

# *cis*-Regulatory Complexity within a Large Non-Coding Region in the *Drosophila* Genome

Mukta Kundu<sup>1</sup>\*, Alexander Kuzin<sup>1</sup>\*, Tzu-Yang Lin<sup>2</sup>, Chi-Hon Lee<sup>2</sup>, Thomas Brody<sup>1\*</sup>, Ward F. Odenwald<sup>1\*</sup>

**1** Neural Cell-Fate Determinants Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, United States of America, **2** Section on Neuronal Connectivity, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, United States of America

## Abstract

Analysis of *cis*-regulatory enhancers has revealed that they consist of clustered blocks of highly conserved sequences. Although most characterized enhancers reside near their target genes, a growing number of studies have shown that enhancers located over 50 kb from their minimal promoter(s) are required for appropriate gene expression and many of these 'long-range' enhancers are found in genomic regions that are devoid of identified exons. To gain insight into the complexity of *Drosophila cis*-regulatory sequences within exon-poor regions, we have undertaken an evolutionary analysis of 39 of these regions located throughout the genome. This survey revealed that within these genomic expanses, clusters of conserved sequence blocks (CSBs) are positioned once every 1.1 kb, on average, and that a typical cluster contains multiple (5 to 30 or more) CSBs that have been maintained for at least 190 My of evolutionary divergence. As an initial step toward assessing the *cis*-regulatory activity of conserved clusters within gene-free genomic expanses, we have tested the *in-vivo* enhancer activity of 19 consecutive CSB clusters located in the middle of a 115 kb gene-poor region on the 3<sup>rd</sup> chromosome. Our studies revealed that each cluster functions independently as a specific spatial/temporal enhancer. In total, the enhancers possess a diversity of regulatory functions, including dynamically activating expression in defined patterns within subsets of cells in discrete regions of the embryo, larvae and/or adult. We also observed that many of the enhancers are multifunctional—that is, they activate expression during multiple developmental stages. By extending these results to the rest of the *Drosophila* genome, which contains over 70,000 non-coding CSB clusters, we suggest that most function as enhancers.

**Citation:** Kundu M, Kuzin A, Lin T-Y, Lee C-H, Brody T, et al. (2013) *cis*-Regulatory Complexity within a Large Non-Coding Region in the *Drosophila* Genome. PLoS ONE 8(4): e60137. doi:10.1371/journal.pone.0060137

**Editor:** Arnar Palsson, University of Iceland, Iceland

**Received:** November 29, 2012; **Accepted:** February 21, 2013; **Published:** April 22, 2013

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

**Funding:** This work was supported by National Institutes of Health (NIH), US government. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: brodyt@ninds.nih.gov (TB); odenwaldw@ninds.nih.gov (WFO)

† These authors contributed equally to this work.

## Introduction

Transcriptional activity of dynamically expressed genes is controlled in part by multiple pattern-specific enhancers that regulate different aspects of a gene's complete spatial/temporal expression, and many of these enhancers are clustered close to their regulated genes (for examples see [1–4]). A survey of *cis*-regulatory DNA surrounding developmental genes indicates that the *Drosophila* genome may harbor more than 50,000 enhancers [5]. Analysis of chromosome DNase-1 hypersensitivity profiles suggests that many genes that are active during embryonic development use multiple enhancers [6]. Other surveys have revealed that transcription factor (TF) high occupancy target regions map to active embryonic enhancers located close to structural genes [7]. These studies have also shown that most enhancers are functionally autonomous, since they correctly regulate heterologous transgene expression dynamics outside of their endogenous chromosomal environment (reviewed by [8] [9]).

Phylogenetic footprinting of vertebrate, *Drosophila* or nematode genomic DNA has revealed that enhancers can be distinguished from other essential gene regions based on their characteristic pattern of conserved sequences [2] [4] [10–14]. Collectively, these

studies have shown that most enhancers are made up of clusters of 5 to 30 or more conserved sequence blocks (CSBs) [15–17]. On average, enhancer CSB clusters span ~1 kb and are flanked by non-conserved DNA of variable length. Self-alignment of conserved sequences within enhancers reveals that their CSBs contain repeat and palindrome sequence elements that make up, on average, over 60% of their sequences [17]. TF DNA-binding site searches of characterized enhancers reveals that while most CSBs contain core docking sites for known TFs, much of their conserved sequences consists of novel repeat, palindromic or single copy sequence elements. Genome-wide systematic manual curation of conserved *Drosophila* DNA has identified over 70,000 non-coding conserved sequence clusters [17]. Thus far, studies that have tested individual clusters closely associated with *Drosophila* developmental genes, such as *nerfin-1* [2], *hunchback* [18], *sloppy-paired* [3] and *castor* [4], have all shown that each cluster is an enhancer and many enhancers are multifunctional in that they regulate embryonic and/or adult developmental gene expression.

While examination of gene neighborhoods reveals that most non-coding regions associated with developmental determinants contain *cis*-regulatory sequences [5], sequence conservation tracks

that span the fly genome [19] reveal that clusters of conserved sequences are not exclusively restricted to gene neighborhoods. Large gene-free regions termed 'gene deserts' in vertebrates are thought to consist of reservoirs of enhancers that function at a distance to regulate gene expression [20] [21]. For example, a 645 kb region that separates the vertebrate Iroquois genes *Irx3* and *Irx5* contains multiple Iroquois-specific enhancers [22]. Likewise, in *Drosophila* a distal non-coding region adjacent to the *Drosophila iroquois* gene complex contains multiple enhancers required for

wild-type *irx* gene expression [23]. In addition, analysis of *cut* gene regulation has identified one of its wing margin enhancers 80 kb away from its proximal promoter region [24]. Other studies have located *invected* gene *cis*-regulatory elements positioned ~78 kb upstream of its transcribed sequence [25] and remote shadow enhancers have been described for *shavenbaby*, a transcript of the *ovo* locus [26].

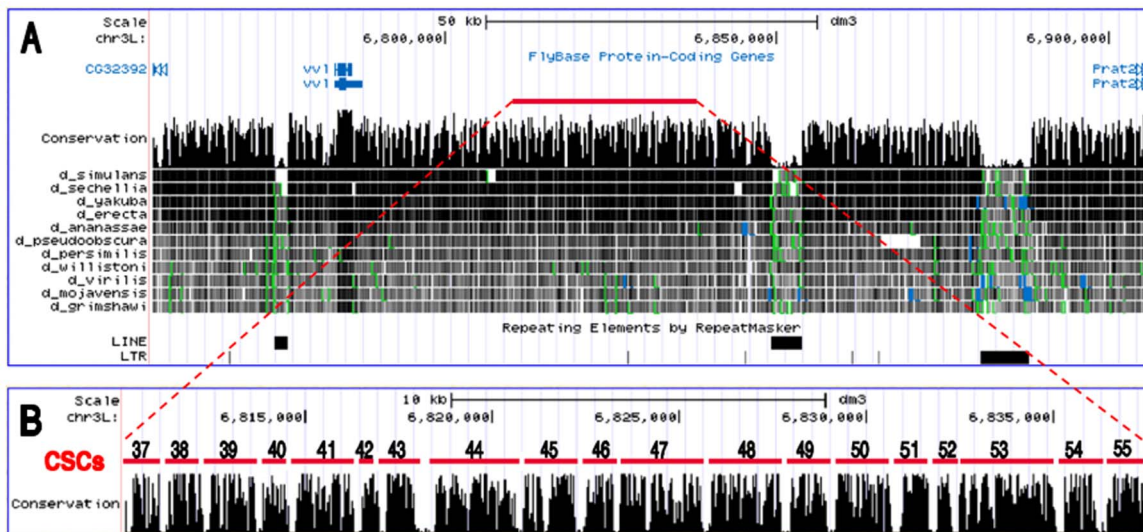
To what extent do clustered CSBs that reside in gene-free genomic expanses function as *cis*-regulatory enhancers? As an

