The Density of Regulatory Information Is a Major Determinant of Evolutionary Constraint on Noncoding DNA in *Drosophila*

Gonzalo Sabarís [1,2,†] Daniela M. Ortíz,[1,†] Ian Laiker,[1,†] Ignacio Mayansky,[1] Sujay Naik,[3] Giacomo Cavalli [2] David L. Stern [4] Ella Preger-Ben Noon,[3] and Nicolás Frankel [1,5,*]

[1]Instituto de Fisiología, Biología Molecular y Neurociencias (IFIBYNE), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad de Buenos Aires (UBA), Buenos Aires 1428, Argentina

[2]Institute of Human Genetics, UMR 9002 CNRS-Université de Montpellier, Montpellier, France

[3]Department of Genetics and Developmental Biology, The Rappaport Faculty of Medicine and Research Institute, Technion—Israel Institute of Technology, Haifa 3109601, Israel

[4]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA

[5]Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales (FCEN), Universidad de Buenos Aires (UBA), Buenos Aires 1428, Argentina

[†]These authors contributed equally.

***Corresponding author:** E-mail: nfrankel@conicet.gov.ar.

Associate editor: Ilya Ruvinsky

Abstract

Evolutionary analyses have estimated that ∼60% of nucleotides in intergenic regions of the *Drosophila melanogaster* genome are functionally relevant, suggesting that regulatory information may be encoded more densely in intergenic regions than has been revealed by most functional dissections of regulatory DNA. Here, we approached this issue through a functional dissection of the regulatory region of the gene *shavenbaby* (*svb*). Most of the ∼90 kb of this large regulatory region is highly conserved in the genus *Drosophila*, though characterized enhancers occupy a small fraction of this region. By analyzing the regulation of *svb* in different contexts of *Drosophila* development, we found that the regulatory information that drives *svb* expression in the abdominal pupal epidermis is organized in a different way than the elements that drive *svb* expression in the embryonic epidermis. While in the embryonic epidermis *svb* is activated by compact enhancers separated by large inactive DNA regions, *svb* expression in the pupal epidermis is driven by regulatory information distributed over broader regions of *svb cis*-regulatory DNA. In the same vein, we observed that other developmental genes also display a dense distribution of putative regulatory elements in their regulatory regions. Furthermore, we found that a large percentage of conserved noncoding DNA of the *Drosophila* genome is contained within regions of open chromatin. These results suggest that part of the evolutionary constraint on noncoding DNA of *Drosophila* is explained by the density of regulatory information, which may be greater than previously appreciated.

***Key words: Drosophila*, noncoding DNA, density of regulatory elements, *shavenbaby*, evolutionary constraint, transcriptional enhancers.**

Introduction

Genomic regions that encode the information for gene regulation have been studied intensely for decades (Davidson 2010; Schaffner 2015; Moore et al. 2020). A clear picture has emerged from these analyses, where gene expression is controlled through the combinatorial binding of transcription factors to regulatory elements such as transcriptional enhancers (Zinzen et al. 2009). Enhancers contain arrangements of transcription factor binding sites that promote transcription at defined spatiotemporal patterns (Levine 2010). Decades of work have shown that enhancers play an important role in evolutionary change (Carroll 2008; Hill et al. 2021) and that their malfunction can cause disease (Claringbould and Zaugg 2021).

The fact that enhancer regions smaller than 1-kb cloned upstream of a core promoter and a reporter gene can recapitulate somewhat faithfully the endogenous expression pattern of a gene has led to the common belief that enhancers are compact regulatory elements. Thus, in the literature, enhancers are often described as compact regulatory elements (Levine 2010; Long et al. 2016; Panigrahi and O'Malley 2021). Nonetheless, some regulatory

Received: July 04, 2023. **Revised:** November 26, 2023. **Accepted:** January 05, 2024

information in vertebrate genomes may be distributed over regions larger than 10 kb, forming so-called super-enhancers (Pott and Lieb 2014).

In *Drosophila*, a few *cis*-regulatory regions (noncoding regions flanking coding DNA) have been dissected in detail (Small and Arnosti 2020). These studies have identified compact regulatory elements. For example, the pattern of 7 stripes of the pair-rule gene *even-skipped* (*eve*) in the early embryo of *Drosophila melanogaster* is determined by several ∼500-bp "minimal enhancers" (i.e. the shortest pieces of DNA that can drive an expression pattern that resembles the endogenous pattern of a gene), each directing expression in 1 or 2 stripes (Fujioka et al. 1999; Sackerson et al. 1999). Results such as these imply that the regulatory information that constitutes an enhancer is confined to a compact DNA region with autonomous function.

In contrast, several lines of evidence suggest that there is important regulatory information in DNA flanking minimal enhancers in *Drosophila*. For example, despite the early work defining minimal enhancers of the *eve* locus, a 16-kb reporter construct of the *eve* locus with a deletion of the minimal stripe 2 element (the enhancer that generates the second of the 7 stripes) retains residual expression at stripe 2 cells (Ludwig et al. 2005). In addition, it was shown that the ∼200 bp that flank the stripe 2 minimal enhancer buffer gene expression under environmental or genetic perturbations (Ludwig et al. 2011; López-Rivera et al. 2020). Furthermore, in *Drosophila erecta*, a species closely related to *D. melanogaster*, DNA outside of the minimal stripe 2 enhancer (tested in an extended version of stripe 2 enhancer of ∼1.6 kb) is needed for accurate expression of *eve* in stripe 2 (Crocker and Stern 2017). In addition, a quantitative analysis of transcriptional activation of the *eve* locus using reporter constructs of different sizes showed that there is regulatory information outside minimal enhancers, which is necessary for creating the native stripe 2 and stripe 7 patterns in the embryo (Halfon 2006; Janssens et al. 2006).

In addition, some *cis*-regulatory regions in the *Drosophila* genome contain enhancers that are not compact. Six of the 7 stripes of the pair-rule gene *hairy* are generated by seemingly compact enhancers, but regulatory information for driving *hairy* stripe 2 is spread over many kilobases (Riddihough and Ish-Horowicz 1991). Similarly, several aspects of the expression of the pair-rule gene *runt* in *D. melanogaster* appear to be mediated by regulatory information scattered over many kilobases (Klingler et al. 1996). The possibility that regulatory information may be spread over kilobases linked to the possibility that a coherent expression pattern might only emerge upon the interaction of scattered subelements (Frankel 2012) could explain the inability to isolate discrete enhancers from some large regulatory regions (Davis et al. 2007).

Evolutionary analyses have estimated that the fraction of functionally relevant nucleotides (nucleotides that evolve under either negative or positive selection and thus have a function) in intergenic regions of the *D. melanogaster* genome is ∼0.6 (Andolfatto 2005; Halligan and Keightley 2006). This estimate suggests that most nucleotides in *cis*-regulatory regions are functional. But how can we explain such a large fraction of functional nucleotides if regulatory elements occupy small segments of noncoding DNA? Is the estimated fraction of functional nucleotides inflated by neutral sites linked to selected sites (Leffler et al. 2012)? Is this noncoding DNA mostly related to functions other than controlling gene expression (such as chromosome replication, chromatin structure, and noncoding RNAs)? How much of the noncoding constraint within the *Drosophila* genome is explained by the presence of regulatory DNA?

Here, we explore this issue by examining the distribution of conserved elements and the function of noncoding DNA in the regulatory region of the *shavenbaby* (*svb*) gene. Svb is a transcription factor that controls the formation of nonsensory cuticular hairs (trichomes) in the larva (Payre et al. 1999) and pupa (Delon et al. 2003; Preger-Ben Noon et al. 2018) of *D. melanogaster*. The *svb* regulatory region has been studied for decades, providing a solid platform for exploring mechanistic and evolutionary aspects of gene regulation (Frankel et al. 2012; Stern and Frankel 2013; Kittelmann et al. 2021; Soverna et al. 2021). Prior comprehensive analyses of the ∼90-kb region upstream of the *svb* first exon using reporter constructs revealed that regulatory activity in the embryo is limited to 7 enhancers, some of which have been dissected to fragments shorter than 1 kb (McGregor et al. 2007; Frankel et al. 2010; Frankel et al. 2011; Crocker et al. 2015; Preger-Ben Noon et al. 2016, 2018). These 7 enhancers are also active in the epidermis of the pupal abdomen and other larval tissues (Preger-Ben Noon et al. 2018).

Here, we verified that regions outside of the known enhancer elements display high sequence conservation (Stern and Frankel 2013). We reasoned that there might be additional regulatory information in the locus controlling *svb* expression in contexts other than the embryo. Thus, we undertook a functional characterization of the *cis*-regulatory region of *svb* in the epidermis of the pupal abdomen. As opposed to achieving *svb* activation through compact enhancers separated by large inactive DNA regions, we found that *svb* expression in the abdominal pupal epidermis results from the activity of large DNA regions with enhancer activity, which occupy most of the ∼90-kb upstream of *svb* first exon.

To assess the generality of our finding, we examined regulatory regions of other developmental genes. We observed that other developmental genes also display dense arrays of open-chromatin regions in their flanking noncoding DNA and that a large fraction of conserved bases lie within these putative regulatory elements. We hypothesized that this pattern might be extrapolated to the whole genome and therefore performed a genome-wide analysis. We found that a large fraction of conserved noncoding DNA throughout the genome is contained within open-chromatin regions. Overall, these results suggest that the widespread conservation of *Drosophila* noncoding DNA can be explained, at least in part, by the distribution of regulatory elements throughout the noncoding genome.

