Genome-wide analyses of Shavenbaby target genes reveals distinct features of enhancer organization

Delphine MENORET[1,2*], Marc SANTOLINI[3*], Isabelle FERNANDES[1,2,4], Rebecca SPOKONY[5], Jennifer ZANET[1,2], Ignacio GONZALEZ[6,7], Yvan LATAPIE[1,2], Pierre FERRER[1,2], Hervé ROUAULT[3,9], Kevin P. WHITE[5], Philippe BESSE[6,7], Vincent HAKIM[3], Stein AERTS[8], Francois PAYRE[1,2#] and Serge PLAZA[1,2#]

[1] Centre de Biologie du Développement, Université de Toulouse, UPS, Toulouse, F-31062, France

[2] CNRS UMR5547, Toulouse, F-31062, France

[3] Laboratoire de Physique Statistique, CNRS, Université Pierre & Marie Curie, Université Denis Diderot, ENS, Paris, France

[4] present address: Department of Biology, McGill University, Montreal, QC, Canada

[5] Institute for Genomics and Systems Biology, Department of Human Genetics, The University of Chicago, Chicago, IL, USA

[6] Université de Toulouse, INSA, Toulouse, France

[7] Institut de Mathématiques, CNRS UMR5219, Toulouse, France

[8] Laboratory of Computational Biology, Center for Human Genetics KU Leuven, Belgium

[9] present address: Howard Hughes Medical Institute, Janelia Farm Research Campus, Ashburn, VA, USA

* The two first authors contributed equally

Corresponding authors: serge.plaza@univ-tlse3.fr, francois.payre@univ-tlse3.fr

Running title: Functional architecture of trichome enhancers

keywords: Drosophila, enhancer, epidermis, shavenbaby, morphogenesis

Abstract

Background: Developmental programs are implemented by regulatory interactions between Transcription Factors (TFs) and their target genes, which remain yet poorly understood. While recent studies have focused on regulatory cascades of TFs that govern early development, little is known on how are selected and controlled the ultimate effectors of cell differentiation. We addressed this question during late *Drosophila* embryogenesis when the finely tuned expression of a TF, Ovo/ Shavenbaby (Svb), triggers the morphological differentiation of epidermal trichomes.

Results: We defined a sizeable set of 39 Svb downstream genes and used *in vivo* assays to delineate 14 enhancers driving their specific expression in trichome cells, with highly similar pattern and dynamics. Coupling computational modeling to functional dissection, we investigated the regulatory logic of these enhancers. Genome-wide approaches, further extending the repertoire of epidermal effectors, support that the regulatory models learned from this first sample are representative of the whole set of trichome enhancers. We find that these “terminal” enhancers harbor remarkable features with respect to their functional architectures. They display weak if any clustering of Svb binding sites. The *in vivo* function of each site relies on its intimate context, with a critical importance of the flanking nucleotides. We identify two additional cis-regulatory motifs, retrieved in a broad diversity of composition and positioning among trichome enhancers, and that critically contribute to their activity.

Conclusion: Our results show Svb directly regulates a large set of terminal effectors of the remodeling of epidermal cells. Furthermore, these data reveal that trichome formation is underpinned by unexpectedly diverse modes of regulation providing fresh insights into the functional architecture of enhancers governing a terminal differentiation program.

Introduction

Many studies have established that transcriptional networks control development, through determining specific programs of genome expression [1]. These Gene Regulatory Networks (GRNs) are implemented by Transcription Factors (TFs) that bind to regulatory DNA sequences, known as enhancers or Cis-Regulatory-Modules (CRMs), to control the transcription of nearby genes. Although recruited to target genes via their DNA binding properties [2], TFs recognize only short and often degenerate motifs (reviewed in [3, 4]). Consequently, thousands of putative Binding Sites (BS) are scattered throughout the genome hampering efficient prediction of cis-regulatory regions [3, 5, 6]. The fine structure of enhancers as well as putative general rule(s) underlying their organization remain(s), however, poorly understood.

