli-specific expression. Motifs in the blue box were identified in regions directing expression in the leg chordotonal organ. The green and orange boxes contain motifs identified in regions driving expression in antenna/leg rings or leg sensory cells, respectively. For each enhancer expression set, all motifs with an E-value <0.01 are represented; the score for each motif is the $-\log_{10}$-transformed E-score. The numbers below the score for each motif represent the number of times a given motif is found within the DNA regions (left) and the percentage of regions containing at least one instance of the motif (right).

expression of site-specific recombinases and nonfly transcription factors (e.g., GAL4, LexA) to ectopically express or knock down any gene of interest in a small number of cells (del Valle Rodríguez et al., 2012; Nern et al., 2011; Pfeiffer et al., 2010). In the future, it will also be of interest to compare the location of active CRMs identified in this screen with other marks of gene expression, such as histone modifications, chromatin accessibility, and transcription factor binding, to gain insights into how CRM activities are controlled.

## EXPERIMENTAL PROCEDURES

### Dissections and Immunostaining

The construction of the *CRM-GAL4* reporter genes and transformants was described previously (Pfeiffer et al., 2008). Third-instar larvae offspring from crosses of *CRM-GAL4* males and *UAS-GFP* virgin females were screened for GFP-positive signals under a fluorescence microscope. Positive larvae were dissected in PBS and fixed in 4% paraformaldehyde for 25 min at room temperature (RT), washed, and blocked with 1% BSA and 0.3% Triton. Disc GFP-positive L3 larvae were stained using standard procedures. The primary antibodies used were rabbit anti-GFP (A6565, Invitrogen, Carlsbad, CA, USA), guinea-pig anti-Distalless and either rabbit anti-Hth or guinea pig anti-Hth (Estella et al., 2008), and mouse anti-engrailed (4D9, Developmental Studies Hybridoma Bank). In some cases, GFP fluorescence was imaged directly, without an antibody.

All secondary antibodies were from Invitrogen: goat anti-rabbit Alexa-488 (A11034), goat anti-guinea pig Alexa-555 (A21435), and goat anti-mouse Alexa-647 (A21236, which also recognizes guinea-pig primary IgGs). In some cases DAPI nuclear counterstain (D3571, Invitrogen) was performed.

Images were acquired with a Zeiss Axioplan equipped with Apotome and Axiovision software (Carl Zeiss, Jena, Germany) or a Leica SP5 confocal and Leica Application Suite (LAS) software. All images were assembled in ImageJ (http://rsb.info.nih.gov/ij), Adobe Photoshop and InDesign (Adobe Systems, San Jose, CA, USA).

### De Novo Discovery of Putative Binding Sites

Overrepresented DNA motifs were identified using the peak-motifs program as part of Regulatory Sequence Analysis Tools (RSAT) (Thomas-Chollier et al., 2011, 2012). Briefly, each set of DNA regions was tested for overrepresented 6-mers using a background model in which subword frequencies were

estimated using a Markov model trained on the DNA regions being analyzed. Overlapping, significant 6-mers were assembled and converted to position-specific scoring matrices (Thomas-Chollier et al., 2012). Significance scores in Figure 7 are $-\log_{10}$(E-value), with the E-value representing a multiple testing corrected p value for the seed 6-mer that was used to generate the position-specific scoring matrix (Thomas-Chollier et al., 2011).

Gene Ontology analysis was done using the AmiGO software from The Ontology project (http://www.geneontology.org/).

## SUPPLEMENTAL INFORMATION

Supplemental Information includes two figures and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2012.09.010.

## LICENSING INFORMATION



## ACKNOWLEDGMENTS

We thank Marc Smith for help with the screen, Gary Struhl and Andrew Tomlinson for advice on disc annotations, Crystal Sullivan for organizing trips to JFRC, and Chris Doe for comments on the manuscript. This work was supported by the Janelia Farm Visitor Program and National Institutes of Health grants GM054510 and GM058575 awarded to R.S.M.