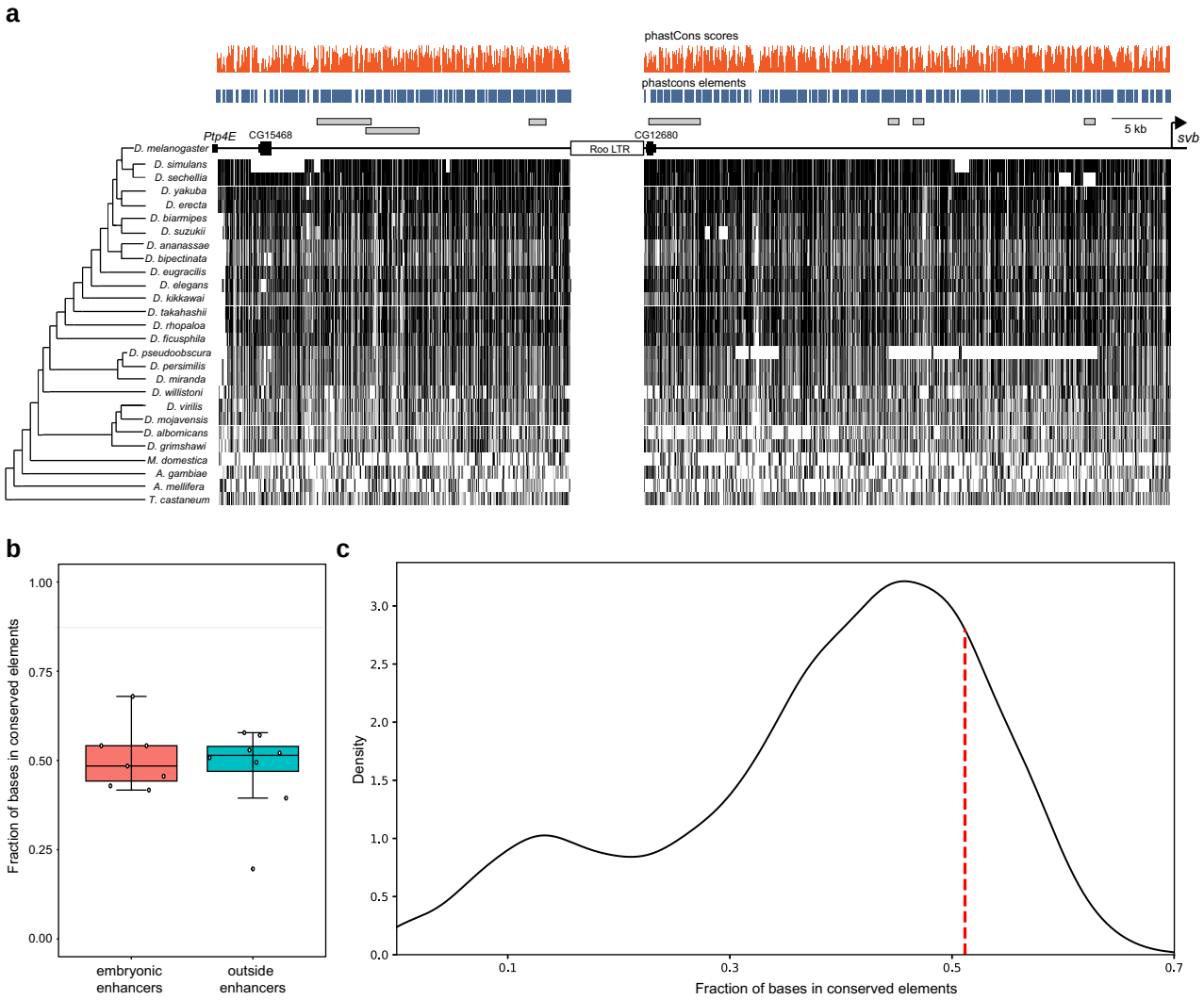

Fig. 1. Widespread sequence conservation in the regulatory region of the *svb* gene. a) Scheme of the *svb* *cis*-regulatory region of *D. melanogaster*, showing the position of the embryonic enhancers (gray boxes) and the pattern of sequence conservation in a multiple sequence alignment of 23 *Drosophila* species and 4 outgroup species. Orange peaks represent phastCons conservation scores per base. Blue boxes represent conserved elements predicted by phastCons. In *D. melanogaster*, the regulatory region of *svb* contains a species-specific transposable element (Roo LTR) (b) Boxplots with the fraction of bases in conserved elements in the 7 enhancers (left) and 8 regions that lie outside enhancers (right). Empty circles indicate values for each DNA fragment. The differences between categories are not significant (Mann–Whitney test, $P = 0.61$). (c) Density plot showing sequence conservation in different noncoding regions of the *D. melanogaster* genome. The *x* axis indicates the fraction of bases within phastCons conserved elements in 10,000 noncoding windows of the *D. melanogaster* genome. The red dashed line indicates the fraction of conserved bases for the *svb* *cis*-regulatory region (0.511).

Results

The Regulatory Region of the *svb* Gene Displays Widespread Sequence Conservation

We examined sequence conservation of the *svb* *cis*-regulatory region and other noncoding genomic regions of similar size using multispecies genome alignments (see Materials and Methods for details) and quantified conservation using the phastCons program (Siepel et al. 2005), which is based on a 2-state phylogenetic hidden Markov model. phastCons uses a multiple alignment and a phylogenetic model to estimate per-base probabilities of negative selection, enabling the prediction of conserved elements as contiguous bases with high probability of being under negative selection. First, we calculated phastCons values throughout the *svb* *cis*-regulatory region (Fig. 1a). We observed similar fractions of bases in phastCons conserved elements in the 7 known enhancers and the regions that lie outside enhancers (Fig. 1b; supplementary table S1, Supplementary Material online). Second, we found that the fraction of bases in phastCons conserved elements in the whole *cis*-regulatory region of *svb* is greater than in 80% of 10,000 randomly chosen windows of noncoding DNA of the same size in the *D. melanogaster* genome (Fig. 1c). We hypothesized that the high sequence conservation could indicate the existence of functional noncoding DNA, so we embarked on a search for additional regulatory elements. Since we had previously identified all enhancers in the ~90-kb upstream of *svb* first exon that drive

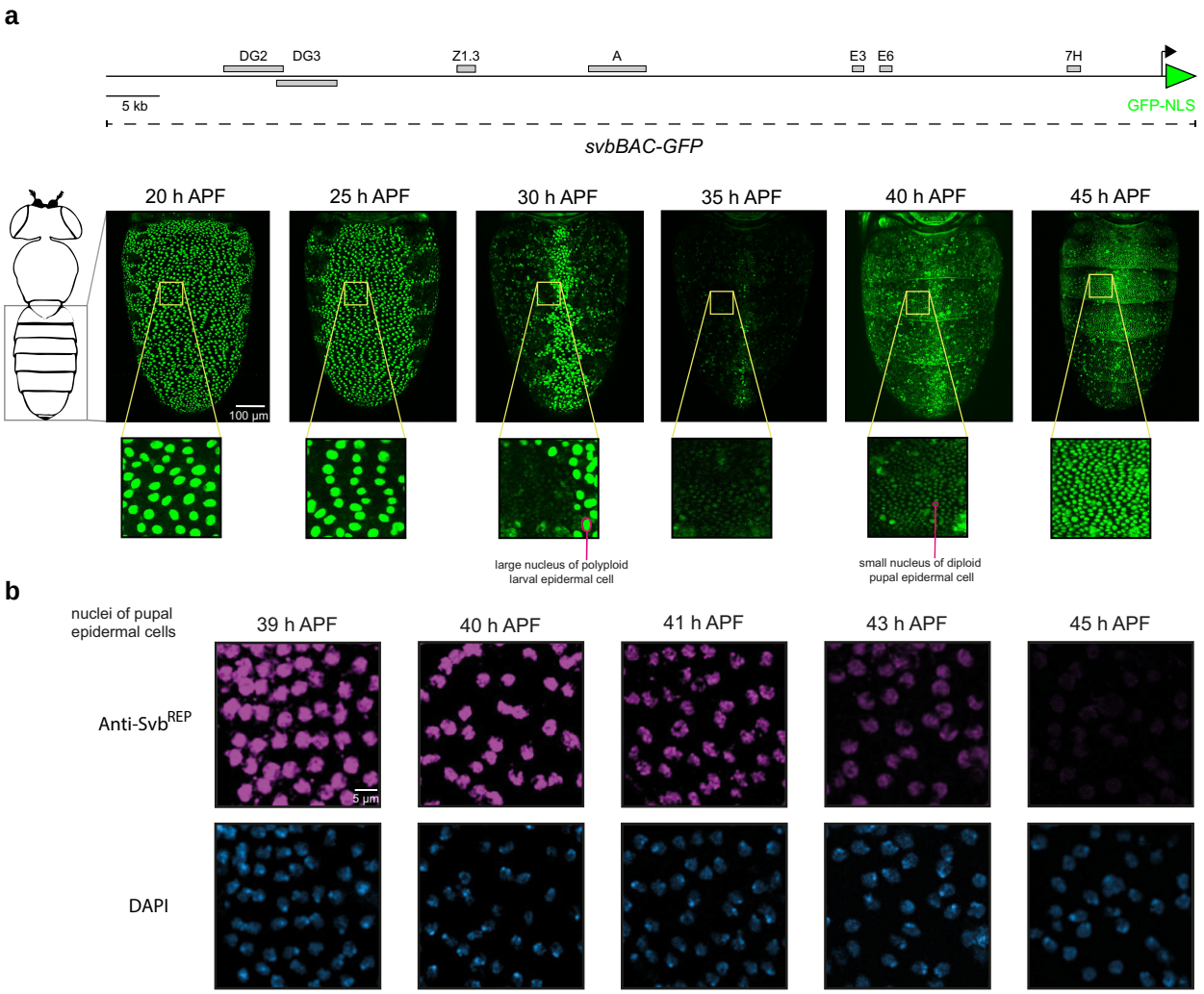

Fig. 2. *Svb* expression and Svb activation in the abdominal pupal epidermis of *D. melanogaster*. a) GFP expression driven by the *svbBAC*-GFP reporter construct (scheme above) in nuclei of larval epidermal cells (large nuclei) and nuclei of pupal epidermal cells (small nuclei) of the pupal abdomen between 20 and 45 h APF. b) The presence of the full-length Svb (transcriptional repressor) in nuclei of the abdominal pupal epidermis between 39 and 45 h APF (magenta). Images below show the DAPI signal (blue) for the same fields.

svb expression in the embryonic epidermis (McGregor et al. 2007; Frankel et al. 2010), we decided to characterize *svb* expression in the pupal abdominal epidermis, another tissue in which Svb is required for trichome production (Preger-Ben Noon et al. 2018).

Svb Is Active in the Abdominal Epidermis of Pupae at ~40 to 45 hAPF

To investigate *svb* expression in the pupal epidermis, we examined expression driven by a bacterial artificial chromosome (BAC) integrated into the *D. melanogaster* genome that carries the *svb* upstream region and a GFP-NLS reporter, named *svb*BAC-GFP (Fig. 2a; Preger-Ben Noon et al. 2018). We observed GFP in larval epidermal cells at 20 h after puparium formation (hAPF) (Fig. 2a), which is consistent with earlier *svb* expression linked to the formation of the puparium. The regulatory elements necessary for driving this larval expression pattern are currently unknown. The GFP signal observed in the nuclei of these large polyploid cells disappears during the histolysis of these cells (Fig. 2a). The pupal abdominal epidermis is derived from histoblast nest cells, which divide and migrate in the early pupa. These diploid cells, which are much smaller than larval epidermal cells, replace the larval epidermis across the whole abdomen. We first detected GFP expression in pupal epidermal cells at 35 hAPF (Fig. 2a). Later, GFP levels increase, and by 45 hAPF, all pupal abdominal epidermal cells display bright GFP signal.