Although animals encode hundreds of TFs, only a few of them have been studied in detail to elucidate the regulatory logic of their target enhancers [7, 8]. In *Drosophila*, current knowledge of enhancer structure mainly comes from works on early development, *e.g.* TFs controlling segmentation and mesoderm specification [9-12]. Within these early acting networks, a number of studies have shown that the local enrichment for BS (homotypic or heterotypic clustering) in evolutionarily conserved regions is a general signature of active enhancers [13-15]. Functionally related enhancers (driving similar expression pattern) often share a combination or code of cis-regulatory motifs, together defining a specific program of expression [11, 16-18]. Whether enhancers rely on a constrained organization of cis-regulatory motifs or can accommodate flexibility in their number, composition and positioning remains debated (reviewed in [4, 19, 20]). While several studies have shown that regulatory codes are efficient to predict expression pattern [9, 11, 16], recent large-scale work suggests that developmental enhancers may have a more flexible architecture [10,

20]. However, in depth analyses of individual enhancers [21-24] have revealed an unexpected level of functional constraint in their intimate architecture. It has been proposed that constrained enhancers could be critical when TFs display limiting concentrations [25], *e.g.* to accurately integrate gradients [26]. On the other hand, enhancers that do not hold integrative properties might be of simpler architecture [27, 28]. Distinguishing between these possibilities thus requires detailed analyses of the structure and regulatory logic of CRM/TF interactions that occurs at late developmental stages.

Here, we focus on a GRN that controls cell morphogenesis during terminal differentiation of the *Drosophila* embryonic epidermis. The subset of epidermal cells that express the TF Ovo/Shavenbaby (Svb) [29] undergo localized changes in cell shape leading to the formation of dorsal hairs and ventral denticles, collectively referred to as trichomes [30]. Svb triggers the expression of various classes of cellular effectors in trichome cells. Developmental and genetic analyses have established that trichome formation relies on their collective action, acting together as a developmental module to promote cell shape reorganization [31-33]. The mechanisms underlying the co-expression of Svb-regulated genes in trichome cells remained yet poorly understood. A first level of regulation resides in the activity of Svb itself that is controlled in a post-translational manner, in response to small peptides encoded by the gene *polished-rice (pri)* [34]. Pri peptides trigger N-terminal truncation of the Svb protein, switching its activity from a full-length repressor to a cleaved activator [34], therefore providing a temporal control to the program of trichome formation [32]. However, little is known concerning how this TF recognizes and selects its target genes. Besides definition of DNA-binding specificity *in vitro* [35] and the identification of few targets regulated by Ovo germline-specific isoforms [35,

36], only a single epidermal enhancer dependent on Svb has been identified so far [31]. Thus, whether or not Svb targets genes that are co-expressed in trichome cells co-opted similar cis-regulatory elements remained an open question.

To address this question, we designed a set of computational modeling coupled to experimental approaches to identify and investigate the cis regulatory logic of Svb-dependent enhancers. By systematic *in vivo* assays, we first identified a robust set of Svb target effectors, specifically expressed in trichome cells at the time of their morphological differentiation. We then searched for and identified 14 Svb-dependent epidermal enhancers driving their expression in trichome cells and investigated their functional organization. Computational analyses and experimental dissection led to a refinement of the Svb BS bound *in vivo* and the identification of two additional motifs required for enhancer activity. Our studies further reveal that the distribution of these *cis*-regulatory motifs does not follow a stereotypical organization. Coupled to chromatin immunoprecipitation (ChIP-seq) and microarray profiling, the models built from these fine scale experiments allow efficient genome-wide identification of new enhancers, which drive the specific expression of trichome effectors. In summary, our results show that enhancers driving co-expression in cells of a late GRN present various composition and respective organization of cis regulatory motifs, extending the idea that co-expressed developmental enhancers can have diverse cis-regulatory architectures [11, 37], including for those mediating terminal stages of cell differentiation.