**Table 1.** *D. melanogaster* CSB clusters within non-coding regions >65 kb.

Chromosome Location	Flanking Genes (5' - 3')	Length (kb)	# of CSCs
chr2R:2,416,250–2,487,059	jing <sup>→</sup> –CG15233 <sup>←*</sup>	70.8	63
chr2R:4,156,848–4,234,973	CG30371 <sup>←</sup> –Pdm3 <sup>→</sup>	78.1	70
chr2R:10,944,185–11,025,474	CG33467 <sup>←</sup> –chn <sup>→</sup>	81.3	59
chr2R:15,893,693–15,999,491	CG16898 <sup>←</sup> –18 w <sup>→</sup>	105.8	89
chr2R:16,243,641–16,316,603	CG11192 <sup>←</sup> –CG12484 <sup>→</sup>	73.0	57
chr2R:17,300,726–17,367,629	Sdc <sup>←</sup> –Sdc <sup>←</sup>	67.0	73
chr2L:1,614,340–1,690,699	RFeSP–chinmo	76.4	65
chr2L:8,564,841–8,631,270	Sema-1a <sup>→</sup> –Sema-1a <sup>→</sup>	66.4	53
chr2L:12,822,253–12,911,000	Kek1 <sup>←</sup> –ACXC <sup>→</sup>	88.7	88
chr2L:17,827,895–17,914,722	CadN2 <sup>←</sup> –CG43271 <sup>←</sup>	86.8	85
chr2L:20,970,616–21,052,985	CG42238 <sup>←</sup> –betaltnt-nu <sup>→</sup>	82.4	66
chr3L:5,013,564–5,097,509	CG12027 <sup>←</sup> –CG34047 <sup>←</sup>	83.9	68
chr3L:5,269,532–5,337,221	Shep <sup>←</sup> –Lama <sup>←</sup>	65.6	61
chr3L:6,788,236–6,903,573	vvf <sup>→</sup> –Prat2 <sup>←</sup>	115.3	90
chr3L:10,691,176–10,765,559	NijA <sup>←</sup> –CG43245 <sup>→</sup>	74.4	78
chr3L:13,658,522–13,771,330	bru3 <sup>←</sup> –CG34243 <sup>←</sup>	112.8	98
chr3L:15,335,752–15,401,048	Toll6 <sup>→</sup> –CG33259 <sup>→</sup>	65.3	76
chr3L:15,722,537–15,801,891	Comm <sup>←</sup> –CG6244 <sup>→</sup>	79.4	66
chr3L:18,297,975–18,389,946	grim <sup>←</sup> –rpr <sup>←</sup>	92.0	99
chr3L:19,140,122–19,229,706	Fz2 <sup>←</sup> –CG33647 <sup>→</sup>	89.6	86
chr3L:21,835,505–21,900,787	CG14563 <sup>←</sup> –mub <sup>→</sup>	65.3	50
chr3L:22,076,308–22,158,687	msopa <sup>→</sup> –Olf413 <sup>→</sup>	99.2	86
chr3R:814,412–910,094	CG2022 <sup>←</sup> –corto <sup>→</sup>	95.7	58
chr3R:2,735,634–2,878,317	Antp <sup>←</sup> –Sod-1 <sup>→</sup>	142.6	140
chr3R:4,269,254–4,335,814	PQBP-1 <sup>←</sup> –OR85b <sup>←</sup>	66.6	57
chr3R:6,262,387–6,337,098	Cyp12e1 <sup>←</sup> –hth <sup>←</sup>	74.7	63
chr3R:7,096,987–7,172,245	CG31386 <sup>←</sup> –KP78b <sup>←</sup>	75.3	63
chr3R:10,258,903–10,335,707	cv-c <sup>←</sup> –HtrA1 <sup>←</sup>	76.8	70
chr3R:10,747,590–10,840,491	CG3837 <sup>←</sup> –CG14861 <sup>→</sup>	92.9	92
chr3R:11,378,494–11,458,604	CG18516 <sup>←</sup> –CG5302 <sup>→</sup>	80.1	65
chr3R:18,671,035–18,741,129	CG4704 <sup>←</sup> –klg <sup>→</sup>	70.1	66
chr3R:19,238,330–19,307,587	CG4374 <sup>←</sup> –CG31225 <sup>→</sup>	69.3	52
chr3R:21,226,341–21,299,237	CG31439 <sup>←</sup> –CG5127 <sup>→</sup>	72.9	70
chr3R:24,179,225–24,287,776	Or98b <sup>→</sup> –beat-VI <sup>→</sup>	108.6	97
chrX:972,189–1,039,710	CG3655 <sup>←</sup> –CG14626 <sup>←</sup>	67.5	62
chrX:3,866,525–3,971,906	CG6414 <sup>←</sup> –CG32790 <sup>→</sup>	105.4	83
chrX:7,004,130–7,091,461	fz4 <sup>←</sup> –CG9650 <sup>→</sup>	87.3	75
chrX:16,018,498–16,105,580	disco-r <sup>←</sup> –disco <sup>←</sup>	87.1	72
chrX:17,218,418–17,291,800	B-H2 <sup>→</sup> –BH-1 <sup>→</sup>	73.4	73

\*Arrows indicate direction of transcription.

doi:10.1371/journal.pone.0060137.t001



**Figure 1. DNA conservation spanning the *Drosophila vvl* locus.** (A) Shown is an UCSC Genome Browser view covering 150 kb of DNA located on the left arm of the *D. melanogaster* 3<sup>rd</sup> chromosome (<http://genome.ucsc.edu>). The linear representation includes the *vvl* transcribed region and flanking DNA starting from the upstream neighboring CG32392 gene and extending to the 3' end of the downstream *Prat2* transcribed sequence. A 12 species *Drosophila* DNA conservation track [19] reveals the presence of conserved sequence clusters throughout the locus. The red bar (positioned 30 kb downstream of the *vvl* transcribed sequence) covers the 27 kb of the non-coding intergenic region that was examined in this study for the presence of independent *cis*-regulatory enhancers. Aligned below the conservation track are identified Line and LTR repeat elements present within the *D. melanogaster* DNA that are not present in the same orthologous positions within many of the other species included in the conservation analysis. (B) An expanded view of the intergenic region studied for its *cis*-regulatory activity (highlighted in panel A) reveals 19 consecutive conserved sequence clusters that were independently tested for their *cis*-regulatory activity. Cluster numbers correspond to their designation in the *cis*-Decoder *D. melanogaster* genome-wide sequence conservation database [17].  
doi:10.1371/journal.pone.0060137.g001

initial step toward addressing this question, we have examined the frequency of conserved sequence clusters and their CSB composition within 39 different *Drosophila* regions lacking exons that are greater than 65 kb in length, located throughout the genome. Comparative analysis of orthologous regions in other *Drosophila* species reveals that within these non-coding regions, such clusters occur on average once every 1.1 kb and, similar to known enhancers, each consists of multiple CSBs that contain unique sets of repeat and palindromic elements. Our comparative analysis also reveals that, like characterized enhancers, subsets of adjacent CSBs clustered within exon-poor regions also form ‘super-blocks:’ regions of invariantly spaced CSBs.

To address the function of ‘gene-distant’ CSB clusters, we tested the *cis*-regulatory function(s) of consecutive clusters within one gene-free 115 kb expanse that contains 90 conserved clusters located on the 3<sup>rd</sup> chromosome between the *vvl* and *Prat2* genes. Earlier studies have identified tracheal and neural enhancers within this region [17] [27]. Our *in vivo* enhancer-reporter studies on 19 consecutive clusters demonstrate that each functions as an enhancer that regulates reporter expression in defined sets of tissues and developmental phases, thus representing a remarkable diversity of expression patterns. Many of these enhancers have proven to be multifunctional, directing expression dynamics at different times during development and in different cell types. Based on these findings, we estimate that the fly genome may contain over 70,000 enhancers and many are likely to be multifunctional.

## Results

### Sequence conservation within *Drosophila* gene-poor genomic expanses

As an initial step toward assessing the frequency and diversity of long-range enhancers in the *Drosophila* genome, we documented the occurrence of CSB clusters within 39 genomic expanses that each span at least 65 kb and are devoid of known or predicted protein encoding sequences (Table 1). The largest of these regions was an expanse of non-coding sequence of ~142 kb associated with the *Antp* gene. None of the gene-poor regions are as large as those in mammals; it is known that the *Drosophila* genome is approximately an order of magnitude more compact, and the density of conserved elements is greater in *Drosophila* [19]. Also included in the phylogenetic footprinting survey were sequences in intragenic regions within large introns. *EvoPrints* covering these non-coding regions located throughout the genome (covering a total of ~3.26 Mb or ~2.7% of the euchromatic genome) revealed a near uniform conserved sequence cluster density of one per every ~1.1 kb even in regions located over 50 kb from transcribed sequences (see Materials and Methods for *EvoPrint* conditions and [17] for database details). Clusters of CSBs were defined as independent when they were separated by at least 150 bp of non-conserved sequences and were resolved with *EvoPrint* conditions that represented a cumulative evolutionary divergence (the length of time that multiple species have evolved separately from one another) in excess of 190 million years. When compared to similar clusters close to transcriptional start sites, we did not detect significant differences between their genome density/spacing, their average number of CSBs, or differences in the degree of evolutionary sequence conservation of their CSBs (Figures 1, 2, S1, S2, S3, S4, and data not shown).