Received: August 15, 2012
Revised: September 14, 2012
Accepted: September 17, 2012
Published online: October 11, 2012

## WEB RESOURCES

The URLs for data presented herein are as follows:

FlyLight Image Database, http://www.janelia.org/gal4-gen1
Gene Ontology, http://www.geneontology.org/
Image J, Image Processing and Analysis in Java, http://rsb.info.nih.gov/ij
REDfly, Regulatory Element Database for Drosophilia v3.2, http://redfly.ccr.buffalo.edu/index.php

## REFERENCES

Bailey, T.L., and Machanick, P. (2012). Inferring direct DNA binding from ChIP-seq. Nucleic Acids Res. Published online May 18, 2012. http://dx.doi.org/10.1093/nar/gks433.

Barolo, S. (2012). Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. Bioessays *34*, 135–141.

Blair, S.S. (2003). Lineage compartments in Drosophila. Curr. Biol. *13*, R548–R551.

Bonn, S., and Furlong, E.E. (2008). cis-Regulatory networks during development: a view of Drosophila. Curr. Opin. Genet. Dev. *18*, 513–520.

Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development *118*, 401–415.

Bras-Pereira, C., Bessa, J., and Casares, F. (2006). Odd-skipped genes specify the signaling center that triggers retinogenesis in Drosophila. Development *133*, 4145–4149.

Brockmann, A., Domínguez-Cejudo, M.A., Amore, G., and Casares, F. (2011). Regulation of ocellar specification and size by twin of eyeless and homothorax. Dev. Dyn. *240*, 75–85.

Bulger, M., and Groudine, M. (2010). Enhancers: the abundance and function of regulatory sequences beyond promoters. Dev. Biol. *339*, 250–257.

Crickmore, M.A., and Mann, R.S. (2006). Hox control of organ size by regulation of morphogen production and mobility. Science *313*, 63–68.

Crickmore, M.A., and Mann, R.S. (2007). Hox control of morphogen mobility and organ development through regulation of glypican expression. Development *134*, 327–334.

Crickmore, M.A., and Mann, R.S. (2008). The control of size in animals: insights from selector genes. Bioessays *30*, 843–853.

de Navas, L.F., Garaulet, D.L., and Sánchez-Herrero, E. (2006). The ultrabithorax Hox gene of Drosophila controls haltere size by regulating the Dpp pathway. Development *133*, 4495–4506.

del Valle Rodríguez, A., Didiano, D., and Desplan, C. (2012). Power tools for gene expression and clonal analysis in Drosophila. Nat. Methods *9*, 47–55.

Estella, C., and Mann, R.S. (2010). Non-redundant selector and growth-promoting functions of two sister genes, buttonhead and Sp1, in Drosophila leg development. PLoS Genet. *6*, e1001001.

Estella, C., McKay, D.J., and Mann, R.S. (2008). Molecular integration of wingless, decapentaplegic, and autoregulatory inputs into Distalless during Drosophila leg development. Dev. Cell *14*, 86–96.

Estella, C., Voutev, R., and Mann, R.S. (2012). A dynamic network of morphogens and transcription factors patterns the fly leg. Curr. Top. Dev. Biol. *98*, 173–198.

Estrada, B., Casares, F., and Sánchez-Herrero, E. (2003). Development of the genitalia in Drosophila melanogaster. Differentiation *71*, 299–310.

Field, L.H., and Matheson, T. (1998). Chordotonal organs of insects. Adv. Insect Physiol. *27*, 1–228.

Frankel, N., Davis, G.K., Vargas, D., Wang, S., Payre, F., and Stern, D.L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature *466*, 490–493.

Gallo, S.M., Gerrard, D.T., Miner, D., Simich, M., Des Soye, B., Bergman, C.M., and Halfon, M.S. (2011). REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in Drosophila. Nucleic Acids Res. *39*(Database issue), D118–D123.