Results

Enrichment of conserved binding sites in Svb downstream genes

Previous work has identified a dozen of genes activated by Svb, each contributing to epidermal cell remodeling [31, 33, 38, 39]. To investigate the cis-regulatory logic of Svb-dependent targets, we first sought to define a larger set of Svb downstream genes appropriate for *in silico* analyses. We therefore analyzed additional candidates selected because of their expression in subsets of epidermal cells (BDGP) using *in situ* hybridization. Among 57 candidates, we identified 21 Svb-dependent genes, *i.e.*, downregulated in *svb* mutants and upregulated following *svb* ectopic expression (Fig. 1A, S1A), while the other 36 epidermal genes were found independent of Svb (Fig. S1B). Together with genes identified previously [31, 33, 38, 39], this constituted a robust set of 39 genes activated by Svb to be expressed in trichome cells. We used these 39 Svb targets to examine whether they displayed an evolutionarily conserved signature in their non-coding regions, when compared with all *Drosophila* genes, or the 36 epidermal genes independent of *svb* as a negative control. The recent method cisTargetX aims at detecting motifs enriched among a group of co-expressed genes, *e.g.* to predict direct targets of a TF [40]. It exploits a library of >3000 motifs, including TF binding sites and ultra-conserved DNA words [41], each motif being ranked with a score representative both of clustering and evolutionary conservation [40]. When applied to Svb targets, 4 of the top 5 motifs match the consensus CnGTT (Fig. 1B, S1C), characteristic of the Ovo/Svb BS CnGTTa as defined *in vitro* [35]. From the 39 input genes, CisTargetX determined an optimal subset of 16 Svb direct targets, having the highest scores for the OvoQ6 motif (Fig. 1B and S1C) [35, 36]. OvoQ6 was specific to Svb targets since it was not

detected in control epidermal genes (Fig. S1C). In contrast, motifs matching the BS of TFs involved in general epidermis differentiation such as Grainy head [42] or Vrille/c-EBP [43] were highly ranked in Svb independent genes (Fig. S1C), whilst detected only at low score in Svb downstream genes. Hence, OvoQ6 motifs appear as a signature of a subset of genes activated by Svb, a result consistent with their direct regulation.

Distribution of Svb BS clusters poorly correlates with enhancer activity.

We then examined the genomic distribution of OvoQ6 motifs within Svb target loci showing significant enrichment when compared to random *Drosophila* genes. We found that each target gene contained evolutionarily conserved OvoQ6 scattered throughout intergenic and intronic regions (Fig. 2A,B), instead of high OvoQ6 clusters enriched locally (even using relaxed conditions of at least 2 sites *per* kb). To delineate which regions mediate epidermal expression, we generated a series of transgenic reporters scanning systematically two Svb downstream genes. We focused on *singed* since it encodes Fascin, a conserved regulator of actin organization [44], and *shavenoid* that encodes a pioneer protein but displays an extreme trichome phenotype upon its inactivation [31]. Although most regions with OvoQ6 sites did not show embryonic expression, we identified three sequences one in *singed* (*snE1)* and two in *shavenoid* (*sha1* and *sha3)* that drove expression in the epidermis, specifically in trichome cells (Fig. 2A,B). Somehow unexpectedly, one of the three sequences, *sha1,* displays a single recognizable OvoQ6 motif (see below) in *D. melanogaster,* as well as in sibling species. The activity of all three regions was lost when introduced in *svb* null mutant background, showing that they are functional Svb target enhancers (Fig. 2A,B). cisTargetX predicts the location of putative