*mojavensis* and *D. grimshawi* genomes. Less or non-conserved DNA is shown as lower case gray letters and the lower-case red-font bases indicate invariant spacer length DNA between CSBs. Colored highlighted conserved sequences within the *vvl-38* (blue), *vvl-39* (yellow), and *vvl-41* (purple) clusters represent repeat elements that are discussed in Text S1. doi:10.1371/journal.pone.0060137.g002

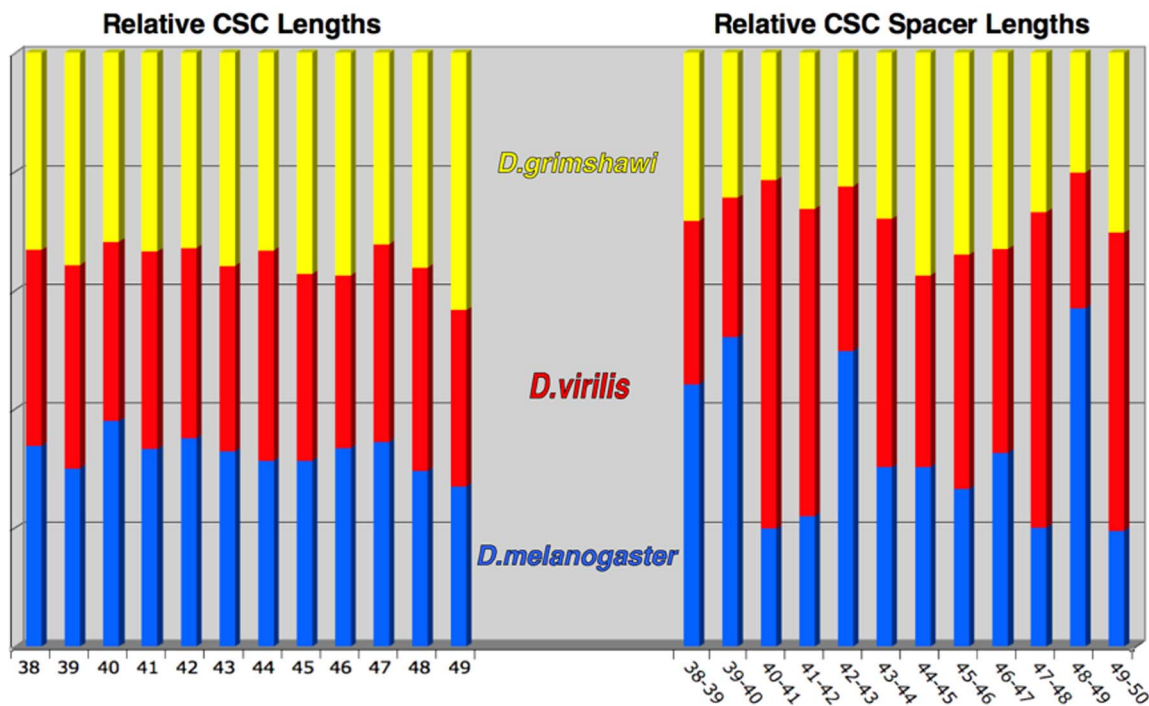
A prominent feature of both gene-proximal and -distal conserved clusters is that the intervening non-conserved spacer regions between adjacent clusters have greater sequence length variability among different species in comparison to regions within clusters [2]. A graphic representation of this is shown in Figure 3. The lower sequence length variability within clusters indicates that there are differences in the structural constraints within clusters in contrast to inter-clustal flanking sequences. Pair-wise alignments of clustered CSBs among different drosophilids reveal that in many cases spacing between adjacent CSBs is not variable (Figure 2; and data not shown). We refer to neighboring CSBs that are separated by a conserved spacer length as a ‘super-block’ and suggest that inherent structural requirements for enhancer function place evolutionary constraints on occurrence of indels between CSBs within a super-block. There are, however, species-specific exceptions to the sequence length constraints observed between CSBs, and the variability observed in an individual species or in a subset of species is informative with regard to functional compartmentalization within an enhancer (see below).

#### *cis*-Regulatory analysis of consecutive CSB clusters within a gene-poor region

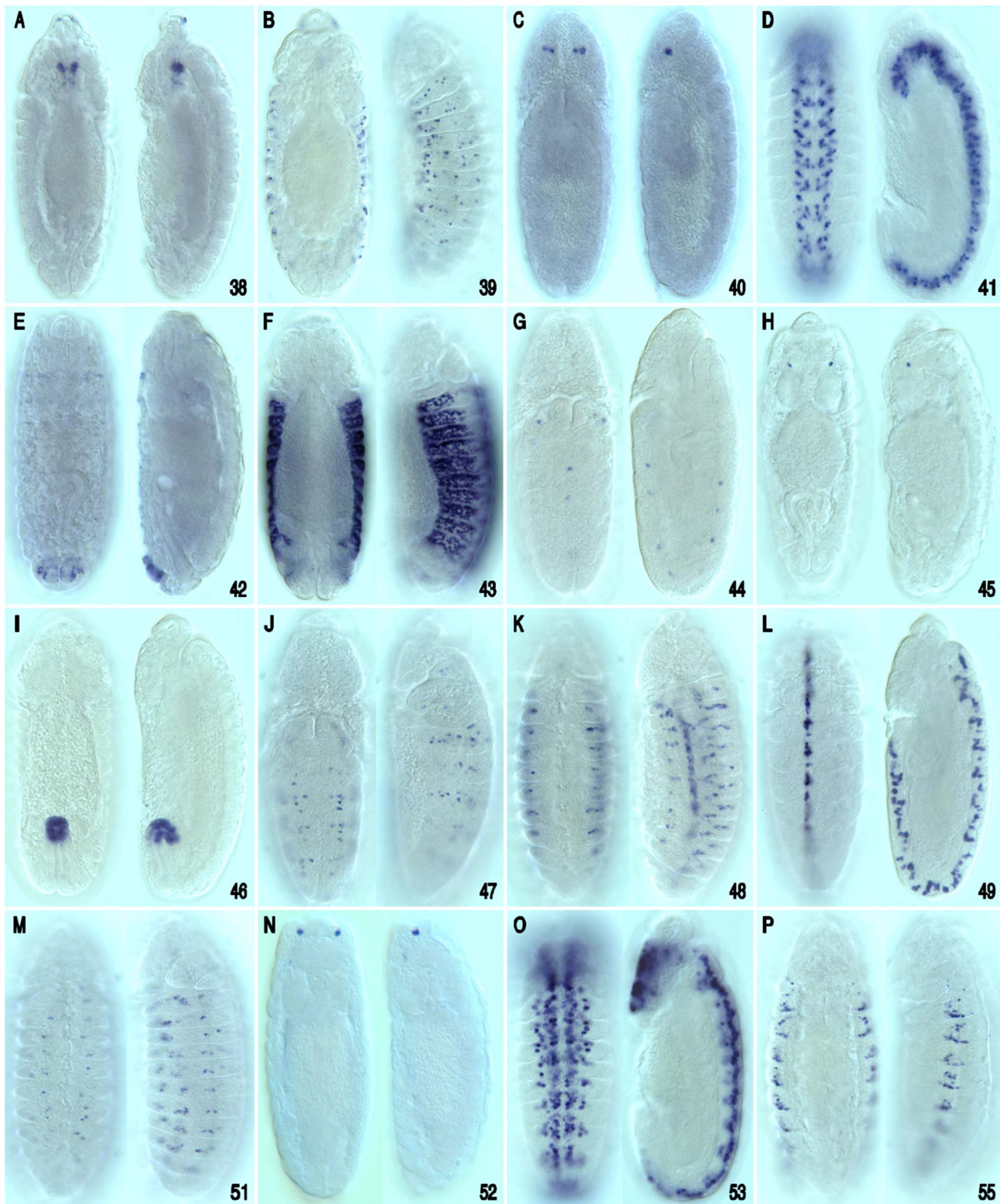
Many of the non-coding regions examined in this study are flanked by developmentally regulated genes. For example, a 115 kb non-coding expanse separates the *ventral veins lacking* (*vvl*) gene, encoding a POU-domain containing transcription factor, and the Phosphoribosylamidotransferase-2 (*Prat2*) gene (Figure 1a).

*EvoPrint* analysis of this region identifies 90 conserved sequence clusters and each contains CSBs that have been conserved for more than 190 million years of cumulative evolutionary divergence (Figures 2 and S1, S2, S3, S4; and data not shown). Both *vvl* and *Prat2* functions are required during multiple phases of development and in a variety of cell types. For example, *vvl* function is required for the correct migration of tracheal cells and glia during embryonic development [28], specification of motor neuron identity in the embryonic ventral cord [29], correct development of the peripheral nervous system [30], embryonic brain neural precursor cell identity [31], development of adult external mechanosensory organs [32], dendritic targeting of olfactory projection neurons in adults [33], [34], the correct temporal identity of optic lobe neurons [35] and for the development of wing imaginal disc cells [36]. In addition, *vvl* may also be required for correct epidermal development, as it is dynamically expressed in the epidermis during embryonic development [37]. *Prat2* is expressed in embryos and larvae, as well as in testis [38], and *in situ* mRNA localization studies reveal expression in the embryonic yolk nuclei [39]. *Prat2* is also required for metamorphosis during pupal development [39].