To determine whether the patterns of GFP transcription driven by the *svb*BAC-GFP reflect the expression of Svb protein, we stained the abdominal epidermis with an antibody that recognizes the N-terminus of Svb (Chanut-Delalande et al. 2014). Full-length Svb protein acts as a transcriptional repressor and is converted into a transcriptional activator upon degradation of its N-terminus (Kondo et al. 2010). We detected Svb repressor in the epidermis of the pupal abdomen at 39 hAPF (Fig. 2b). Svb is present in all cells of the pupal abdominal epidermis, which coincides with the pattern of GFP expression reported by *svb*BAC-GFP (data not shown). Starting at approximately 41 hAPF, the intensity

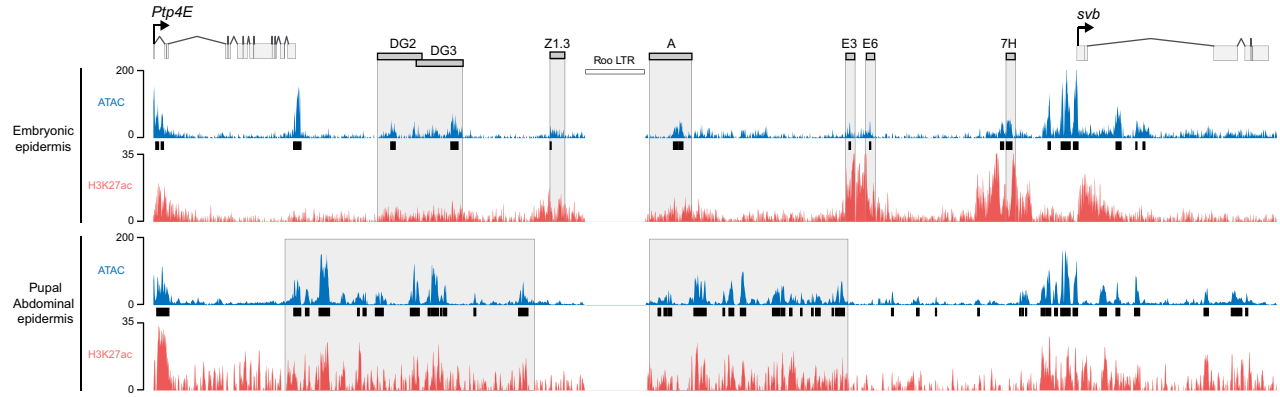

Fig. 3. Contrasting chromatin landscapes in the *svb* regulatory region between embryonic and pupal epidermis. Chromatin landscape of the *svb* locus in the embryonic epidermis (above) and the pupal abdominal epidermis (below). ATAC-seq profiles are colored in blue and ATAC-seq peaks are indicated with black boxes below profiles. H3K27ac profiles are colored in red. Embryonic enhancers of *svb* are represented with gray boxes. Gray shading marks the position of embryonic enhancers in embryonic tracks. In pupal tracks, gray shading marks the location of regions with high density of ATAC-seq peaks and H3K27ac signal.

of Svb repressor staining decreases, until it becomes almost undetectable at 45 hAPF (Fig. 2b). These results suggest that Svb is converted into a transcriptional activator between 40 and 45 hAPF, which implies that Svb target genes are activated during this interval, triggering trichome development. Although we do not provide direct evidence of the maturation of Svb into an activator, our data are consistent with a previous study that identified the beginning of trichome formation in the abdominal epidermis at approximately 45 hAPF (Mangione and Martín-Blanco 2018).

Chromatin Landscapes of the *svb* Regulatory Region in the Embryo and Pupa Are Sharply Different

To characterize the regulatory landscapes of *svb* in the epidermis of the embryo and pupal abdomen, we assayed open chromatin and the presence of the histone mark H3K27ac, which together provide evidence for the existence of active enhancers (Moore et al. 2020). We performed fluorescence-activated cell sorting (FACS) to isolate *svb*-expressing epidermal cells from late embryos and performed chromatin immunoprecipitation (ChIP)-seq to quantify the genome-wide H3K27ac signal in these cells. In addition, we characterized open-chromatin regions by retrieving computationally defined clusters of single-cell assay for transposase-accessible chromatin with sequencing (scATAC-seq) data corresponding to epidermal cells of the late embryo (Cusanovich et al. 2018). We analyzed these data in the context of previous findings, which identified 7 enhancers scattered across the *svb* regulatory region that together can fully recapitulate the native *svb* expression pattern in the embryo (in these experiments, the entire ~90-kb upstream of *svb* was tested for embryonic enhancers through reporter constructs). We observed that genomic regions showing high levels of open chromatin and H3K27ac enrichment are largely coincident with the locations of the 7 previously characterized embryonic enhancers (Fig. 3). In fact, there is a consistent overlap of ATAC peaks and flanking acetylation signals specifically for enhancer regions that have been dissected to minimal elements (Z1.3, E3, E6, and 7H; Fig. 3). ATAC-seq peaks and acetylation signals within enhancers that have not been dissected to small elements (DG2, DG3, and A) are potential predictors of the position of minimal elements (Fig. 3). Thus, the chromatin landscape of the *svb* locus in epidermal cells of the embryo reveals that embryonic enhancers are small islands in a large regulatory region, which is consistent with prior knowledge from reporter assays.

To characterize the chromatin landscape of the *svb* locus in epidermal cells of the pupal abdomen, we dissected the epidermis of the pupal abdomen between 38 and 45 hAPF and performed ATAC-seq and CUT&RUN for the H3K27ac mark. Strikingly, we observed large areas of the *svb* upstream region displaying ATAC-seq peaks and elevated H3K27ac signal (Fig. 3). Notably, pupal ATAC peaks do not overlap with embryonic ATAC peaks (Fig. 3). The fact that chromatin of small embryonic enhancers (Z1.3, E3, E6, and 7H) appears closed in the pupal epidermis is a puzzling result, because these enhancers were shown to drive reporter expression in the epidermis of the pupa and because the deletion of 7H, E6, and Z1.3 in *svb*BAC-GFP alters GFP expression in the epidermis of the pupa (Preger-Ben Noon et al. 2018). It remains possible that these elements function with positioned nucleosomes. In sum, these data show that the chromatin landscapes of the *svb* regulatory region in epidermal cells of the embryo and epidermal cells of the pupa are different. Regulatory DNA driving pupal epidermal expression is spread across large regions that are not active in the embryo. Furthermore, open-chromatin data from other tissues in which *svb* is expressed suggest that the regulatory landscape of this gene might be highly variable between contexts (supplementary fig. S1, Supplementary Material online). Given that the chromatin structure in epidermal cells of the pupal abdomen suggests that the *svb cis*-regulatory region contains multiple enhancers for

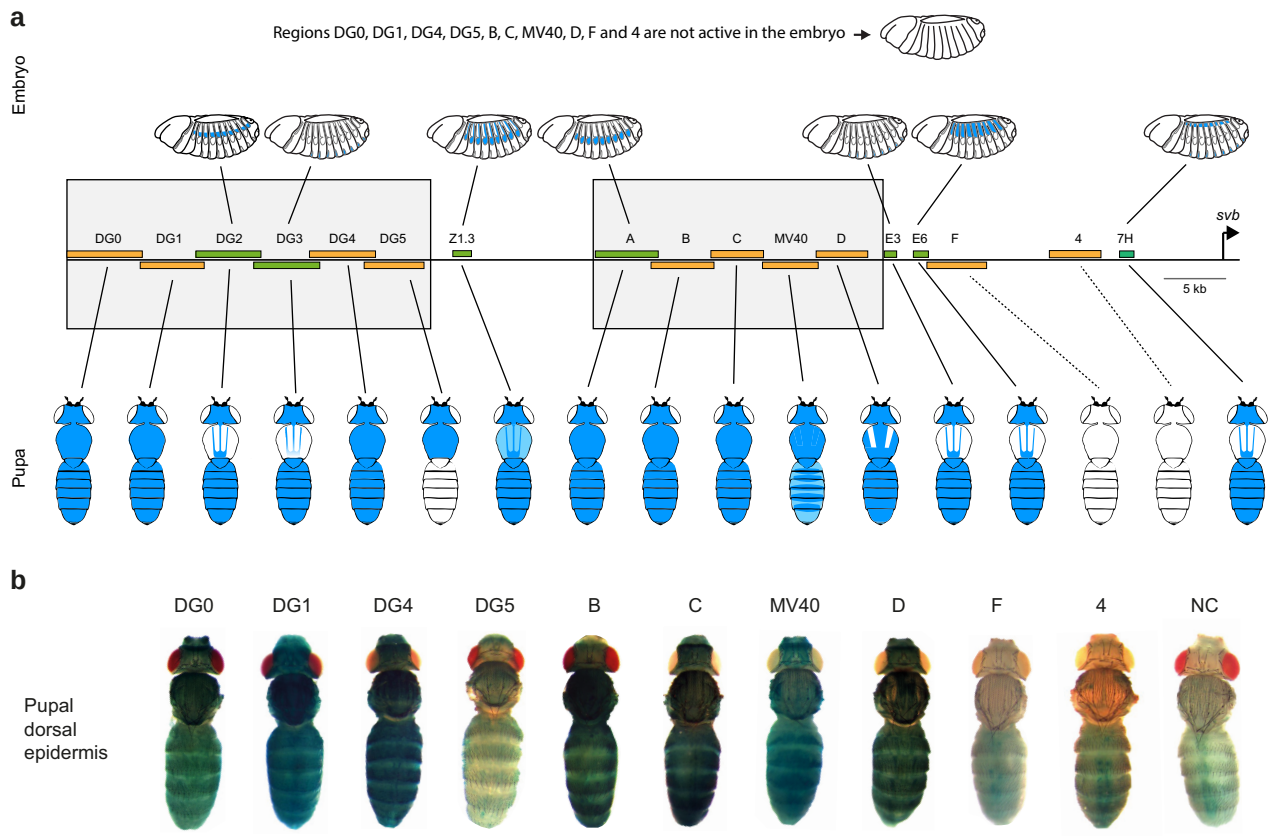

Fig. 4. Reporter constructs within the candidate regulatory regions have enhancer activity in the pupa. a) Schematic of expression patterns driven by *svb* reporter constructs in the embryo and pupal epidermis (the panel integrates expression data obtained in McGregor et al. 2007; Frankel et al. 2010; Preger-Ben Noon et al. 2018; and this study). The diagram of the *svb* locus shows the position of the 7 enhancers that are active in the embryo, larva, and pupa (green boxes) and 10 fragments without enhancer activity in the embryo (orange boxes) that were examined for pupal enhancer activity in this study. The position of candidate regulatory regions is indicated with gray boxes. b) X-Gal staining of whole pupae (dorsal view) for the 10 fragments without enhancer activity in the embryo (orange boxes) and a negative control (NC).

the pupal epidermis, we sought to further validate these candidate enhancer regions with reporter constructs.

Reporter Gene Assays Validate Novel Pupal and Larval Enhancers of the *svb* Gene

Previously, we found that the 7 *svb* enhancers with embryonic activity are active also in the pupal epidermis (Preger-Ben Noon et al. 2018). Here, we examined reporter constructs that encompass most of the remaining *svb* upstream region (Fig. 4a). We analyzed the activity of these fragments, which are cloned upstream of the *lacZ* reporter gene, in whole pupae. We observed that all 8 fragments within candidate enhancer regions in the pupa (fragments DG0, DG1, DG4, DG5, B, C, D, and MV40) drove reporter expression in the pupal abdominal epidermis (Fig. 4a and b; supplementary fig. S2, Supplementary Material online). All of these fragments are also active in the epidermis of the thorax and head (Fig. 4; supplementary fig. S2, Supplementary Material online). Two fragments that lie outside of the candidate enhancer regions (fragments F and 4) did not display enhancer activity in the pupal epidermis (Fig. 4a and b; supplementary fig. S2, Supplementary Material online). Altogether, these analyses confirm that multiple DNA regions encode regulatory information capable of driving reporter expression in the pupal epidermis.