enhancers within each gene [40] and two out of three enhancers defined *in vivo* matched these predictions, in one case (*sha3*) to the highest rank for this gene (Fig. 2C). We therefore investigated whether evolutionarily conserved OvoQ6 sites were sufficient to predict trichome enhancers and assayed 18 additional regions (Fig. 2C) taken from the top 100 predictions. Transgenic reporter assays identified 4 novel sequences from *CG15589*, *cypher*, *dusky-like* and *neyo* driving expression in the epidermis. We verified in each case that they were specifically expressed in all (*dyl2, nyo1*) or subsets (*15589, cyrA*) of trichome cells where Svb is active. Consistently, these four enhancers depended on Svb since they displayed a strong reduction of their expression in the absence of *svb* (Fig. 2C). Hence, analysis of Svb downstream targets shows that they are enriched in OvoQ6 BS, a feature well conserved across *Drosophila* species. However, putative trichome enhancers predicted from evolutionary conservation and clustering of OvoQ6 sites were validated only at a rate of 28% (6/21, Fig. 2C), most tested regions being devoid of activity in embryos, suggesting that other criteria distinguishes enhancers from negative regions.

We noticed that OvoQ6 clusters failed to predict a number of active enhancers. This was the case of *sha1* (Fig. 2) or *Emin*, an epidermal enhancer previously identified for *miniature* [31]. Examination with Cluster-Buster [45] or Swan [46] did not detect supplementary OvoQ6 in *sha1* or *Emin* sequences (even in *D. melanogaster* only), explaining that these enhancers, containing a single Svb BS, escape from *in silico* predictions. 6 additional enhancers identified during initial stages of this study using alternative prediction criteria (see Fig. S1C) were not highly ranked by cisTargetX, because they lack BS clustering and/or evolutionary conservation. These data therefore show that BS clustering is not an absolute requisite for Svb regulation (Fig.

2C), suggesting that additional sites are required to discriminate between enhancers and inactive regions.

***De novo* motif discovery identifies a specific signature of Svb BS active *in vivo*.** To search for these putative sites, we compared the two sets of experimentally tested regions, *i.e.* the 14 enhancers (positive) and 25 inactive regions (negative), using *Imogene*, an algorithm designed for *de novo* motif discovery [47]. Briefly, we systematically searched, *ab initio*, for 10bp motifs that are evolutionarily conserved across *Drosophilidae* and display a distribution within each region statistically different from background sequences. We then evaluated how well each motif discriminated between enhancers and inactive regions and ranked these *de novo* motifs accordingly (Fig. 3A). Strikingly, the most discriminative motif overlaps OvoQ6 (CnGTTa), with a similar core consensus but extended to adjacent nucleotides (**A**CHGTT**AK**). A second discriminative motif (WAGAAAGCSR) called blue motif was also found, and will be studied below.

The **A**CHGTT**AK** motif, hereafter called svbF7, was sufficient to detect 10 out of 14 enhancers (Fig. 3B). The proportion of svbF7-positive enhancers reached 13/14, when relaxing the penalty imposed for poor conservation [47]. In contrast, svbF7 was found in only 6/25 negative regions (Fig. 3B), even when lowering the threshold (data not shown). Once added to the cisTargetX library, svbF7 is the most significant motif found in the set of 39 Svb downstream genes (Fig. S1C,D). It also increased the accuracy of enhancer predictions, with three additional positives *32159, Emin* and *EminB* while 9 negatives were removed from the top100 cisTargetX regions (Fig. S1C). Hence, svbF7 performs better than OvoQ6 or any other related motifs [48] (Fig. 3B,S1D). To evaluate whether this slight extension of the Svb BS

was relevant for activity, we substituted nucleotides flanking the core CnGTTa in the single svbF7 of *Emin*, *i.e.* altering the svbF7 motif without disrupting OvoQ6 consensus (Fig. 3C). When assayed *in vivo*, different patterns of flanking substitutions including a single point mutation of the 5' A residue were sufficient to strongly reduce *Emin* expression (Fig. 3C). This demonstrated the functional importance of flanking nucleotides within the svbF7 motif for CRM activity. Hence, the computational analysis of Svb-dependent enhancers has discovered a refined nucleotide sequence required for *in vivo* regulation.

Trichome enhancers use different combinations of cis-regulatory motifs.