The dynamic expression of both of these genes may be regulated in part by multiple close-range enhancers that reside near their minimal promoters. Indeed, *vvl* has over 30 CSB clusters positioned within 25 kb of its transcribed DNA (Figure 1A; and data not shown). To assess the *cis*-regulatory nature of CSB clusters positioned beyond the local confines of these structural



**Figure 3. Evolutionary constraints on CSB cluster structure.** Multi-species analysis of CSB clusters and their flanking spacer regions reveals that the less-conserved spacer DNA has greater evolutionary sequence length variability when compared to their flanking CSB clusters. Shown, are percentage base pair length differences between *D. melanogaster* (blue) *D. virilis* (red) and *D. grimshawi* (yellow) *vvl* clusters 38 through 49 and the percent differences within their flanking spacer regions (each column represents 100%). doi:10.1371/journal.pone.0060137.g003



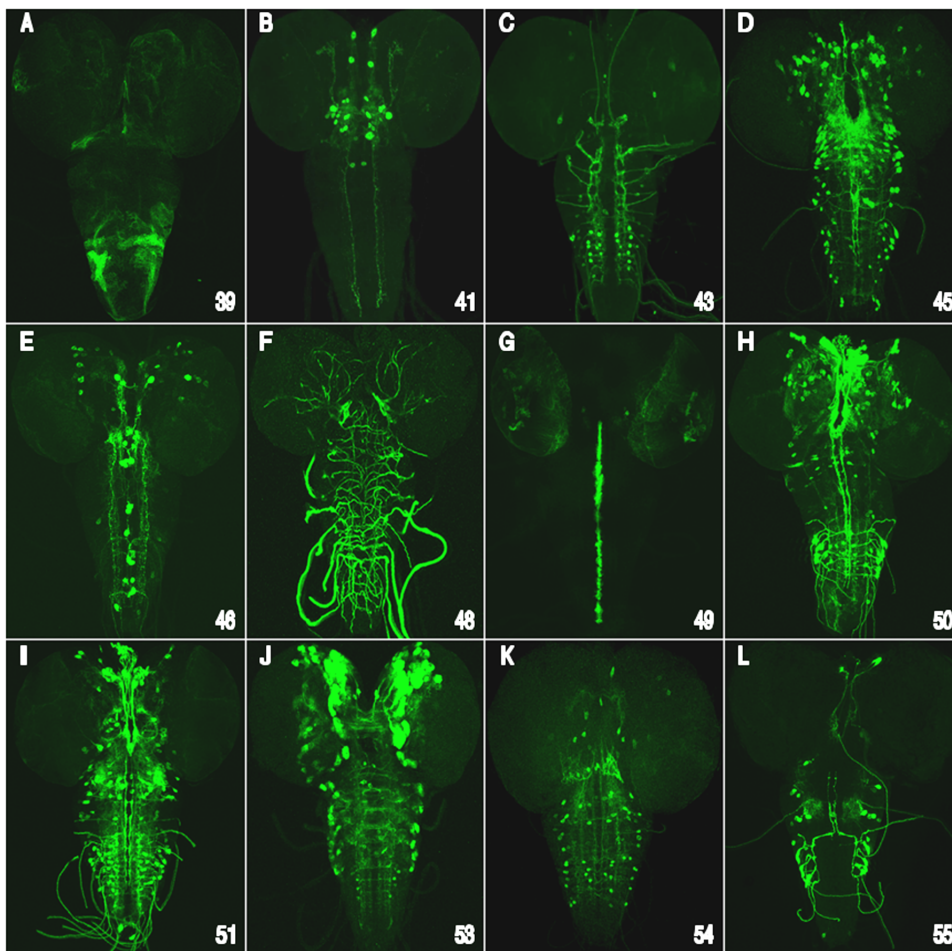
**Figure 4. Conserved cluster *cis*-regulatory enhancer activity during embryonic development.** Enhancer/reporter transgene expression analysis during embryonic development reveals that many of the tested CSB clusters are functionally independent embryonic enhancers that direct expression in different spatial/temporal patterns within the developing embryo. Shown are enhancer-reporter embryo expression patterns for 16 of the 19 consecutive clusters tested. Whole-mount mRNA stained embryos (staging according to Hartenstein and Campos-Ortega [58]; dorsal or ventral views adjacent to lateral views are shown for each cluster-reporter transgene; anterior up) to reveal peak reporter mRNA expression detected by a digoxigenin labeled Gal4 riboprobe for each of the cluster/enhancer-reporter constructs. The numbers in the lower right corner of each panel correspond to the clusters shown in Figure 1 and Figures S1, S2, S3, S4 and described in Table 2. (A) Dorsal and lateral view of a stage 13 embryo. *vvl-38* activates transgene reporter expression in a small cluster of cells within or near the developing antenno-maxillary complex and within a cluster of anterior gut epidermal cells positioned adjacent to the cephalic lobes. (B) Dorsal and lateral surface views of a stage 13 embryo. *vvl-39* drives expression in putative PNS cells. (C) Dorsal and lateral views of a stage 10 embryo. *vvl-40* activates expression in two adjacent NBs within each cephalic brain lobe. (D) ventral and lateral view of a stage 11 embryo. *vvl-41* drives expression in a set of NBs after they have generated their first GMC progeny. (E) Dorsal and lateral view of a stage 15 embryo. *vvl-42* drives expression in cells of the gut ectoderm. (F) Dorsal and later surface view of stage 13 embryo. *vvl-43* drives expression in late lateral ectodermal cells. (G) Dorsal and lateral view of a stage 10 embryo. *vvl-44* drives expression in a

single midline cell per segment and in segmentally repeated lateral cells, possibly PNS cells. (H) Dorsal and lateral view of a stage 14 embryo. *vv1-45* drives expression in a bilateral pair of brain neurons. (I) Dorsal and dorsal-lateral views of a stage 13 embryo. *vv1-46* drives expression in the posterior midgut. (J) Dorsal and lateral view of a stage 11 embryo *vv1-47* drives in a few unidentified cells per hemisegment in the neuroectoderm and CNS. (K) Deep ventral and lateral view of a stage 10 embryo. *vv1-48* drives expression in segmentally repeated clusters that appear to be tracheal placodes. (L) Ventral and lateral view of stage 11 embryo. The *vv1-49* cluster activates reporter expression in ventral cord midline glial cells (also shown in Figure 8). (M) Ventral and lateral view of a stage 13 embryo. *vv1-51* drives expression in segmentally repeated putative neurons in the peripheral nervous system. (N) Dorsal and lateral views of a stage 14 embryo. The *vv1-52* cluster activates reporter expression in two bilaterally symmetrical cells within the antenno-maxillary complex. (O) Ventral and lateral views of a stage 12 embryo. *vv1-53* cluster drives expression in CNS NBs (both brain and ventral cord) during late NB lineage development. (P) Ventral and lateral surface views of a stage 14 embryo. *vv1-55* activates expression in cells that line tracheal branches.

doi:10.1371/journal.pone.0060137.g004

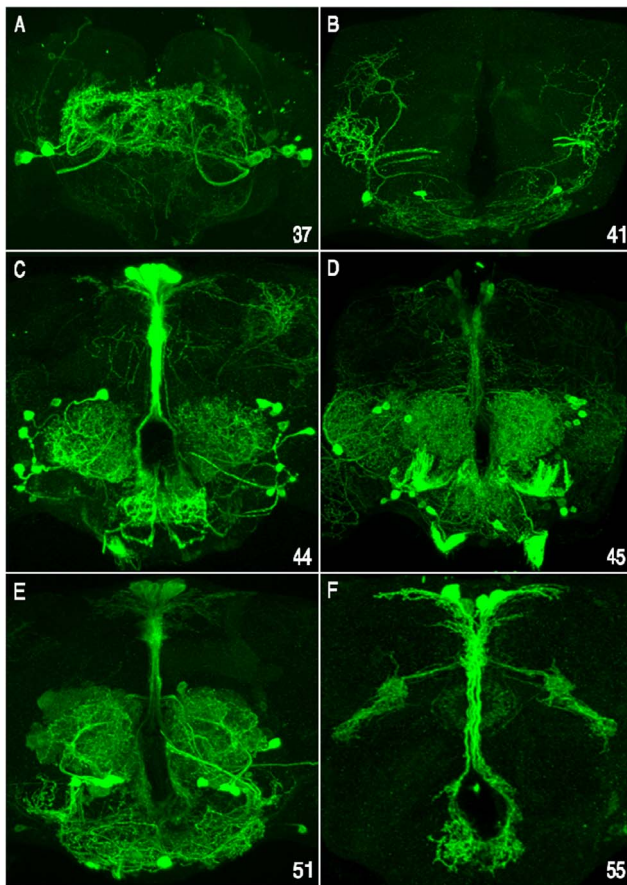
genes, we individually tested 19 consecutive conserved clusters within a 27 kb non-coding region (located 30 kb downstream of *vv1* and 57 kb from the 3' end of *Prat2*) for their ability to regulate transgene reporter expression during embryonic development, in 3<sup>rd</sup> instar larvae, and in the adult brain (Figures 4, 5, and 6). Each of the tested clusters (*vv1-37* to *vv1-55*, so named because of their proximity to *vv1*), coincides with phastcon peaks present in the

UCSC genome browser conservation track (Figure 1B) [19]. Figure 2 shows a relaxed *EvoPrint* of five of the consecutive clusters (*vv1-37* to *vv1-41*) and highlights sequences that are conserved in all, or all but one, of the 12 species used in the analysis. A relaxed *EvoPrint* of the remaining clusters is shown in Figures S1, S2, S3, and S4. Super-blocks, multiple CSBs separated by conserved spacer lengths, are indicated by red-colored lowercase bases