In an earlier study, we demonstrated that 7 *svb* enhancers (DG2, DG3, A, Z1.3, E3, E6, and 7H) drive expression in the embryo, larva, and pupa (Preger-Ben Noon et al. 2018). We therefore wondered whether the newly characterized fragments with pupal expression may also be active in other developmental contexts. Preliminary evidence from reporter constructs suggested that *svb* is expressed in the brain, imaginal discs, and foregut of the larva (Preger-Ben Noon et al. 2018), but the presence of the protein in these organs had not been confirmed. We used the antibody against the N-terminus of the Svb protein and confirmed that Svb is indeed present in the brain, imaginal discs, and foregut of the larva (supplementary fig. S3, Supplementary Material online). We tested the activity of reporter constructs in these organs and in the epidermis of third instar larvae (supplementary fig. S4, Supplementary Material online). We observed that 7 of the 8 constructs that had pupal activity also drove expression in at least 1 larval organ (supplementary fig. S4, Supplementary Material online). Six constructs drove expression in the foregut, 3 in the epidermis, and 3 in the brain. None of these reporter constructs drove expression in imaginal discs, even though some of

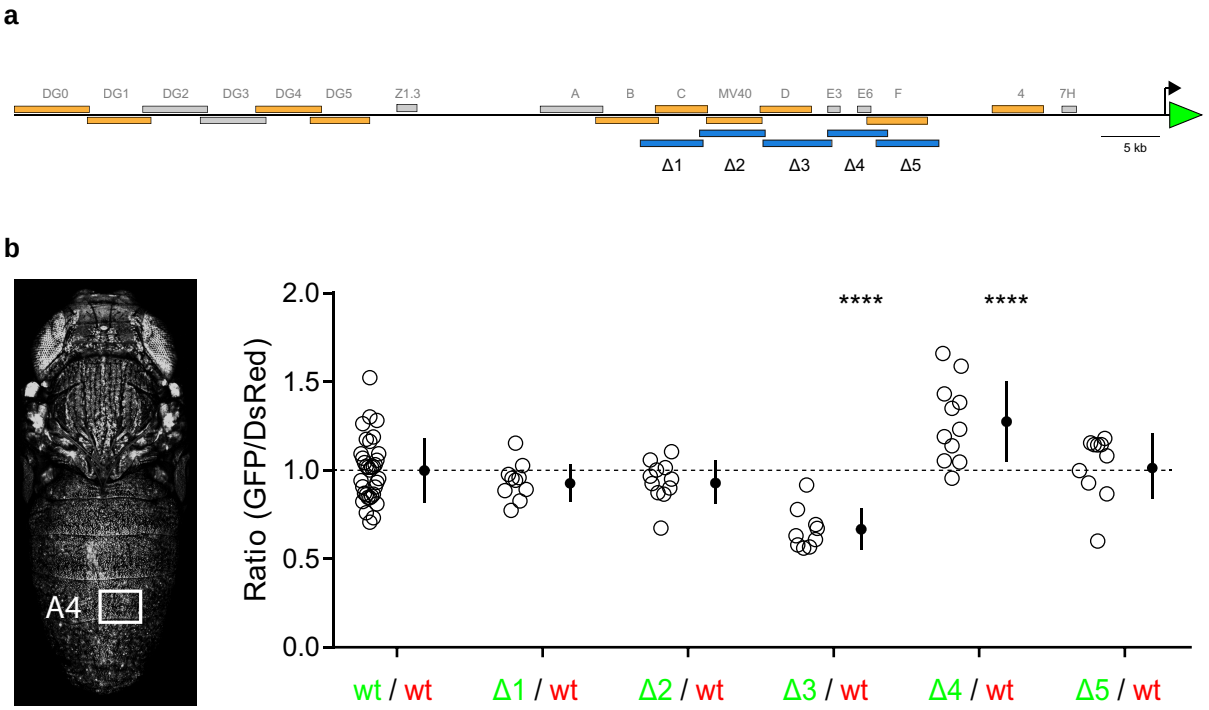

Fig. 5. DNA within one of the large regions with presumptive regulatory activity bears pupal enhancer activity in vivo. a) Position of deletions (blue boxes) that were generated in a BAC reporter construct carrying the regulatory region of *svb* and a GFP reporter (*svb*BAC-GFP). A wild-type *svb*BAC-DsRed BAC was used to normalize GFP levels. b) Effect of deletions on pupal expression. The GFP/DsRed ratio was measured in part of abdominal segment A4 (rectangle). Open circles indicate the average ratio (GFP/DsRed) for each individual. Closed black circles and vertical lines indicate mean and 1 SD, respectively. The dashed line marks the mean ratio of wild-type constructs. Statistical significance was calculated using 1-way ANOVA and Dunnett's pairwise comparisons (****$P < 0.0001$).

these regions appear to have open chromatin in this organ (supplementary fig. S1, Supplementary Material online). It is possible that these open regions are not active, as has been observed for other regions of open chromatin when assayed in imaginal discs (McKay and Lieb 2013).

Novel Pupal Enhancers Are Required for *svb* Expression

Reporter constructs provide evidence that multiple DNA fragments within the *svb cis*-regulatory region contain information for driving the wild-type *svb* expression pattern. To determine whether these regions are required for in vivo activity, we used BAC recombineering to generate 5 mutant versions of *svb*BAC-GFP, each containing a deletion of approximately 5 kb. Four deletions were made in regions with enhancer activity in reporter assays (Δ1 to Δ4) and 1 deletion was made in a region with no enhancer activity (Δ5) (Fig. 5a). These 5 versions of *svb*BAC-GFP were integrated into a specific attP site of the *D. melanogaster* genome. To normalize the fluorescence signal, we compared the GFP signal from mutant *svb*BACs with the DsRed signal from a wild-type *svb*BAC (*svb*BAC-DsRed) integrated into a different attP site, to avoid transvection (Mellert and Truman 2012). We quantified the expression of GFP BACs carrying deletions and *svb*BAC-DsRed in the epidermis of the dorsal abdomen. For each pupa, we calculated the average intensity of GFP fluorescence in segmented nuclei and normalized it with the average intensity of DsRed fluorescence in the same nuclei (Fig. 5b; see Materials and Methods for details). Two deletions, both from regions with enhancer activity in reporter assays, altered GFP expression (Δ3 and Δ4). Remarkably, while Δ3 reduced expression level, deletion of a neighboring region (Δ4) increased expression (Fig. 5b). The increase in GFP expression in Δ4 was also observed when deleting the pleiotropic enhancer E6 in *svb*BAC-GFP, which is included in the fragment deleted in Δ4 (Preger-Ben Noon et al. 2018). Δ1 and Δ2 diminished mean expression levels slightly but not significantly (Fig. 5b). The deletion of a region with no enhancer activity (Δ5) had no significant effects on GFP expression (Fig. 5b).

These results reveal, first, that DNA sequences outside of previously characterized enhancers are required for wild-type *svb* expression in the pupal epidermis. Second, deletions of 2 regions with enhancer activity in reporter assays do not alter gene expression significantly. The activity of these regions may be buffered by other DNA regions with similar expression patterns, as has been observed for *svb* embryonic expression (Frankel et al. 2010). However, we cannot rule out that these 2 regions do not contain regulatory activity in vivo. Finally, the increase in expression in Δ4 may be explained with a model of enhancer–promoter competition (Bothma et al. 2015) or, alternatively, with a scenario in which Δ4 contains regulatory

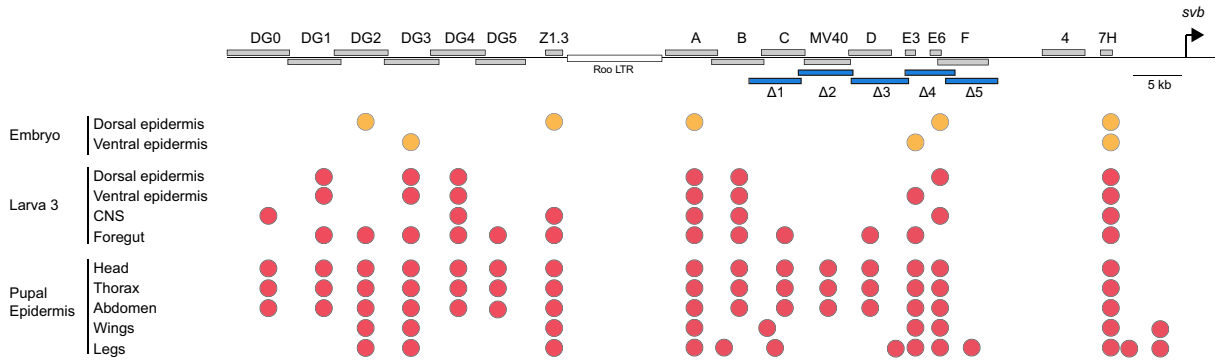

Fig. 6. A myriad of enhancers in a large regulatory region. Summary of enhancer activities in the *svb* locus in embryo, larva, and pupa. Orange circles indicate embryonic activity and red circles indicate larval/pupal activity. Expression data derive from previous works (McGregor et al. 2007; Frankel et al. 2010; Kittelmann et al. 2018; Preger-Ben Noon et al. 2018) and this study.

elements that both enhance and silence gene expression. Altogether, these findings suggest that despite the observed redundancy in the activity of *lacZ* constructs, DNA pieces within *svb* regulatory region may not have an equivalent function in vivo.

Regulatory Function Appears as a Major Determinant of Noncoding Sequence Conservation in the *svb* Locus and throughout the *Drosophila* Genome

We built an atlas of regulatory elements for the *svb* locus (Fig. 6) using data from this work together with comparable experimental data collected in our previous studies in embryo, larva, and pupa (McGregor et al. 2007; Frankel et al. 2010; Kittelmann et al. 2018; Preger-Ben Noon et al. 2018). Embryonic enhancers do not occupy much of the *svb* upstream region (we know that DNA regions between embryonic enhancers bear no activity in the embryo because these regions were scrutinized for embryonic expression with the same reporter constructs that were used in this study). However, if we also consider *svb* expression in different tissues of the larva and pupa, the regulatory elements associated with these contexts cover most of this noncoding DNA (Fig. 6). This atlas shows, prominently, that regulatory information for driving *svb* expression is scattered throughout the whole *cis*-regulatory region (Fig. 6). This regulatory architecture may contribute to the high sequence conservation of the *svb cis*-regulatory region.