Having shown the role of svbF7 in *Emin*, we investigated its functional significance in other enhancers. We focused on enhancers containing from 1 to 3 predicted SvbF7 sites, to address the importance of single versus clustered BS for trichome cell expression. As observed for *Emin*, disruption of the single svbF7 site abolished the activity both of *sha1* and of *nyo1* (Fig. 4A,B). The mutation of svbF7 also decreased the activity of *tyn2*, albeit weakly and only in ventral cells (arrowhead, Fig. 4C). In this enhancer, we detected however a second putative site that appears less conserved across species. Its inactivation strongly reduced expression (Fig. 4), showing that this site mainly contributes to *tyn2* activity. For *sha3* and *dyl2* that contain 2 or 3 svbF7 respectively, simultaneous inactivation of these sites abrogated expression (Fig. 4D-E). The individual disruption of svbF7 led nonetheless to varying defects. The two svbF7 sites of *sha3* are partly redundant, with a similar and limited impact of individual KO when compared to the simultaneous KO (Fig. 4D-G). In contrast, a single svbF7 plays a major role in *dyl2* activity, whereas the two others contribute marginally to expression pattern or levels (Fig. 4E,H). Hence, the disruption of svbF7

leads to a reduced expression for all enhancers that have been tested, confirming the functional importance of this motif. Nevertheless, the introduction of two copies of the svbF7 motif within negative regions (*sha2* and *12063*) was not sufficient to promote expression in trichome cells. In addition, the individual inactivation of multiple svbF7 sites has different consequences on enhancer activity, suggesting that additional elements are likely to modulate, locally, the *in vivo* function of svbF7.

We thus searched for additional cis-regulatory motifs and evaluated their contribution to the activity of trichome enhancers. In a first approach, we performed a systematic mutagenesis of the *Emin* enhancer by linker scanning (Fig. 5A). In addition to svbF7 whose inactivation abolished *Emin* activity (F7mt), the mutation of three regions (8mt, 9mt and 10mt) strongly decreased epidermal expression, two others (3mt, 4mt) affecting only the *Emin* pattern ventrally (Fig. 5A). These results show that while Svb acts as a main switch for *Emin* activity, other motifs are required for complete expression. Interestingly our *de novo* motif discovery identified a second discriminative motif (WAGAAAGCSR), hereafter called the blue motif, enriched in positive regions and evolutionarily conserved in 7 out of 14 enhancers (Fig. 3A,B, 5B). Mutations that disrupted the blue motif (9mt & 8mt) of *Emin* displayed the strongest effect, besides svbF7 KO (Fig. 5A). These unbiased data show that the blue motif represents an element that, in addition to svbF7, is critical for *Emin* activity. To further test its contribution to the activity of trichome enhancers, we mutated the blue motif in two other enhancers that contain a single occurrence of it (Fig. 5B). As observed for *Emin*, disruption of the blue motif reduced *snE1* expression (Fig. 5C). Furthermore, the blue motif plays a key role in *sha3* activity, its inactivation abolishing expression (Fig. 5C) similarly to the simultaneous mutation of both svbF7 sites (Fig. 4D). In addition, we noticed that one important region for *Emin* expression (10mt, Fig.

5A) matches a 8mer (TTATGCAA), previously predicted as a regulatory element from discovery of ultra-conserved DNA words in the genome of distant *Drosophila* species [41]. Although not sufficient by itself to discriminate between active enhancers and negative regions (data not shown), this motif, which we called yellow motif, was nevertheless retrieved in 6 additional trichome enhancers (Fig. 5B). To further assay *in vivo* the role of yellow motifs, we generated mutant versions of the *17058* & *nyo1* enhancers that disrupt their yellow motif. As observed for *Emin*, mutation of the yellow motif led to a strong decrease in the expression driven by both *nyo1* and *17058* (Fig. 5D), showing that the yellow motif represents a functional cis-regulatory element in a subset of enhancers.

Taken together, these data therefore support that svbF7 is a main feature of Svb targets, this motif being shared by the vast majority (13/14) of active enhancers. Our analyses have discovered two additional cis-regulatory elements, the blue and yellow motifs, present in overlapping subsets of trichome enhancers (9/14 & 7/14, respectively). While the 3 motifs show various patterns and combinations, functional assays demonstrated that each of them contributes to the *in vivo* activity of this sample of trichome enhancers.