**Figure 5. Expression of enhancer/reporter transgenes in the larval CNS.** (A–L) During 3<sup>rd</sup>-instar larval development, most enhancer-*Gal4* transgenes from *vv1-37* to *vv1-55* (twelve are illustrated) activate UAS/GFP-CD8 tagged reporter expression in neural precursors, neurons or glia within sub-regions of the cephalic lobes and in the thoracic ventral cord. Shown are stacked images of dorsal views of dissected CNS preparations from wandering third-instar larva (anterior up). (A) *vv1-39* activates reporter expression in a subset of brain and ventral cord glia. (B) *vv1-41* drives reporter expression in a set of subesophageal ganglion (SOG) interneurons. (C–E) *vv1-43*, *-45* and *-46* activate expression in different subsets of ventral cord and/or brain neurons. (F) *vv1-48* drives expression in cells that line the tracheal tubes associated with the brain and ventral cord. (G) *vv1-49* activates reporter expression in CNS midline cells, presumably glia. (H and I) *vv1-50* and *-51* drive expression in subsets of brain and ventral cord neurons. (J) *vv1-53* activates reporter expression in brain and ventral cord NB lineages and in their neurons. (K) The *vv1-54* cluster drives reporter expression in subsets of brain and ventral cord neurons. (L) *vv1-55* activates reporter expression in a subset of both brain and ventral cord neurons. Based on the presence of membrane tagged GFP with in axons that exit the ventral cord, many of the neurons are most likely motor neurons.

doi:10.1371/journal.pone.0060137.g005



**Figure 6. Expression of enhancer/reporters within the adult brain.** Many of the tested conserved regions (six are illustrated) activate reporter expression in neurons or glia positioned within different sub-regions of the central brain. Shown are ventral (A) or anterior (B–F) views of adult brain. (A) *vvl-37* drives expression in several SOG neurons whose axons project across the midline. (B) *vvl-41* drives expression in several SOG neurons that project across the midline or dorsally. (C–F) *vvl-44*, *-45*, *-51* and *-55* all activate reporter expression in putative insulin-producing neurons (IPCs) [59] (C) *vvl-44* drives expression in IPCs and a set of lateral neurons whose dendrites fill the olfactory lobe. (D) *vvl-45* drives expression in a set of ventral brain neurons whose dendrites fill the olfactory lobe and the lateral brain. (E) *vvl-51* drives expression in IPCs and a set of ventral neurons whose axons and dendrites project into the olfactory lobe. (F) *vvl-55* drives expression in IPCs and in presumptive ellipsoid body neurons. doi:10.1371/journal.pone.0060137.g006

between CSBs (Figure 2). Also highlighted in the *EvoPrint* are prominent repeat sequences detected by *cis*-Decoder CSB cluster self-alignments (Figures 2 and S1, S2, S3, and S4).

We tested each of the CSB clusters independently using gypsy-insulated enhancer/reporter transgenes [17] [40]. To control for chromosomal integration-specific events that could influence reporter expression, we employed the *phiC31* mediated site-specific integration system to insure that all reporter transgenes were inserted into the same chromosomal environment [41] [42]. Remarkably, all 19 clusters functioned as spatial/temporal specific enhancers and 10 of these clusters generated distinct expression patterns in all three tested developmental phases (embryonic, larval and adult brain; summarized in Table 2 and Figures 4, 5, 6 and Text S1). In addition, five of the enhancers were active in just two of the developmental windows examined, while only four

enhancers restricted their regulatory behavior to a single developmental phase.

The majority of the reporter expressions observed were remarkably complex bilaterally symmetrical patterns that encompassed a diversity of cell types whose identities, in some cases, are unknown. Seventeen of the clusters activated unique pattern-specific expression during embryonic development in a wide diversity of tissues and cell types, including CNS neuroblasts (NBs), neurons and/or glia, PNS precursor cells, ectodermal cells and cells lining sub-regions of the gut and trachea (Figure 4; and Text S1). Fourteen of the clusters generated cell and/or region specific expression patterns within the 3<sup>rd</sup> instar larva CNS (Figure 5; and data not shown), while thirteen drove reporter expression within overlapping sub-regions and/or cell types of the adult brain (Figure 6; and data not shown). The *cis*-regulation of adjacent enhancers exhibited unique non-overlapping dynamic expression patterns (Figure 4). In addition, *cis*-Decoder analysis revealed that each of the CSB clusters contained unique combinations of repeat and palindromic elements (Figures 2 and Figures S1, S2, S3, and S4). High-resolution views of each of the expression patterns illustrated in Figures 4, 5, and 6 are available at the *cis*Patterns website (<http://cispatterns.ninds.nih.gov/>). Although many of the enhancer expression patterns matched sub-patterns of *vvl* expression, we are unable to state with certainty that these enhancers regulate *vvl* expression.

#### Species-specific variability in CSB spacing within enhancers

Our previous studies have shown that intervening non-conserved spacer regions between adjacent CSB clusters exhibit greater inter-species sequence length variability when compared to sequence length variability between CSBs within clusters [2]. Analysis of the spatial distribution of CSBs within the *vvl-49* cluster in different orthologous DNAs revealed that the *D. grimshawi* cluster has an additional 466 bp of non-conserved DNA within its central region that was not found in the other species (Figure 7). To determine if the *D. grimshawi* insertion indicated that the 1<sup>st</sup> and 2<sup>nd</sup> halves of the *vvl-49* cluster represent two closely spaced enhancers or semi-autonomous functional sub-domains of a single enhancer, we tested the corresponding *D. melanogaster* cluster halves for independent embryonic enhancer activity. The *vvl-49* enhancer activates transgene expression in a subset of ventral cord midline cells during embryonic stage 11 and expression persists throughout development (Figures 4, 6 and 7). Midline expression is most likely mediated via the midline cell-identity TFs Single-minded and Tango; their consensus DNA-binding site is present within four of the *vvl-49* CSBs with two binding sites residing in each cluster half (Figure 7). Figure 8 illustrates the temporal progression of the full *vvl-49* enhancer expression and the expression driven by the 1<sup>st</sup> and 2<sup>nd</sup> halves of the *D. melanogaster vvl-49* CSB cluster, *vvl-49a* and *vvl-49b* (see Figure 7A for CSB boundaries). Expression of the full enhancer was maintained from stage 11 through stage 15 (stages 11–13 illustrated in Fig. 8A). While the onset and timing of expression at stage 11 for the upper half of the cluster (*vvl-49a*) was essentially identical to the full cluster, subsequent reporter expression rapidly declined, so that by stage 13 there was only weak expression in midline cells (Figure 8B). In contrast, reporter expression driven by the lower half (*vvl-49b*) was detected in the ventral cord midline, but in considerably fewer cells than expression of the full cluster or its 1<sup>st</sup> half (Figure 8C). Taken together, the different *cis*-regulatory behaviors of the upper and lower halves indicate that the upper half of the cluster may function to establish midline expression and the lower half to maintain full midline expression. Additional studies that address



**Table 2.** *cis*-Regulatory activity of consecutive *vvl* conserved sequence clusters.

Cluster	Embryo	Larva	Adult	Figure
<i>vvl-37</i>	Negative	Negative	Subset of brain neurons	6A
<i>vvl-38</i>	antenna-maxillary complex & anterior gut	Negative	negative	4A
<i>vvl-39</i>	PNS glia	Putative ventral cord glia	Putative glia	4B, 5A
<i>vvl-40</i>	Pair of cephalic lobe NBs	Negative	Putative glia	4C
<i>vvl-41</i>	CNS neuroblast late lineage	Subset of brain neurons	Subset of brain neurons	4D, 5B, 6B
<i>vvl-42</i>	Posterior gut and ectoderm	Negative	Negative	4E
<i>vvl-43</i>	Late ectoderm	Ventral cord neurons	Subset of brain neurons	4F, 5C
<i>vvl-44</i>	Midline & PNS precursors	Negative	Central brain neurons and IPCs	4G, 6C
<i>vvl-45</i>	At stage 15, single neuron per cephalic lobe	Subset of brain and ventral cord neurons	Subset of central brain neurons and IPCs	4H, 5D, 6D
<i>vvl-46</i>	Gut	Ventral midline and brain neurons	Subset of brain neurons	4I, 5E
<i>vvl-47</i>	CNS and Neuroectoderm	A few cells in the SOG	Negative	4J
<i>vvl-48</i>	Trachea	Tracheal tubes associated with the brain & ventral cord	Optic lobe and central brain trachea	4K, 5F
<i>vvl-49</i>	Ventral cord midline glia	Ventral cord midline glia	Putative glia	4L, 5G,8
<i>vvl-50</i>	Negative	Subset of brain and ventral cord neurons	Brain and optic lobe neurons	5H
<i>vvl-51</i>	Ventral cord and PNS	Subset of brain and ventral cord neurons including motor neurons	Subset of Optic lobe, brain neurons and IPCs	4M, 5I, 6E
<i>vvl-52</i>	Anterior Tip	Negative	Negative	4N, 5J
<i>vvl-53</i>	Late temporal network NBs and neurons	Brain lineages including NBs and ventral cord neurons	Negative	4O
<i>vvl-54</i>	Gut ring	Subset of brain and ventral cord neurons	Subset of central Brain neurons	5K
<i>vvl-55</i>	placode cells or PNS neurons	motor neurons	Subset of central brain neurons and IPCs	4L, 6F

doi:10.1371/journal.pone.0060137.t002

the function(s) of individual CSBs within *vvl-49* are required to understand how the two halves interact to maintain full enhancer activity.