Istrail, S., and Davidson, E.H. (2005). Logic functions of the genomic cis-regulatory code. Proc. Natl. Acad. Sci. USA *102*, 4954–4959.

Jan, Y.N., Ghysen, A., Christoph, I., Barbel, S., and Jan, L.Y. (1985). Formation of neuronal pathways in the imaginal discs of Drosophila melanogaster. J. Neurosci. *5*, 2453–2464.

Jenett, A., Rubin, G.M., Ngo, T.-T.B., Shepherd, D., Murphy, C., Dionne, H., Pfeiffer, B.D., Hibbard, K.L., Cavallaro, A., Hall, D., et al. (2012). A GAL4-driver line resource for Drosophila neurobiology. Cell Rep. *2*. Published online October 11, 2012. http://dx.doi.org/10.1016/j.celrep.2012.09.010.

Kumar, J.P. (2011). My what big eyes you have: how the Drosophila retina grows. Dev. Neurobiol. *71*, 1133–1152.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565–570.

Mann, R.S., and Morata, G. (2000). The developmental and molecular biology of genes that subdivide the body of Drosophila. Annu. Rev. Cell Dev. Biol. *16*, 243–271.

Manning, L., Purice, M.D., Roberts, J., Pollard, J.L., Bennett, A.L., Kroll, J.R., Dyukareva, A.V., Doan, P.N., Lupton, J.R., Strader, M.E., et al. (2012). Annotated embryonic CNS expression patterns of 5000 GMR Gal4 lines: a resource for manipulating gene expression and analyzing cis-regulatory motifs. Cell Rep. *2*. Published online October 11, 2012. http://dx.doi.org/10.1016/j.celrep.2012.09.010.

Morata, G. (2001). How Drosophila appendages develop. Nat. Rev. Mol. Cell Biol. *2*, 89–97.

Nern, A., Pfeiffer, B.D., Svoboda, K., and Rubin, G.M. (2011). Multiple new site-specific recombinases for use in manipulating animal genomes. Proc. Natl. Acad. Sci. USA *108*, 14198–14203.

Perry, M.W., Boettiger, A.N., Bothma, J.P., and Levine, M. (2010). Shadow enhancers foster robustness of Drosophila gastrulation. Curr. Biol. *20*, 1562–1567.

Pfeiffer, B.D., Jenett, A., Hammonds, A.S., Ngo, T.T., Misra, S., Murphy, C., Scully, A., Carlson, J.W., Wan, K.H., Laverty, T.R., et al. (2008). Tools for neuroanatomy and neurogenetics in Drosophila. Proc. Natl. Acad. Sci. USA *105*, 9715–9720.

Pfeiffer, B.D., Ngo, T.T., Hibbard, K.L., Murphy, C., Jenett, A., Truman, J.W., and Rubin, G.M. (2010). Refinement of tools for targeted gene expression in Drosophila. Genetics *186*, 735–755.

Rauskolb, C., and Irvine, K.D. (1999). Notch-mediated segmentation and growth control of the Drosophila leg. Dev. Biol. *210*, 339–350.

Struhl, G. (1982). Genes controlling segmental specification in the Drosophila thorax. Proc. Natl. Acad. Sci. USA *79*, 7380–7384.

Thomas-Chollier, M., Defrance, M., Medina-Rivera, A., Sand, O., Herrmann, C., Thieffry, D., and van Helden, J. (2011). RSAT 2011: regulatory sequence analysis tools. Nucleic Acids Res. *39*(Web Server issue), W86–W91.

Thomas-Chollier, M., Herrmann, C., Defrance, M., Sand, O., Thieffry, D., and van Helden, J. (2012). RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. Nucleic Acids Res. *40*, e31.

Wolff, T., and Ready, D.F. (1991). The beginning of pattern formation in the Drosophila compound eye: the morphogenetic furrow and the second mitotic wave. Development *113*, 841–850.