Genome-wide prediction of Shavenbayby target enhancers.

To address whether these cis-regulatory motifs were a relevant signature of the genome-wide set of enhancers regulated by Svb, we undertook ChIP-seq to obtain an extensive cartography of Svb binding sites in epidermal cells. To improve specificity, we used a Svb::GFP transgene driven in ventral and dorsal trichome cells by two complementary *svb* cis-regulatory regions [34], likely at levels comparable to endogenous since it rescues *svb* mutant phenotypes [49]. ChIP-Seq data indicated

that Svb was bound to almost 6000 genomic sites, a large number of binding events being a feature shared by several *Drosophila* TFs [6, 8, 15]. Analysis of ChIP peaks with *i*-cisTarget [50] showed that svbF7 and OvoQ6 are the most enriched motifs. A strong cross correlation between conserved svbF7 and the center of ChIP peaks confirmed the importance of this motif (Fig. 6A). As observed in our pilot analysis of enhancers, we did not detect high svbF7 clustering, multiple svbF7 motifs being rarely found within genome-wide ChIP peaks. Blue motifs (and to a lesser extent yellow motifs) also displayed a significant but weaker correlation with Svb peaks, consistent with wider genomic distribution (Fig. 6A).

With the large number of Svb bound regions detected by ChIP-Seq, it was unlikely that all of them were functional, as being involved in the regulation of target genes [5, 15]. Therefore, in order to identify the entire set of genes regulated by Svb, we performed microarray profiling, comparing wild type to mutant embryos. In mRNA samples prepared from *svb* whole embryos, we often detected only a modest reduction in the levels of validated targets (Fig. 6B, S3), challenging unambiguous identification of Svb downstream genes. In the absence of *pri,* Svb behaves as a dominant repressor [34] and consistently we observed a stronger decrease in the levels of known Svb targets in *pri* mutants (Fig. 6B,C,S3), therefore providing an additional criteria to identify the genes regulated by Svb. Henceforth, we selected the genes down-regulated in *svb* mutants and that also displayed a further (>2 fold) reduction of their expression in *pri* mutants, as benchmarked for known Svb targets. This defined a set of 150 genes encompassing 16/39 Svb targets validated *in vivo* (Fig. S1A), as well as 42 additional epidermal candidates (Fig. S3). Among these, we examined 23 genes by *in situ* hybridization and confirmed that 21 of them required Svb to be expressed in trichome cells (Fig. 6B,C and S4). These results therefore

show that microarray profiling has defined a representative set of genes activated by Svb in trichome cells.

Focusing on this genomic set of Svb-regulated genes, we found 172 peaks associated with 85 genes (Fig. S3), including 11 out of 14 active enhancers (Fig. S7). Within the whole set of relevant Svb-bound regions, we retrieved the characteristic features of cis motifs as defined previously. Although retrieved in many Svb-bound regions (Fig. 6A, S5), the enrichment in yellow motifs within ChIP peaks associated with Svb regulated does not reliably reach a significant threshold, consistent with a broad genomic distribution [41]. In contrast, we found clear association of svbF7 motifs and to a lesser extent of blue motifs (Fig. S5). Importantly, these motifs were not detected in peaks associated with a control set of genes independent of Svb (Fig. S5), strongly supporting that they are hallmarks of Svb-target enhancers. As an independent way to evaluate this conclusion, we used *ab initio* analysis of ChIP peaks using PeakMotif [51]. This identified the motif ACAGTTA characteristic of peaks associated with Svb downstream genes that extensively matches svbF7 (Fig. S6). A second sequence (TGAAAAG) partly matching the blue motif was also detected in about 50% of peaks, again only in Svb-regulated genes and not among control genes (Fig. S6).