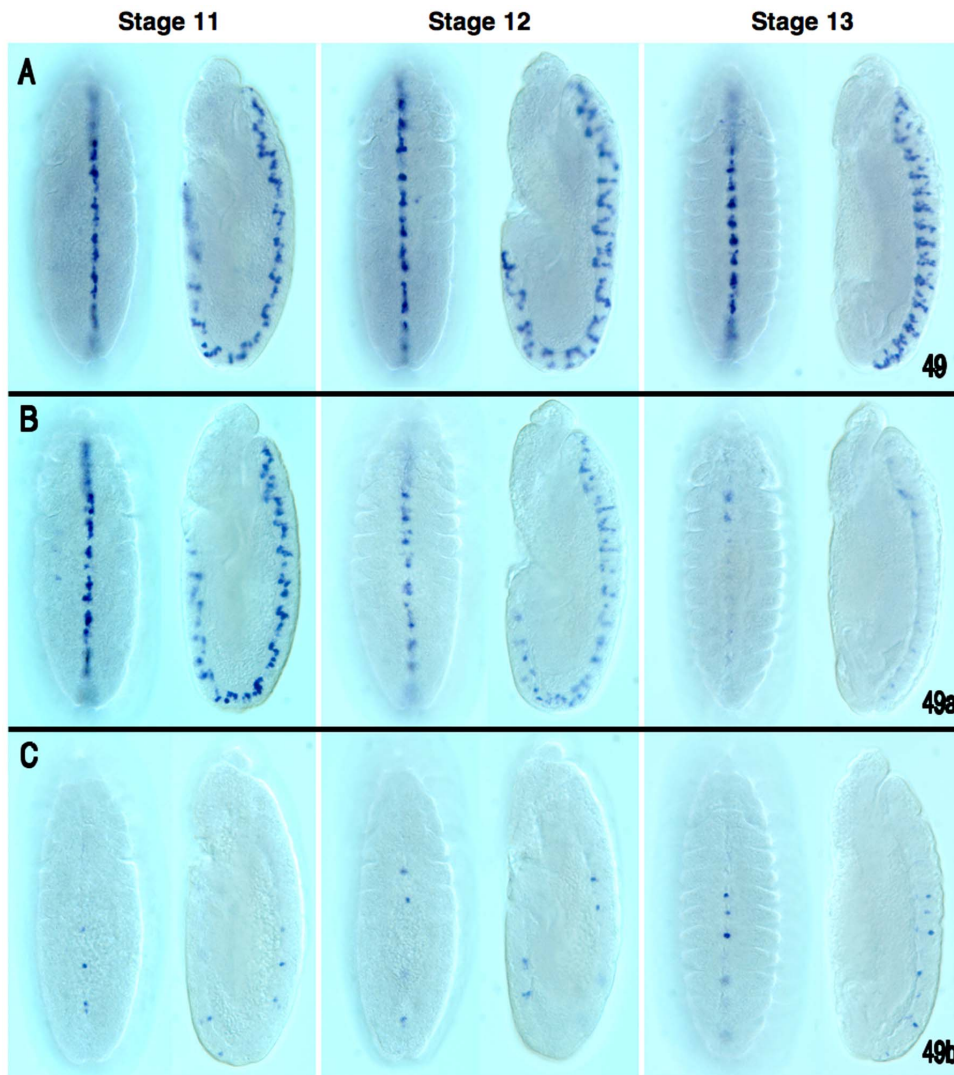
## Discussion

The principal finding of this study is that many *Drosophila* intergenic regions and large introns are home to a remarkable diversity of *cis*-regulatory enhancers. Conservation tracks that span the *Drosophila* genome demonstrate that highly conserved clusters of CSBs are not confined to sequences within or adjacent to genes (Figure 1 and [17]). *EvoPrint* analysis of gene-distant or large intron genomic expanses reveal that most of the phastcon “peaks” [19] within these regions correspond to individual conserved clusters that are separated from each other by poorly conserved DNA (Figures 1B and 2). The evolutionarily constrained spacing of CSBs within these clusters and their high degree of repeat and palindromic element coverage (that frequently exceeds 60%) are both features of characterized enhancers and strongly suggests that many of these novel sequences also participate in transcriptional regulation. Based on these observations and our earlier work on identifying repeat and palindromic elements within non-coding *Drosophila* CSB clusters [17], we estimate that the *Drosophila* genome contains over 70,000 functionally distinct enhancers, and many of these are most likely multifunctional, capable of directing gene expression during different developmental phases and in different cell types.

The use of multiple enhancers to regulate different aspects of *Drosophila* gene expression is well documented for dynamically expressed developmental determinants such as segmentation and neural genes [2–4] [43] [44]. Our survey of genomic regions surrounding other well-characterized cell-identity determinants (Table 1) combined with the functional analysis of *vvl*-associated enhancers highlights a high level of *cis*-regulatory complexity acting on developmental determinants. These results suggest that the diversity of enhancers within the tested region is not an exception, in that large numbers of clustered CSBs are also found throughout the genome. Many genes contain large introns that harbor multiple CSB clusters, including genes constituting the Hox clusters (e.g., *Antp* and *Ubx*), the neural determinants *pdm-1* and *pdm-2*, the *gsb* genes, and other genes associated with enhancer fields such as *jing*, *engrailed*, *18 w*, and *ds* (unpublished observations and Table 1). It would appear that in many cases there is a requirement for a large number of multiple independent enhancers to regulate the different spatial/temporal expression dynamics of developmentally important genes. It has been proposed that remote enhancers interact with proximal promoter sequences by enhancer/promoter tethering (see for example [25] [45]; reviewed by [9]). Without further analysis, which we believe is beyond the scope of this study, it is uncertain whether the enhancers identified in this study regulate *vvl*, *Prat2*, or other more remote genes.

During the *cis*-Decoder analysis of the different *vvl/Prat2* associated enhancers, we noticed that groups of neighboring CSBs





**Figure 8. Expression analysis of *vvl-49* midline enhancer sub-domains.** Temporal expression of *vvl-49* enhancer/reporter transgenes during embryonic development. (A) The entire *vvl-49* cluster drives reporter expression in a set of midline cells continuously from stages 11 through 13. (B) At stage 11, the *D. melanogaster vvl-49a* sub-fragment onsets expression in a subset of midline cells that is indistinguishable from that of the whole cluster, however, expression progressively declines in stage 12 and stage 13 embryos. (C) Expression of the *vvl-49b* sub-region activates expression in only a subset of the midline cells compared to the full *vvl-49* enhancer activity. doi:10.1371/journal.pone.0060137.g008

[17]. This previous search for NB enhancers identified *vvl-41* as belonging to a family of enhancers based on the shared presence of POU homeodomain and bHLH TF binding sites, often in overlapping and adjacent juxtaposition within CSBs. This current study has identified the *vvl-53* NB enhancer as an additional member of this enhancer family as it also shares conserved repeat elements with the late temporal network NB enhancers (see Text S1).

Many of the enhancers described in this study are multifunctional, in that they drive expression in two or more temporal windows or in developmentally different cells. Previous studies have described multifunctional enhancers in *Drosophila* and have dealt with the question of how gene expression patterns evolve [50]. It was suggested that novel *cis*-regulatory functions evolve by employing the hidden activities of pre-existing regulatory sequences, indicating that only a few mutations are sufficient to modulate enhancer behavior. Evolutionary flexibility of enhancer sequence is also evidenced by the high divergence of DNA sequence and TF

binding site position within the *sparkling* eye enhancer among the *Drosophila* species (reviewed by [51]). Our finding of a species-unique insertion in the *vvl-49* CSB cluster indicates some sequence flexibility between enhancer CSBs may not adversely affect their regulatory behavior; additional studies are required to understand the significance of this insertion. Some of the clusters analyzed in our study are missing one or more CSBs in the different species providing additional evidence of structural changes within these enhancers. For example, the 6th CSB (CCAAATACATAATTA) of the *vvl-43* enhancer is present in all *Drosophila* species examined in this study except for *D. willistoni* (Figure S1 and data not shown). However, the significance of the species-unique variations will only be understood by testing the effect of these changes on enhancer regulatory behavior.

The structural aspects of enhancers, including the presence multiple CSBs, their integration into super-blocks, and their content of repeat and palindromic elements, suggest that probing enhancer architecture is key to understanding the mechanism of

their *cis*-regulatory behavior. Our findings are compatible with a flexible enhanceosome model where an enhancer's *cis*-regulatory activity is responsive to different combinations of TFs that are expressed in cells undergoing different developmental programs. These studies also suggest that sequence conservation within enhancers is the norm, and that this evolutionary phenotype can be used to delimit distinct enhancers and to identify important elements necessary for their function.