We wondered if this link between the density of regulatory information and the extent of sequence conservation is a general phenomenon. Hence, we decided to explore a possible relationship between noncoding sequence conservation and open chromatin in other developmental genes and throughout the *Drosophila* genome. Although the presence of open chromatin is not directly equivalent to the presence of regulatory DNA (open chromatin might also be related to genome replication or chromatin structure), it is likely that a large fraction of open chromatin in noncoding DNA is related to the regulation of gene expression. To determine whether sequence conservation can be explained by the density of putative regulatory elements, we intersected ATAC-seq peaks for 79 developmental contexts with the collection of phastCons noncoding conserved elements of the *Drosophila* genome. We first inspected the regulatory regions of well-studied developmental genes that possess large noncoding regions in *cis* (Nelson et al. 2004) and that are known to be active across multiple developmental contexts (*Dll*, *en*, *Ets98B*, *fkh*, *hth*, *vvl*, and *svb*; Fig. 7). These large regulatory regions were defined by the positions of validated regulatory elements smaller than 2 kb (see Fig. 7c and supplementary fig. S5, Supplementary Material online). We observed that these regulatory regions have a high density of ATAC-seq peaks (putative regulatory elements), since 34% to 50% of their bases lie within open chromatin when summing all 79 developmental contexts (Fig. 7a). We also found that between 49% and 63% of their bases are found in conserved elements (Fig. 7a). With these data, we examined the fraction of conserved bases that fall within ATAC-seq peaks. We calculated how this percentage changes with the number of contexts that are considered (Fig. 7b). We observed that this parameter first grows rapidly and that the slope decreases when about 20 contexts have been included in the analysis. However, curves do not seem to reach an asymptote (Fig. 7b). When considering all contexts, 39% to 57% of the conserved bases in these regulatory regions fall within open chromatin (Fig. 7b). To rule out that the overlap between open chromatin and conserved bases is a mere result of the abundance of these 2 features in the genome, we performed an odds ratio analysis with the number of conserved and nonconserved bases that occur outside open-chromatin peaks. Since regulatory elements can evolve rapidly (Swanson et al. 2010), not all their bases are expected to be conserved. However, an odds ratio may indicate whether there is an enrichment of conserved bases in putative regulatory elements (open-chromatin regions). Indeed, we observed that conserved bases are more likely to be found in open chromatin than nonconserved bases (for all regulatory regions, odds ratios are significantly higher than 1, Fisher's exact test $P < 0.00001$; see supplementary fig. S6A, Supplementary Material online).

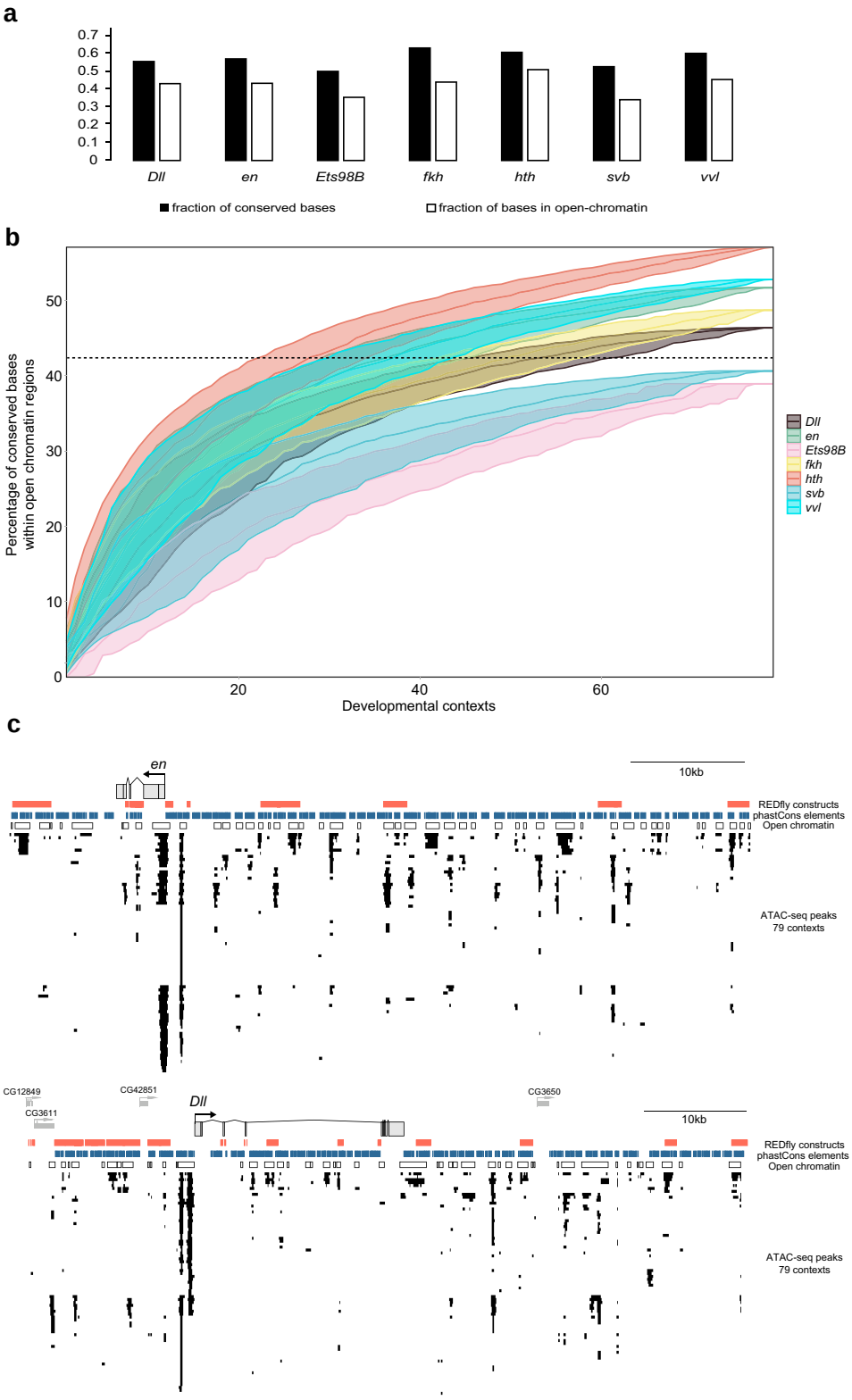

Fig. 7. Extensive overlap between putative regulatory elements and phastCons elements in regulatory regions of developmental genes and throughout the noncoding *Drosophila* genome. a) Fraction of conserved bases (black bars) and fraction of bases in open chromatin considering all developmental contexts (white bars) in regulatory regions of developmental genes. b) Overlap between phastCons elements and open chromatin in regulatory regions of developmental genes considering different number of developmental contexts. Curves vary slightly depending on the order of sampling of contexts. Colored lines indicate average values. Shaded bounding areas indicate upper and lower values. The dotted line indicates the average overlap for 10,000 random windows, considering all 79 contexts. c) Overlap between conservation and open chromatin in the regulatory regions of the developmental genes *engrailed* (above) and *Distal-less* (below). Black boxes indicate regions of open chromatin for each of the 79 developmental contexts. White boxes indicate regions of open chromatin for all 79 developmental contexts. *phastCons* elements are marked in blue. Red boxes indicate regions that are part of reporter constructs with validated regulatory activity (REDFly constructs). These data were used to generate the graph of a).

We also calculated the percentage of bases in conserved elements that are within ATAC-seq peaks in the 10,000 random windows of the genome that we had generated for the analysis of Fig. 1c (each window is the size of the regulatory region of *svb*). We considered all 79 developmental contexts for this calculation. We found that, on average, 42.4% of the bases in conserved elements are within putative regulatory DNA (dotted line in Fig. 7b). This average is quite similar to the value obtained for the *svb* regulatory region counting all 79 developmental contexts (40.6%). In 82.5% of these 10,000 windows, we also observed that conserved bases are more likely to be found in open chromatin than nonconserved bases (odds ratios are higher than 1 in 82.5% of the windows; see supplementary fig. S6B, Supplementary Material online). These results support the hypothesis that regulatory DNA contributes to the high levels of sequence conservation observed in noncoding regions of the *D. melanogaster* genome.

Discussion

At many loci in the *Drosophila* genome, noncoding regions display high levels of sequence conservation that cannot be explained by the density and distribution of known *cis*-regulatory elements. This is clearly demonstrated at the *svb* locus, where the 7 embryonic enhancers that we identified previously (McGregor et al. 2007; Frankel et al. 2010) cannot account for the high level of sequence conservation. In this work, we showed that the density of regulatory information can explain, at least in part, sequence conservation in the regulatory region of the *svb* gene. Through functional analyses, we demonstrated that the regulatory landscapes of the *svb* locus in embryo and pupa are different. While the activation of *svb* in the embryo is driven by compact enhancers, pupal *svb* expression is generated by extensive regions of regulatory information. Whether regulatory information for pupal expression is continuous over the extensive regions or is constituted by clusters of small and discrete enhancers remains to be elucidated. A more detailed analysis of the function of *svb* regulatory elements in the pupa is needed to determine whether and how the regulatory architecture of *svb* in the pupa differs from that of the embryo.

Close examination of trichome morphology reveals that sizes and shapes of trichomes vary over the embryo, larva, and adult, and previous work has shown that Svb levels modulate trichome number and size (Delon et al. 2003). Thus, it is conceivable that different levels of *svb* are needed in the embryo and pupa to generate trichomes with wild-type size. It is possible that the integrated output of multiple elements with redundant expression patterns ensures the correct transcriptional output for the *svb* gene in the pupal epidermis. Such a feature has been observed during mouse embryogenesis, where multiple redundant enhancer elements act in concert to activate the genes *Ihh* (Will et al. 2017) and *Fgf8* (Hörnblad et al. 2021). Similarly, the transcriptional activity of locus control regions and super-enhancers is achieved through the action of multiple small enhancer elements (Grosveld et al. 2021).

We explored whether the pattern of conservation and density of regulatory elements observed in the *svb* locus is replicated in other developmental genes and throughout the *Drosophila* genome. We found that other regions also have a high density of putative regulatory elements (ATAC-seq peaks). By analyzing the overlap between conserved elements and ATAC-seq peaks, we found that more than 40% of the bases in conserved elements of the *Drosophila* genome fall within open chromatin. Furthermore, in some regulatory regions of developmental genes, this percentage rises above 50%. This finding begs the question as to whether patterns of DNA conservation can be explained solely by the presence of regulatory DNA. To get an answer to this question, it will be necessary to obtain ATAC-seq data for many more developmental contexts and to determine what fractions of open chromatin are devoted to regulatory functions or genome structure/replication functions.

The density of putative regulatory activity that we detected at *Drosophila* developmental genes appears different from that for vertebrate developmental genes. For example, enhancers of the *HoxD* cluster and *Shh* and *Pax6* genes in mice are separated by tens of kilobases of noncoding DNA that do not seem to encode a function (Montavon et al. 2011; Anderson et al. 2014; Buckle et al. 2018). Similarly, patterns of phylogenetic footprinting in noncoding DNA of vertebrates indicate that conserved regions are separated by large nonconserved sequences (Santini et al. 2003; Sun et al. 2006; Navratilova et al. 2009; Peterson et al. 2009). In contrast to vertebrates, the high density of regulatory information in noncoding DNA of *Drosophila* may simply result from the compactness of its genome (Nelson et al. 2004). The different regulatory architectures observed in *Drosophila* and vertebrate genes open the question as to whether there are fundamental differences in the mechanisms of gene activation between these clades. Is regulatory information in compact genomes just pressed more tightly? Or is high regulatory density also associated to a different mode of transcriptional regulation?