Hence, we interpret these results to imply that svbF7, and to a lesser extent the blue and/or yellow motifs, would allow predicting the location of additional trichome enhancers (Fig. 7A). To evaluate this, we tested ChIPed regions containing svbF7 alone *(12017, 14395)*, svbF7 in association with either the blue motif *(mey2, EminC, actn, 12017-2)* or the yellow motif *(31022, 4914),* or all three motifs together *(9095, 11175)* (Fig. 7B and S7). We found that 8/10 (80%) of these regions act as Svb-dependent enhancer when assayed *in vivo* (Fig. 7B). Indeed, they drove robust

expression, specifically in trichome cells, and their activity was reduced in *svb* mutant embryos (Fig. 7B). Moreover, these data confirm that trichome enhancers are generally built from different combinations of the three cis-regulatory motifs. For example, only a subset of newly predicted trichome enhancers relies on the blue motif, since *mey2*, *EminC, 9095* and *11175* contain conserved blue motifs whereas *12017, 31022* and *4914* do not (Fig. 7B , S7). In the case of the *actn* enhancer*,* there are four partly degenerate blue motifs in the sequence from *D. melanogaster* and sibling species, while not retrieved in more distant species suggesting a turnover of cis-regulatory motifs (Fig. S8). However, aside a couple of fast evolving enhancers, we found in many cases a remarkable conservation of the pattern of svbF7, blue and yellow motifs within individual enhancers across distantly related *Drosophila* species (Fig. 8, S8).

Therefore, the regulatory signatures learned from modeling and experimental dissection of a subset of enhancers helps understanding how the Svb TF selects the genomic set of its direct targets. Furthermore, they collectively allow efficient identification of cis-regulatory regions that specify the program of trichome-specific expression in response to Svb.

Discussion

It is well established that the Shavenbaby TF determined the trichome fate [29, 32, 52], however little was known on the repertoire of its direct target genes and mechanistic insights into the functional organization of trichome enhancers were lacking. Combining functional dissection, computational modeling and genome-wide profiling, we provide here a molecular map of the ultimate repertoire of genes and cis-regulatory elements implementing the network of trichome differentiation.

Physical elements of the GRN governing trichome formation

Our results identify a high-confidence set of more than 150 genes activated by Svb in trichome cells. We confirmed 60 of those, showing complete or partial down-regulation in the absence of active Svb protein. While most genes are expressed in all trichome cells, some are restricted to trichome subsets suggesting that they can contribute to the diversity in trichome shape and organization observed along the body [52]. Functional annotation (Gene Ontology and manual curation) indicates that Svb controls terminal players of trichome differentiation. In addition to novel factors of F-Actin organization [31, 39], ECM remodeling [31, 33], cuticle formation [31, 38] and pigmentation [31], we identify enzymes involved in oxidation-reduction, proteolysis and cell trafficking, further extending the repertoire of cellular functions involved in the terminal differentiation of trichome cells. Hence, a major role of Svb in trichome formation is to directly activate the expression of a battery of cell morphogenesis effectors. In support of this, ChIP-Seq peaks are present in >70% of these Svb-dependent effector genes. Experimental assays further validated 22 functional enhancers driving the expression of genes encoding factors involved in cytoskeletal or ECM reorganization, sugar binding, proteolysis and additional enzymes.

Recent work has established that apparently redundant, or shadow, enhancers ensure robust expression of transcription factors [53, 54]. For example, the transcription of *svb* itself involves separate enhancers that buffer the trichome pattern against variations in the genetic background and external conditions [53]. It has been proposed that shadow enhancers are required to drive an acute expression of some key developmental regulators [55]. We define within both *shavenoid* and *miniature* separable enhancers (*sha1*, *sha3* & *Emin, EminB, EminC*) that mediate Svb regulation. These data indicate that apparently redundant enhancers may not be limited to regulatory factors operating at high hierarchic positions in gene networks. Instead, we provide evidence that several "blue collar" effector genes display a similar regulatory architecture, suggesting that multiple enhancers represent an overlooked feature of the successive tiers of gene networks.

Binding site clustering as a general signature of active enhancers?