## Materials and Methods

### Comparative genomic analysis

Gene-poor regions within the *Drosophila* chromosomes were identified with the UCSC genome browser (<http://genome.ucsc.edu>) [19]. Using *Drosophila melanogaster* as the reference species, analysis was carried out as described previously, using a relaxed *EvoPrint* [52]. Integrity of CSB clustering defined in this manner was tested by *EvoPrinting* homologous regions of other species, particularly of *D. willistoni*, *D. virilis*, *D. mojavensis* and *D. grimshawi*.

To reveal adjacent CSBs separated by invariant spacing, *D. melanogaster* CSB clusters were submitted for *EvoPrint* analysis. Orthologous regions of *D. willistoni*, *D. virilis*, *D. mojavensis* and *D. grimshawi* genomic sequences were curated from the *EvoPrint* scorecard. Each species' genomic sequence was subject to *EvoPrint* analysis. Pair-wise alignments of the *D. melanogaster* *EvoPrint* against the other species' *EvoPrints* were performed using the Gene-wise DNA block aligner ([53]: <http://www.ebi.ac.uk/Tools/Wise2/Dbiform.html>), which was designed to reveal colinear-conserved blocks that are flanked by non-conserved sequences of varying lengths. Common blocks shared by all pair-wise alignments were termed super-blocks.

### Enhancer-reporter transgenes

Genomic CSB clusters, detected by *EvoPrint* analysis, were amplified according to procedures described previously [4]. Primer sequences for each genomic fragment are provided in Table S1. PCR-amplified genomic fragments were inserted into the Invitrogen pCRII-TOPO vector for sequence verification. To test their *cis*-regulatory activity, fragments were transferred into a modified pCa4B site-specific integration vector termed pBullfinch-Gal4 [4] [17]. All transgenes were integrated on the 3<sup>rd</sup> chromosome at the *atp2* integration site [42]. Details of the cloning steps and vector sequence are available upon request.

### Embryo *in situ* localization of mRNA

Embryo collection and fixation were performed according to the procedures described by [54]. For *in situ* hybridization detection of reporter expression, we used the Berkeley *Drosophila* Genome project embryo *in situ* hybridization protocol (<http://www.fruitfly.org/about/methods/RNAinsitu.html>) adapted for 1.6 ml Eppendorf tubes. Gal4 mRNA expression detected by a DIG probe, generated using a Roche protocol and reagents. Staining was visualized using anti-FITC Fab fragments coupled to alkaline phosphatase. After whole-mount *in situ* hybridization, embryos were viewed in 70% glycerol/30% phosphate-buffered saline (PBS), and photographed using a Nikon microscope equipped with Nomarski (DIC) optics. Embryo developmental stages were determined by morphological criteria [55]. All details are available upon request.

### Immunohistochemistry and confocal imaging of larval and adult brains

In order to visualize CSB cluster enhancer activity in the larval and adult CNS, our cluster/GAL4 enhancer reporter lines were

crossed to the UAS-mCD8::GFP reporter line [56]. Larval CNS dissection was performed as described previously [56]; immunohistochemistry used a rabbit anti-GFP antiserum (1:1,500, Invitrogen, San Diego, CA). Confocal imaging was performed using a Zeiss LSM710 and Plan-Apochromat objective 10× (numerical aperture = 0.45). Serial optical sections (1,024×1,024 pixel resolution) were taken at 1 μm intervals along the dorso-ventral axis. The confocal image stacks were analyzed using ImageJ software (NIH, Bethesda, MD). For analysis of adult brain, at least three flies of mixed genders were collected after hatching and used for the analysis. Brain dissection and immunohistochemistry were performed as described previously [57] using a rabbit anti-GFP (1:300, Torrey Pines Biolabs, East Orange, NJ). Confocal imaging was performed using a Zeiss LSM510 META and plan Neofluar objective 40× (numerical aperture = 1.3). Serial optical sections (512×512 pixel resolution) were taken at 1 μm intervals along the rostral-caudal axis. The confocal image stacks were analyzed using Imaris (Bitplane, Zurich, Switzerland) software.

## Supporting Information

**Figure S1 Gene-distant conserved sequence clusters are made up of multiple conserved sequence blocks.** A *D. melanogaster* relaxed *EvoPrint* spanning 6.6 kb of the tested region that includes *vvl* clusters 42 through 46 (indicated by vertical bars in left margin). This genomic region is located between the *vvl* and *Prat2* genes. CSB clusters are resolved by their flanking less-conserved inter-cluster sequences. Capital letters represent bases in the *D. melanogaster* reference sequence that are conserved in all, or all but one, of the following orthologous regions within the *D. simulans*, *D. sechellia*, *D. erecta*, *D. yakuba*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. willistoni*, *D. virilis*, *D. mojavensis* and *D. grimshawi* genomes. Less or non-conserved DNA is shown as lower case gray letters. Colored highlighted sequences represent conserved repeat and/or palindromic elements discussed in the Text S1. (TIF)

**Figure S2 Gene-distant conserved sequence clusters are made up of multiple conserved sequence blocks.** A *D. melanogaster* relaxed *EvoPrint* spanning 6.6 kb of the tested region that includes *vvl* clusters 46 through 50 (indicated by vertical bars in left margin). For additional information see legend for Figure S1. (TIF)

**Figure S3 Gene-distant conserved sequence clusters are made up of multiple conserved sequence blocks.** A *D. melanogaster* relaxed *EvoPrint* spanning 6.6 kb of the tested region that includes *vvl* clusters 50 through 54 (indicated by vertical bars in left margin). For additional information see legend for Figure S1. (TIF)

**Figure S4 Gene-distant conserved sequence clusters are made up of multiple conserved sequence blocks.** A *D. melanogaster* relaxed *EvoPrint* spanning 1 kb of the tested region that includes *vvl* clusters 54 and 55 (indicated by vertical bars in left margin). For additional information see legend for Figure S1. (TIF)

**Table S1 PCR primers used to clone *vvl-37* through *vvl-55* enhancers.** (DOC)

## Text S1 *cis*-Regulatory behavior and structural analysis of 19 consecutive gene-distant CSB clusters.

(DOC)

### Acknowledgments

The authors thank Jermaine Ross and Antonios Ekatomatis for their technical assistance and Judy Brody for editorial expertise.