Altogether, our results suggest that the noncoding *Drosophila* genome is dense with regulatory information. A future challenge is to determine how this plethora of regulatory information is integrated in space and time to achieve precise regulatory outputs.

Materials and Methods

Fly Strains

Enhancer-*lacZ* reporter lines B, C, MV40, D, F, and 4 are described in McGregor et al. (2007), while enhancer-*lacZ* reporter lines DG0, DG1, DG4, and DG5 are described in Frankel et al. (2010). *svb*BAC-GFP and *svb*BAC-DsRed lines are described in Preger-Ben Noon et al. (2018). We used BAC recombineering (Wang et al. 2009) to delete regions of approximately 5 kb in the context of the *svb*BAC-GFP. All primers and constructs that were used for BAC recombineering are listed in supplementary table S2, Supplementary Material online. These constructs were integrated into the fly genome through attP/attB recombination (Rainbow

Transgenic Flies Inc.). The different versions of *svb*BAC-GFP were integrated in attP site VK00033. *svb*BAC-DsRed was integrated in attP site VK00037.

X-GAL Stainings

Third-instar larvae were dissected in PBS and fixed in PBS with 4% formaldehyde for 10 min. Staged pupae were removed from the pupal case and then fixed in PBS with 4% formaldehyde for 15 min. After washing in PBT (PBS + 0.1% Triton X-100), samples were incubated with X-Gal solution (5 mM K4[Fe + 2(CN)6], 5 mM K3[Fe + 2(CN)6], 1 mg/ml X-Gal in PBT) at 37°C for 1 h. The samples were mounted and imaged with bright-field microscopy. We used a fly line that does not carry a *lacZ* reporter as a negative control.

Immunofluorescence in Pupa and Larva

Larva 3 tissues were dissected, fixed, and stained using a standard protocol with anti-Svb1s (1:300) and anti-rabbit Alexa-488 (1:300; Thermo Fisher Scientific). The pupal abdominal epidermis was dissected, fixed, and stained with anti-Svb1s (1:300) and anti-rabbit Alexa-488 (1:300; Thermo Fisher Scientific) as described in http://gompel.org/wp-content/uploads/2015/09/2003-12-pupal_epidermis.pdf.

Microscopy and Image Analysis

Pupae of the desired stages were removed from the pupal case and placed in a microscope slide for imaging. For live GFP imaging in the pupa, the GFP signal was measured over a *z* stack in a confocal microscope. Images were analyzed using ImageJ software (http://rsb.info.nih.gov/ij/). We used the sum projection of the *z* stacks to analyze qualitatively the GFP levels between different stages. To analyze the effect of enhancer deletions in *svb*BACs, we measured GFP and DsRed levels in pupae carrying *svb*BAC-GFP (wild type and deletions) and svbBAC-DsRed (wild type). GFP and DsRed signals were measured sequentially over a *z* stack in a confocal microscope. Images were analyzed using ImageJ software (http://rsb.info.nih.gov/ij/). First, the background was subtracted using a 50-pixel rolling-ball radius in each slice of the confocal *z* stack. Then, we calculated the sum projection of the *z* stacks for each channel to compare GFP versus DsRed levels. Segmentation masks were applied with Ilastik 1.2.0 software (http://ilastik.org) to the sum projections of the GFP channel. We measured the fluorescence mean intensities of each nucleus with the "Analyze particles" tool in ImageJ. Then, we calculated the average of the fluorescence mean intensity of all segmented nuclei. Last, we computed the ratio GFP/DsRed in each nucleus and calculated the average ratio for all segmented nuclei.

Multiple Sequence Alignment of the *svb Cis*-Regulatory Region and Quantification of Evolutionary Conservation

We defined the *svb* regulatory region in *D. melanogaster* as the segment between the *svb* TSS and the last base of the coding region of the gene *Ptp4E* (92,355 bp). Coordinates of the *svb* regulatory region in most species were obtained from an existing multispecies genome alignment (27-way alignment in the USCS Genome Browser). We filtered the MAF file with MafFilter (Dutheil et al. 2014) to keep the sequences orthologous to the *svb* regulatory region of *D. melanogaster*. Since we could not find the whole regulatory region of *svb* for *Drosophila persimilis*, *Drosophila mojavensis*, *Musca domestica*, *Apis mellifera*, *Anopheles gambiae*, and *Tribolium castaneum* in the alignment file, we searched for the coordinates of the complete region using BLASTP with the protein sequences of Ptp4E and Svb as queries (these 2 coding regions flank the regulatory region of *svb* in *D. melanogaster*). For the non-*Drosophila* species, we observed that *Ptp4E* was >200 kb away from *svb* first exon. Given that the largest *svb* regulatory region in *Drosophila* species is 150 kb (*D. mojavensis*), we defined the *svb* regulatory region in non-*Drosophila* as the 150-kb sequence upstream of the *svb* TSS. We extracted the sequences of the *svb* regulatory region from all 26 species with BEDTools (Quinlan and Hall 2010). We generated a multiple alignment of the *svb* regulatory region using the local multiple sequence aligner TBA from multiz package (Blanchette et al. 2004). To perform the multialignment, TBA generates a series of pairwise alignments to "seed" the multiple alignment process. We performed the pairwise alignments using an optional "blastz specs file" with the following parameters: --hspthresh = 1500 –gappedthresh = 1500. Measurements of evolutionary conservation in the *svb* regulatory region were performed with phastCons from the PHAST package (Siepel et al. 2005) with the following parameters: expected-length = 45, target-coverage = 0.3, rho = 0.3. The nonconserved model was downloaded from the University of California Santa Cruz (UCSC) site (https://hgdownload.soe.ucsc.edu/goldenPath/dm6/phastCons27way/). To identify conserved elements, we ran phastCons with the --most-conserved --score parameters. We removed poorly conserved elements (elements < 25 bp and with log odds scores < 60). Tracks were visualized using the UCSC Genome Browser. Genome-wide measurements of evolutionary conservation were performed using phastCons data based on the existing multispecies genome alignment from UCSC (phastCons27way track). Again, we removed poorly conserved elements (elements < 25 bp and with log odds scores < 60). We removed all exons and transposons not overlapping phastCons elements from the alignment; 10,000 random windows of 82,184 bp (the length of the *svb* regulatory region after removing exons and the Roo transposable element) were selected from the *D. melanogaster* genome. The fraction of conserved bases (bases within conserved elements/total number of bases) was calculated using Bedmap in BEDOPS suite (Neph et al. 2012) with parameters --echo --delim '\t' –-bases-uniq, with the windows bed file as reference file and the conserved elements bed file as map file.

Overlap between Conserved Elements and Open Chromatin in Developmental Genes and throughout the Genome

To study the relationship between sequence conservation and regulatory activity in noncoding regions of the genome,

we downloaded processed bulk ATAC-seq peaks from ChIP-ATLAS (https://chip-atlas.org/) and single-cell ATAC-seq embryonic data from scEnhancer (http://enhanceratlas.net/scenhancer). We also included ATAC-seq peaks from the abdominal pupal epidermis generated in this work. We merged peaks from the same developmental context when more than 1 source was available. The whole data set consisted of 79 distinct tissues and cell types, including embryonic cell types, and larval, pupal, and adult tissues (supplementary table S3, Supplementary Material online).

We defined regulatory regions for 7 classical developmental genes (*Dll*, *en*, *Ets98B*, *fkh*, *hth*, *svb*, and *vvl*) by using REDfly (redfly.ccr.buffalo.edu) elements smaller than 2 kb. For each gene, the regulatory region was defined as the DNA region between the 2 most distant validated regulatory elements for the gene. We calculated the overlap between conserved bases (bases in conserved phastCons elements) and open-chromatin regions; 500 growth curves were generated for each regulatory region by sampling an increasing number of developmental contexts (from all 79 contexts) in different orders.

To obtain odds ratios, we first calculated the number of conserved bases within (A) and outside (B) open chromatin (odds = A/B). Next, we calculated the number of nonconserved bases within (C) and outside (D) open chromatin (odds = C/D). The odds ratio for each regulatory region = AD/BC. We performed the same calculation for the 10,000 random windows. We excluded 61 random windows because they had 0 conserved or nonconserved bases. For the 7 regulatory regions, we used Fisher's exact tests to determine the statistical significance of the enrichment of conserved bases in open chromatin. To correct for multiple comparisons, we used the Bonferroni method.

ATAC-seq in Pupal Abdominal Epidermis

For each replicate ($n = 3$), we removed abdominal cuticles from 50 pupae 38 to 45 hAPF. We carefully removed internal organs with forceps in cold PBS, to retain only epidermal cells, which are attached to the cuticle. We followed the Omni-ATAC protocol (Corces et al. 2017) but used a lysis buffer based on IGEPAL detergent (lysis buffer: 10 mM Tris–HCl, pH = 7.5, 10 mM NaCl, 3 mM MgCl2, and 0.1% IGEPAL CA-630). Library concentration was quantified with a KAPA Library Quantification Kit (Roche), and quality control was performed with the High-Sensitivity DNA Analysis Kit in a Bioanalyzer 2100 (Agilent). Libraries were sequenced on a NextSeq 550 system (Illumina) with 75-bp PE reads.

CUT&RUN against H3K27ac in Pupal Abdominal Epidermis

For each replicate ($n = 2$), we removed abdominal cuticles from 10 pupae 38 to 45 hAPF. We generated 2 replicates for the histone H3K27ac antibody (39134, Active Motif) and 1 control replicate for the Normal Rabbit IgG antibody (Cat. 2729S, Cell Signaling). Both antibodies were used at a 1:100 dilution. After dissection, we washed pupal abdomens by replacing PBS with 1 ml of wash + buffer (20 mM HEPES pH 7.5, 150 mM NaCl, and 0.5 mM spermidine with Roche complete EDTA-free protease inhibitor) and centrifuging twice at 12,000 × *g* for 5 min. We resuspended pupal abdomens in 15 µl of BioMag Plus Concanavalin-A-conjugated magnetic beads (ConA beads, Polysciences, Inc.) and followed a CUT&RUN protocol for *Drosophila* tissues (https://dx.doi.org/10.17504/protocols.io.umfeu3n). Libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB) following instructions but with the following changes: (i) adaptors were diluted 1:10 in water for adaptor ligation (step 2), (ii) the size selection of the adaptor-ligated DNA in step 3A was omitted (we proceeded directly to step 3B), and (iii) we performed 14 cycles of PCR with 10 s of annealing/extension for enrichment of short DNA fragments. Libraries were sequenced on a NovaSeq 6000 system with 150-bp PE reads.