### References

- Fujioka M, Emi-Sarker Y, Yusibova GL, Goto T, Jaynes JB (1999) Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. *Development* 126: 2527–2538.
- Kuzin A, Kundu M, Ekatomatis A, Brody T, Odenwald WF (2009) Conserved sequence block clustering and flanking inter-cluster flexibility delineate enhancers that regulate *neftin-1* expression during *Drosophila* CNS development. *Gene Expr Patterns* 9: 65–72.
- Fujioka M, Jaynes JB (2012) Regulation of a duplicated locus: *Drosophila sloppy paired* is replete with functionally overlapping enhancers. *Dev Biol* 362: 309–319.
- Kuzin A, Kundu M, Ross J, Koizumi K, Brody T, et al. (2012) The cis-regulatory dynamics of the *Drosophila* CNS determinant *castor* are controlled by multiple sub-pattern enhancers. *Gene Expr Patterns* 12: 261–272.
- Pfeiffer BD, Jenett A, Hammonds AS, Ngo TT, Misra S, et al. (2008) Tools for neuroanatomy and neurogenetics in *Drosophila*. *Proc Natl Acad Sci U S A* 105: 9715–9720.
- Thomas S, Li XY, Sabo PJ, Sandstrom R, Thurman RE, et al. (2011) Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. *Genome Biol* 12: R43.
- Kvon EZ, Stampfel G, Yanez-Cuna JO, Dickson BJ, Stark A (2012) HOT regions function as patterned developmental enhancers and have a distinct cis-regulatory signature. *Genes Dev* 26: 908–913.
- Bachmann A, Knust E (2008) The use of P-element transposons to generate transgenic flies. *Methods Mol Biol* 420: 61–77.
- Spitz F, Furlong EE (2012) Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* 13: 613–626.
- Odenwald WF, Rasband W, Kuzin A, Brody T (2005) EVOPRINTER, a multigenomic comparative tool for rapid identification of functionally important DNA. *Proc Natl Acad Sci U S A* 102: 14700–14705.
- Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, et al. (2006) In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 444: 499–502.
- Brody T, Rasband W, Baler K, Kuzin A, Kundu M, et al. (2007) *cis*-Decoder discovers constellations of conserved DNA sequences shared among tissue-specific enhancers. *Genome Biol* 8: R75.
- Loots G, Ovcharenko I (2007) ECRbase: database of evolutionary conserved regions, promoters, and transcription factor binding sites in vertebrate genomes. *Bioinformatics* 23: 122–124.
- Hardison RC (2000) Conserved noncoding sequences are reliable guides to regulatory elements. *Trends Genet* 16: 369–372.
- Kuo CL, Chen ML, Wang K, Chou CK, Vernooij B, et al. (1998) A conserved sequence block in murine and human T cell receptor (TCR)  $\alpha$  region is a composite element that enhances TCR  $\alpha$  enhancer activity and binds multiple nuclear factors. *Proc Natl Acad Sci U S A* 95: 3839–3844.
- Chen ML, Kuo CL (2001) A conserved sequence block in the murine and human T cell receptor  $\alpha$  loci interacts with developmentally regulated nucleoprotein complexes in vitro and associates with GATA-3 and octamer-binding factors in vivo. *Eur J Immunol* 31: 1696–1705.
- Brody T, Yavatkar AS, Kuzin A, Kundu M, Tyson LJ, et al. (2012) Use of a *Drosophila* genome-wide conserved sequence database to identify functionally related cis-regulatory enhancers. *Dev Dyn* 241: 169–189.
- Hirono K, Margolis JS, Posakony JW, Doe CQ (2012) Identification of hunchback cis-regulatory DNA conferring temporal expression in neuroblasts and neurons. *Gene Expr Patterns* 12: 11–17.
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15: 1034–1050.
- Nobrega MA, Ovcharenko I, Afzal V, Rubin EM (2003) Scanning human gene deserts for long-range enhancers. *Science* 302: 413.
- Ovcharenko I, Loots GG, Nobrega MA, Hardison RC, Miller W, et al. (2005) Evolution and functional classification of vertebrate gene deserts. *Genome Res* 15: 137–145.
- de la Calle-Mustienes E, Feijoo CG, Manzanares M, Tena JJ, Rodriguez-Seguel E, et al. (2005) A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts. *Genome Res* 15: 1061–1072.
- Maeso I, Irimia M, Tena JJ, Gonzalez-Perez E, Tran D, et al. (2012) An ancient genomic regulatory block conserved across bilaterians and its dismantling in tetrapods by retrogene replacement. *Genome Res* 22: 642–655.
- Jack J, Dorsett D, Delotto Y, Liu S (1991) Expression of the *cut* locus in the *Drosophila* wing margin is required for cell type specification and is regulated by a distant enhancer. *Development* 113: 735–747.
- Kwon D, Mucci D, Langlais KK, Americo JL, DeVido SK, et al. (2009) Enhancer-promoter communication at the *Drosophila engrailed* locus. *Development* 136: 3067–3075.
- Frankel N, Davis GK, Vargas D, Wang S, Payne F, et al. (2010) Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* 466: 490–493.
- Sotillos S, Espinosa-Vazquez JM, Foglia F, Hu N, Hombria JC (2010) An efficient approach to isolate STAT regulated enhancers uncovers STAT92E fundamental role in *Drosophila* tracheal development. *Dev Biol* 340: 571–582.
- Anderson MG, Perkins GL, Chittick P, Shrigley RJ, Johnson WA (1995) drifter, a *Drosophila* POU-domain transcription factor, is required for correct differentiation and migration of tracheal cells and midline glia. *Genes Dev* 9: 123–137.
- Certel SJ, Thor S (2004) Specification of *Drosophila* motoneuron identity by the combinatorial action of POU and LIM-HD factors. *Development* 131: 5429–5439.
- Inbal A, Levanon D, Salzberg A (2003) Multiple roles for *u-turn/ventral veinless* in the development of *Drosophila* PNS. *Development* 130: 2467–2478.
- Meier S, Sprecher SG, Reichert H, Hirth F (2006) ventral veins lacking is required for specification of the tritocerebrum in embryonic brain development of *Drosophila*. *Mech Dev* 123: 76–83.
- Miller SW, Avidor-Reiss T, Polyanovsky A, Posakony JW (2009) Complex interplay of three transcription factors in controlling the tormogen differentiation program of *Drosophila* mechanoreceptors. *Dev Biol* 329: 386–399.
- Komiyama T, Johnson WA, Luo L, Jefferis GS (2003) From lineage to wiring specificity. POU domain transcription factors control precise connections of *Drosophila* olfactory projection neurons. *Cell* 112: 157–167.
- Komiyama T, Sweeney LB, Schuldiner O, Garcia KC, Luo L (2007) Graded expression of semaphorin-1a cell-autonomously directs dendritic targeting of olfactory projection neurons. *Cell* 128: 399–410.
- Hasegawa E, Kitada Y, Kaido M, Takayama R, Awasaki T, et al. (2011) Concentric zones, cell migration and neuronal circuits in the *Drosophila* visual center. *Development* 138: 983–993.
- Certel K, Hudson A, Carroll SB, Johnson WA (2000) Restricted patterning of vestigial expression in *Drosophila* wing imaginal discs requires synergistic activation by both Mad and the drifter POU domain transcription factor. *Development* 127: 3173–3183.
- Kambadur R, Koizumi K, Stivers C, Nagle J, Poole SJ, et al. (1998) Regulation of POU genes by *castor* and *hunchback* establishes layered compartments in the *Drosophila* CNS. *Genes Dev* 12: 246–260.
- Tomancak P, Beaton A, Weiszmam R, Kwan E, Shu S, et al. (2002) Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol* 3: RESEARCH0088.
- Ji Y, Clark DV (2006) The purine synthesis gene *Prat2* is required for *Drosophila* metamorphosis, as revealed by inverted-repeat-mediated RNA interference. *Genetics* 172: 1621–1631.
- Kuzin A, Kundu M, Brody T, Odenwald WF (2011) Functional analysis of conserved sequences within a temporally restricted neural precursor cell enhancer. *Mech Dev* 128: 165–177.
- Groth AC, Fish M, Nusse R, Calos MP (2004) Construction of transgenic *Drosophila* by using the site-specific integrase from phage  $\phi$ C31. *Genetics* 166: 1775–1782.
- Markstein M, Pitsouli C, Villalta C, Celniker SE, Perrimon N (2008) Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat Genet* 40: 476–483.
- Perry MW, Boettiger AN, Levine M (2011) Multiple enhancers ensure precision of gap gene-expression patterns in the *Drosophila* embryo. *Proc Natl Acad Sci U S A* 108: 13570–13575.
- Bejerano G, Siepel AC, Kent WJ, Haussler D (2005) Computational screening of conserved genomic DNA in search of functional noncoding elements. *Nat Methods* 2: 535–545.
- Ho MC, Schiller BJ, Akbari OS, Bac E, Drewell RA (2011) Disruption of the abdominal-B promoter tethering element results in a loss of long-range enhancer-directed *Hox* gene expression in *Drosophila*. *PLoS One* 6: e16283.
- Liu F, Posakony JW (2012) Role of architecture in the function and specificity of two Notch-regulated transcriptional enhancer modules. *PLoS Genet* 8: e1002796.

## Author Contributions

Conceived and designed the experiments: AK TB WFO. Performed the experiments: MK AK TYL. Analyzed the data: TB CHL WFO. Wrote the paper: TB WFO.

47. Papatsenko D, Goltsev Y, Levine M (2009) Organization of developmental enhancers in the *Drosophila* embryo. *Nucleic Acids Res* 37: 5665–5677.
48. Panne D (2008) The enhancosome. *Curr Opin Struct Biol* 18: 236–242.
49. Miele L (2011) Transcription factor RBPJ/CSL: a genome-wide look at transcriptional regulation. *Proc Natl Acad Sci U S A* 108: 14715–14716.
50. Rebeiz M, Jikomes N, Kassner VA, Carroll SB (2011) Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. *Proc Natl Acad Sci U S A* 108: 10036–10043.
51. Evans NC, Swanson CI, Barolo S (2012) Sparkling insights into enhancer structure, function, and evolution. *Curr Top Dev Biol* 98: 97–120.
52. Yavatkar AS, Lin Y, Ross J, Fann Y, Brody T, et al. (2008) Rapid detection and curation of conserved DNA via *enhanced-BLAT* and *EvoPrinterHD* analysis. *BMC Genomics* 9: 106.
53. Jareborg N, Birney E, Durbin R (1999) Comparative analysis of noncoding regions of 77 orthologous mouse and human gene pairs. *Genome Res* 9: 815–824.
54. Patel NH (1994) Imaging neuronal subsets and other cell types in whole-mount *Drosophila* embryos and larvae using antibody probes. *Methods Cell Biol* 44: 445–487.
55. Campos-Ortega JA, Hartenstein V (1985) *The Embryonic Development of Drosophila melanogaster*. Springer-Verlag, Berlin.
56. Lee T, Luo L (1999) Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* 22: 451–461.
57. Gao S, Takemura SY, Ting CY, Huang S, Lu Z, et al. (2008) The neural substrate of spectral preference in *Drosophila*. *Neuron* 60: 328–342.
58. Hartenstein VC, Campos-Ortega J (1984) Early neurogenesis in wildtype *Drosophila melanogaster*. *Roux's Arch Dev Biol* 193: 308–325.
59. Luo J, Becnel J, Nichols CD, Nassel DR (2012) Insulin-producing cells in the brain of adult *Drosophila* are regulated by the serotonin 5-HT1A receptor. *Cell Mol Life Sci* 69: 471–484.