ChIP-seq against H3K27ac in *svb*+ Cells of the Embryo

Stage 14 to 15 embryos from a line containing E10::GFP and 7::DsRed transgenes (Preger-Ben Noon et al. 2016) were cross-linked, and dissociated and isolated nuclei were immunostained with anti-GFP and anti-DsRed antibodies and the appropriate secondary antibodies. The E10::GFP and 7::DsRed nuclei, which constitute the majority of *svb*-expressing nuclei, were then isolated by FACS. Chromatin from 250,000 nuclei of each cell subpopulations ($n = 3$) was isolated and used for ChIP with anti-H3K27ac and anti-H3 antibodies (Abcam) using the iDeal ChIP-seq Kit (Diagenode). Libraries were prepared using the Ovation Ultralow V2 DNA-Seq Library Preparation Kit (NuGen) according to the manufacturer's instructions. Libraries were sequenced on a NextSeq 550 system (Illumina) with 50-bp SE reads.

Source of ATAC-seq Data from Embryonic Epidermal Cells, Wing Imaginal Disc, Eye-Antenna Imaginal Disc, Larval Brain, and T2 Pupal Leg

For embryonic epidermal cells, the pooled single-cell ATAC-seq fastq files were downloaded from the Gene Expression Omnibus (accession number GSE101581). We filtered reads from epidermal cells from 10- to 12-h embryos (stages 14 to 15) (Cusanovich et al. 2018) using files available at https://descartes.brotmanbaty.org/bbi/fly-chromatin-accessibility and https://github.com/shendurelab/fly-atac. Raw fastq files from ATAC-seq experiments from the wing imaginal disc, eye-antenna imaginal disc, third instar larva brain, and pupa T2 leg were downloaded from Gene Expression Omnibus (accession numbers GSE102841, GSE59078, GSE102441, and GSE113240, respectively).

Mapping and Analysis of NGS Data

The fastq files were processed with a custom Python pipeline (available at https://github.com/laiker96/fastq_to_bam). Briefly, we ran bbduk.sh (sourceforge.net/projects/bbmap/) to trim adapters and aligned adapter-corrected reads to the *D. melanogaster* reference genome dm6 with BWA (Li and Durbin 2009) to obtain SAM files. SAM files were compressed

to BAM files with SAMtools (http://www.htslib.org), and duplicated reads were marked with MarkDuplicates from the Picard suite (https://broadinstitute.github.io/picard/). BAM files were filtered based on MAPQ values (≥20) and SAM flags 4, 256, 1024, and 2048 (unmapped, secondary, duplicate, and supplementary reads) with SAMtools. In paired-end samples, we only kept properly paired reads (SAM flag 2). The number of reads and alignment rate for each NGS sample is summarized in supplementary table S4, Supplementary Material online. Pearson correlation coefficients between replicates, computing read counts in 1-kb windows, were between 0.94 and 0.99 for ATAC-seq of the pupal epidermis, between 0.79 and 0.90 for H3K27ac ChIP-seq of *svb* positive cells, and 0.96 for H3K27ac CUT&RUN of the pupal epidermis. We merged BAM files corresponding to the H3K27ac ChIP-seq of 7 cells and E10 cells of the embryo after removing duplicates and subsampling to the size of the smaller data set. To compare chromatin aperture between contexts, we normalized ATAC-seq experiments taking into account differences in the signal-to-noise ratio between different libraries using S3norm (Xiang et al. 2020). Because we used data from both paired-end and single-end sequencing experiments, we created raw count bedgraph files corresponding to the regulatory region of *svb* using only the first mapped mate from paired-end experiments and used these files as input for the S3norm normalization. The S3 normalized bedgraph files were converted to bigwig files with bedGraphToBigWig (Kent et al. 2010). We fed filtered BAM files to MACS2 (Zhang et al. 2008) to call peaks in ATAC-seq data. For pupal ATAC-seq data, we merged BAM files from the different replicates. MACS2 was used with the following parameters: -g dm -f BAM --keep-dup all --shift -50 --extsize 100 --nomodel. To normalize H3K27ac data, we used IgG (CUT&RUN) and histone H3 (ChIP-seq). H3K27ac experiments were normalized with bamCompare from the DeepTools suite (Ramírez et al. 2016) with --scaleFactorsMethod SES and a bin size of 1. Genome browser plots were generated with the pyGenomeTracks package (Lopez-Delisle et al. 2021).

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank François Payre for sharing the anti-Svb antibody. This work was supported by grants PICT 2016-0414 and PICT 2019-00497 (Agencia, Argentina) to N.F.

Author Contributions

G.S., D.M.O., E.P.-B.N., S.N., and N.F. generated the experimental data. I.L. and I.M. performed the in silico analyses. G.S., D.M.O., I.L., I.M., G.C., D.L.S., E.P.-B.N., and N.F. designed the experiments. N.F. conceived the study and wrote the manuscript with input from all authors. Figures and tables were designed by G.S., D.M.O., I.L., I.M., and N.F.

Conflict of Interest. The authors declare no competing interests.

Data Availability

Raw and processed data generated in this paper are available in Gene Expression Omnibus under accession number GSE253028.

References

Anderson E, Devenney PS, Hill RE, Lettice LA. Mapping the Shh long-range regulatory domain. *Development*. 2014:**141**(20):3934–3943. https://doi.org/10.1242/DEV.108480.

Andolfatto P. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature*. 2005:**437**(7062):1149–1152. https://doi.org/10.1038/nature04107.

Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, *et al.* Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res*. 2004:**14**(4):708–715. https://doi.org/10.1101/gr.1933104.

Bothma JP, Garcia HG, Ng S, Perry MW, Gregor T, Levine M. Enhancer additivity and non-additivity are determined by enhancer strength in the *Drosophila* embryo. *eLife*. 2015:**4**:e07956. https://doi.org/10.7554/ELIFE.07956.001.

Buckle A, Nozawa Rs, Kleinjan DA, Gilbert N. Functional characteristics of novel pancreatic Pax6 regulatory elements. *Hum Mol Genet*. 2018:**27**(19):3434–3448. https://doi.org/10.1093/HMG/DDY255.

Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*. 2008:**134**(1):25–36. https://doi.org/10.1016/j.cell.2008.06.030.

Chanut-Delalande H, Hashimoto Y, Pelissier-Monier A, Spokony R, Dib A, Kondo T, Bohère J, Niimi K, Latapie Y, Inagaki S, *et al.* Pri peptides are mediators of ecdysone for the temporal control of development. *Nat Cell Biol*. 2014:**16**(11):1035–1044. https://doi.org/10.1038/ncb3052.

Claringbould A, Zaugg JB. Enhancers in disease: molecular basis and emerging treatment strategies. *Trends Mol Med*. 2021:**27**(11):1060–1073. https://doi.org/10.1016/J.MOLMED.2021.07.012/ATTACHMENT/84988D99-2316-4D3F-9DA1-26ADCA75651E/MMC1.XLSX.

Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods*. 2017:**14**(10):959–962. https://doi.org/10.1038/NMETH.4396.

Crocker J, Abe N, Rinaldi L, McGregor AP, Frankel N, Wang S, Alsawadi A, Valenti P, Plaza S, Payre F, *et al.* Low affinity binding site clusters confer Hox specificity and regulatory robustness. *Cell*. 2015:**160**(1-2):191–203. https://doi.org/10.1016/j.cell.2014.11.041.

Crocker J, Stern DL. Functional regulatory evolution outside of the minimal even-skipped stripe 2 enhancer. *Development*. 2017:**144**(17):3095–3101. https://doi.org/10.1242/DEV.149427.

Cusanovich DA, Reddington JP, Garfield DA, Daza RM, Aghamirzaie D, Marco-Ferreres R, Pliner HA, Christiansen L, Qiu X, Steemers FJ, *et al.* The *cis*-regulatory dynamics of embryonic development at single-cell resolution. *Nature*. 2018:**555**(7697):538–542. https://doi.org/10.1038/NATURE25981.

Davidson EH. *The regulatory genome: gene regulatory networks in development and evolution*. Oxford (UK): Elsevier Science; 2010.

Davis GK, Srinivasan DG, Wittkopp PJ, Stern DL. The function and regulation of ultrabithorax in the legs of *Drosophila*

melanogaster. *Dev Biol*. 2007:**308**(2):621–631. https://doi.org/10.1016/J.YDBIO.2007.06.002.

Delon I, Chanut-Delalande H, Payre F. The Ovo/Shavenbaby transcription factor specifies actin remodelling during epidermal differentiation in *Drosophila*. *Mech Dev*. 2003:**120**(7):747–758. https://doi.org/10.1016/S0925-4773(03)00081-9.

Dutheil JY, Gaillard S, Stukenbrock EH. MafFilter: a highly flexible and extensible multiple genome alignment files processor. *BMC Genomics*. 2014:**15**(1):1–10. https://doi.org/10.1186/1471-2164-15-53/FIGURES/5.

Frankel N. Multiple layers of complexity in *cis*-regulatory regions of developmental genes. *Dev Dyn*. 2012:**241**(12):1857–1866. https://doi.org/10.1002/dvdy.23871.

Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature*. 2010:**466**(7305):490–493. https://doi.org/10.1038/nature09158.

Frankel N, Erezyilmaz DF, McGregor AP, Wang S, Payre F, Stern DL. Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature*. 2011:**474**(7353):598–603. https://doi.org/10.1038/nature10200.

Frankel N, Wang S, Stern DL. Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution. *Proc Natl Acad Sci U S A*. 2012:**109**(51):20975–20979. https://doi.org/10.1073/pnas.1207715109.

Fujioka M, Emi-Sarker Y, Yusibova GL, Goto T, Jaynes JB. 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*. 1999:**126**(11):2527–2538. https://doi.org/10.1242/DEV.126.11.2527.

Grosveld F, Van Staalduinen J, Stadhouders R. Transcriptional regulation by (super)enhancers: from discovery to mechanisms. *Annu Rev Genomics Hum Genet*. 2021:**22**(1):127–146. https://doi.org/10.1146/ANNUREV-GENOM-122220-093818.

Halfon MS. (Re)Modeling the transcriptional enhancer. *Nat Genet*. 2006:**38**(10):1102–1103. https://doi.org/10.1038/NG1006-1102.

Halligan DL, Keightley PD. Ubiquitous selective constraints in the *Drosophila* genome revealed by a genome-wide interspecies comparison. *Genome Res*. 2006:**16**(7):875–884. https://doi.org/10.1101/GR.5022906.

Hill MS, Zande PV, Wittkopp PJ. Molecular and evolutionary processes generating variation in gene expression. *Nat Rev Genet*. 2021:**22**(4):203–215. https://doi.org/10.1038/s41576-020-00304-w.

Hörnblad A, Bastide S, Langenfeld K, Langa F, Spitz F. Dissection of the Fgf8 regulatory landscape by in vivo CRISPR-editing reveals extensive intra- and inter-enhancer redundancy. *Nat Commun*. 2021:**12**(1):1. https://doi.org/10.1038/S41467-020-20714-Y.

Janssens H, Hou S, Jaeger J, Kim AR, Myasnikova E, Sharp D, Reinitz J. Quantitative and predictive model of transcriptional control of the *Drosophila melanogaster* even skipped gene. *Nat Genet*. 2006:**38**(10):1159–1165. https://doi.org/10.1038/ng1886.

Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics*. 2010:**26**(17):2204. https://doi.org/10.1093/BIOINFORMATICS/BTQ351.

Kittelmann S, Buffry AD, Franke FA, Almudi I, Yoth M, Sabaris G, Couso JP, Nunes MDS, Frankel N, Gómez-Skarmeta JL, *et al*. Gene regulatory network architecture in different developmental contexts influences the genetic basis of morphological evolution. *PLoS Genet*. 2018:**14**(5):e1007375. https://doi.org/10.1371/journal.pgen.1007375.

Kittelmann S, Noon EP-B, McGregor AP, Frankel N. A complex gene regulatory architecture underlies the development and evolution of cuticle morphology in *Drosophila*. *Curr Opin Genet Dev*. 2021:**69**:21–27. https://doi.org/10.1016/j.gde.2021.01.003.

Klingler M, Soong J, Butler B, Peter Gergen J. Disperse versus compact elements for the regulation of runt stripes in *Drosophila*. *Dev Biol*. 1996:**177**(1):73–84. https://doi.org/10.1006/DBIO.1996.0146.

Kondo T, Plaza S, Zanet J, Benrabah E, Valenti P, Hashimoto Y, Kobayashi S, Payre F, Kageyama Y. Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Science*. 2010:**329**(5989):336–339. https://doi.org/10.1126/science.1188158.

Leffler EM, Bullaughey K, Matute DR, Meyer WK, Ségurel L, Venkat A, Andolfatto P, Przeworski M. Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol*. 2012:**10**(9):e1001388. https://doi.org/10.1371/JOURNAL.PBIO.1001388.

Levine M. Transcriptional enhancers in animal development and evolution. *Curr Biol*. 2010:**20**(17):2010, R754–R763. https://doi.org/10.1016/j.cub.2010.06.070.

Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009:**25**(14):1754–1760. https://doi.org/10.1093/BIOINFORMATICS/BTP324.

Long HK, Prescott SL, Wysocka J. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell*. 2016:**167**(5):1170–1187. https://doi.org/10.1016/J.CELL.2016.09.018.

Lopez-Delisle L, Rabbani L, Wolff J, Bhardwaj V, Backofen R, Grüning B, Ramírez F, Manke T. PyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics*. 2021:**37**(3):422–423. https://doi.org/10.1093/BIOINFORMATICS/BTAA692.

López-Rivera F, Foster Rhoades OK, Vincent BJ, Pym ECG, Bragdon MDJ, Estrada J, DePace AH, Wunderlich Z. A mutation in the *Drosophila melanogaster* eve stripe 2 minimal enhancer is buffered by flanking sequences. *G3 (Bethesda)*. 2020:**10**(12):4473–4482. https://doi.org/10.1534/G3.120.401777.

Ludwig MZ, Manu , Kittler R, White KP, Kreitman M. Consequences of eukaryotic enhancer architecture for gene expression dynamics, development, and fitness. *PLoS Genet*. 2011:**7**(11):11. https://doi.org/10.1371/JOURNAL.PGEN.1002364.

Ludwig MZ, Palsson A, Alekseeva E, Bergman CM, Nathan J, Kreitman M. Functional evolution of a *cis*-regulatory module. *PLoS Biol*. 2005:**3**(4):e93. https://doi.org/10.1371/JOURNAL.PBIO.0030093.

Mangione F, Martín-Blanco E. The Dachsous/Fat/Four-jointed pathway directs the uniform axial orientation of epithelial cells in the *Drosophila* abdomen. *Cell Rep*. 2018:**25**(10):2836–2850.e4. https://doi.org/10.1016/J.CELREP.2018.11.036.

McGregor AP, Orgogozo V, Delon I, Zanet J, Srinivasan DG, Payre F, Stern DL. Morphological evolution through multiple *cis*-regulatory mutations at a single gene. *Nature*. 2007:**448**(7153):587–590. https://doi.org/10.1038/nature05988.

McKay DJ, Lieb JD. A common set of DNA regulatory elements shapes *Drosophila* appendages. *Dev Cell*. 2013:**27**(3):306–318. https://doi.org/10.1016/J.DEVCEL.2013.10.009.

Mellert DJ, Truman JW. Transvection is common throughout the *Drosophila* genome. *Genetics*. 2012:**191**(4):1129–1141. https://doi.org/10.1534/GENETICS.112.140475.

Montavon T, Soshnikova N, Mascrez B, Joye E, Thevenet L, Splinter E, De Laat W, Spitz F, Duboule D. A regulatory archipelago controls Hox genes transcription in digits. *Cell*. 2011:**147**(5):1132–1145. https://doi.org/10.1016/J.CELL.2011.10.023.

Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, Adrian J, Kawli T, Davis CA, Dobin A, Kaul R, *et al*. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*. 2020:**583**(7818):699–710. https://doi.org/10.1038/s41586-020-2493-4.

Navratilova P, Fredman D, Hawkins TA, Turner K, Lenhard B, Becker TS. Systematic human/zebrafish comparative identification of *cis*-regulatory activity around vertebrate developmental transcription factor genes. *Dev Biol*. 2009:**327**(2):526–540. https://doi.org/10.1016/J.YDBIO.2008.10.044.

Nelson CE, Hersh BM, Carroll SB. The regulatory content of intergenic DNA shapes genome architecture. *Genome Biol*. 2004:**5**(4):R25. https://doi.org/10.1186/GB-2004-5-4-R25.

Neph S, Scott Kuehn M, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano MT, Vierstra J, Thomas S *et al*. BEDOPS: high-performance genomic feature operations.

Bioinformatics. 2012:**28**(14):1919–1920. https://doi.org/10.1093/BIOINFORMATICS/BTS277.

Panigrahi A, OMalley BW. Mechanisms of enhancer action: the known and the unknown. *Genome Biol*. 2021:**22**(1):108. https://doi.org/10.1186/s13059-021-02322-1.

Payre F, Vincent A, Carreno S. Ovo/Svb integrates wingless and DER pathways to control epidermis differentiation. *Nature*. 1999:**400**(6741):271–275. https://doi.org/10.1038/22330.

Peterson BK, Hare EE, Iyer VN, Storage S, Conner L, Papaj DR, Kurashima R, Jang E, Eisen MB. Big genomes facilitate the comparative identification of regulatory elements. *PLoS One*. 2009:**4**(3):e4688. https://doi.org/10.1371/JOURNAL.PONE.0004688.

Pott S, Lieb JD. What are super-enhancers? *Nat Genet*. 2014:**47**(1):8–12. https://doi.org/10.1038/ng.3167.

Preger-Ben Noon E, Davis FP, Stern DL. Evolved repression overcomes enhancer robustness. *Dev Cell*. 2016:**39**(5):572–584. https://doi.org/10.1016/j.devcel.2016.10.010.

Preger-Ben Noon E, Sabarís G, Ortiz DM, Sager J, Liebowitz A, Stern DL, Frankel N. Comprehensive analysis of a *cis*-regulatory region reveals pleiotropy in enhancer function. *Cell Rep*. 2018:**22**(11):3021–3031. https://doi.org/10.1016/j.celrep.2018.02.073.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010:**26**(6):841–842. https://doi.org/10.1093/BIOINFORMATICS/BTQ033.

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. DeepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016:**44**(W1):W160–W165. https://doi.org/10.1093/NAR/GKW257.

Riddihough G, Ish-Horowicz D. Individual stripe regulatory elements in the *Drosophila* hairy promoter respond to maternal, gap, and pair-rule genes. *Genes Dev*. 1991:**5**(5):840–854. https://doi.org/10.1101/GAD.5.5.840.

Sackerson C, Fujioka M, Goto T. The even-skipped locus is contained in a 16-kb chromatin domain. *Dev Biol*. 1999:**211**(1):39–52. https://doi.org/10.1006/DBIO.1999.9301.

Santini S, Boore JL, Meyer A. Evolutionary conservation of regulatory elements in vertebrate Hox gene clusters. *Genome Res*. 2003:**13**(6A):1111–1122. https://doi.org/10.1101/GR.700503.

Schaffner W. Enhancers, enhancers—from their discovery to today's universe of transcription enhancers. *Biol Chem*. 2015:**396**(4):311–327. https://doi.org/10.1515/HSZ-2014-0303/ASSET/GRAPHIC/HSZ-2014-0303_FIG4.JPG.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, *et al*. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. 2005:**15**(8):1034–1050. https://doi.org/10.1101/GR.3715005.

Small S, Arnosti DN. Transcriptional enhancers in Drosophila. *Genetics*. 2020:**216**(1):1–26. https://doi.org/10.1534/GENETICS.120.301370.

Soverna AF, Rodriguez NC, Korgaonkar A, Hasson E, Stern DL, Frankel N. *Cis*-regulatory variation in the shavenbaby gene underlies intraspecific phenotypic variation, mirroring interspecific divergence in the same trait. *Evolution*. 2021:**75**(2):427–436. https://doi.org/10.1111/evo.14142.

Stern DL, Frankel N. The structure and evolution of *cis*-regulatory regions: the Shavenbaby story. *Philos Trans R Soc B Biol Sci* . 2013:**368**(1632):20130028. https://doi.org/10.1098/rstb.2013.0028.

Sun H, Skogerbø G, Chen R. Conserved distances between vertebrate highly conserved elements. *Hum Mol Genet*. 2006:**15**(19):2911–2922. https://doi.org/10.1093/HMG/DDL232.

Swanson CI, Evans NC, Barolo S. Structural rules and complex regulatory circuitry constrain expression of a notch- and EGFR-regulated eye enhancer. *Dev Cell*. 2010:**18**(3):359–370. https://doi.org/10.1016/j.devcel.2009.12.026.

Wang S, Zhao Y, Leiby M, Zhu J. A new positive/negative selection scheme for precise BAC recombineering. *Mol Biotechnol*. 2009:**42**(1):110–116. https://doi.org/10.1007/S12033-009-9142-3.

Will AJ, Cova G, Osterwalder M, Chan WL, Wittler L, Brieske N, Heinrich V, de Villartay JP, Vingron M, Klopocki E, *et al*. Composition and dosage of a multipartite enhancer cluster control developmental expression of Indian hedgehog. *Nat Genet*. 2017:**49**(10):1539–1545. https://doi.org/10.1038/NG.3939.

Xiang G, Keller CA, Giardine B, An L, Li Q, Zhang Y, Hardison RC. S3norm: simultaneous normalization of sequencing depth and signal-to-noise ratio in epigenomic data. *Nucleic Acids Res*. 2020:**48**(8):e43. https://doi.org/10.1093/NAR/GKAA105.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, *et al*. Model-based analysis of ChIP-seq (MACS). *Genome Biol*. 2008:**9**(9):1–9. https://doi.org/10.1186/GB-2008-9-9-R137.

Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EEM. Combinatorial binding predicts spatio-temporal *cis*-regulatory activity. *Nature*. 2009:**462**(7269):65–70. https://doi.org/10.1038/nature08531.