Early acting enhancers often comprise multiple BS for a given transcription factor [56, 57]. For example, conserved BS clusters have identified target enhancers of Dorsal [13] or Bicoid [58] and feature functional Twist-bound regions [15]. Of note, most algorithms developed for enhancer detection extensively use motif clustering as an important predictor [59]. We found a clear enrichment in putative Svb BS (OvoQ6 motif) in its downstream genes. However, only a small proportion of these motifs mediate *in vivo* regulation. There is very limited, if any, clustering of Svb BS in ChIP peaks associated with Svb target genes, and even genome-wide. Within the trichome enhancers we validated experimentally, 13 out of 22 display a single Svb site. Furthermore, for the enhancers *tyn2*, *sha3* and *dyl2* that contain 2-3 Svb BS, the inactivation of individual sites has often limited consequences, as also reported for

other TFs [60]. Even if some sites have been missed by computational approaches, the presence of multiple BS within a short region is not a deterministic feature of active Svb-dependent enhancers.

These findings highlight a paradoxical discrepancy between the enrichment of putative BS accumulated in Svb downstream genes and the limited number of those acting as cis-regulatory elements. Is there a role for this evolutionary accumulation of Svb-like motifs in Svb targets? For example, these sites presumably of weaker affinity (at least *in vivo*) can increase the local concentration of the TF facilitating regulation through a few BS stably bound *in vivo*, as it has been suggested on thermodynamics grounds [61] or to explain the existence of thousands of binding events that are transcriptionally inactive [5, 15].

Trichome enhancers rely on diverse combinations of cis-regulatory motifs

We found that the motif bound by Svb *in vivo* is more constrained than the consensus defined from *in vitro* [35] or one-hybrid approaches [48]. This shows that slight sequence differences, not detected *in vitro*, can play a key role within genomic context [62], for instance revealing the influence of co-factors [63].

In addition, other motifs influence which Svb BS are functional as regulatory elements, a notion well in line with recent results on the *in vivo* specificity of Hox factors [64]. Our statistical approaches identified a wider spread “blue” motif. Importantly, only half of the enhancers comprise blue motif(s), indicating that there are several ways to build Svb-responsive enhancers. Indeed, the systematic dissection of Emin disclosed an additional motif (TTATGCAA) ultra-conserved across *Drosophilidae* [41] and contributing to its activity. This “yellow” motif is retrieved in half of the trichome enhancers, with or without blue motifs. It is however barely

enriched in Svb-bound regions and therefore was not predicted by computational analyses, showing the importance of unbiased functional dissection to disclose the full spectrum of cis-regulatory elements. Indeed, the disruption of either blue or yellow motifs strongly affects enhancer function in all tested cases, providing experimental evidence of their cis-regulatory activity.

Trichome enhancers thus display various combinations of motifs, from those containing only Svb BS (5/22), Svb plus yellow (4/22), Svb plus blue (6/22) or all three together (7/22). These different motif compositions do not appear to correlate with distinct subclasses of gene function (DM unpublished). Furthermore, multiple enhancers from the same gene can harbor distinct combinations, as exemplified by *shavenoid* and to a lesser extent by *miniature* (Fig. 8, S6). Several studies have shown that motif composition may correlate with a given spatio-temporal pattern, *e.g.* for neurogenic or muscular GRNs [11, 16]. Since most trichome enhancers are often active in the very same population of cells, with highly similar dynamics, it is surprising to observe such diversity in their motif compositions. There are yet four enhancers restricted to dorsal trichome cells, but again they accommodate different motif compositions, with *EminB* & *4702B* that contain blue motifs *vs cyrA* & *31559* without. These data thus indicate that trichome enhancers display diverse distribution of functional motifs, supporting that distinct cis-regulatory architectures drive highly similar spatio-temporal expression.

Flexibility in cis-regulatory motifs among enhancers *vs* across species

Although highly constrained sequences, such as the interferon-B enhanceosome, do not seem widely spread [20], developmental enhancers may yet require some "grammar" for motif positioning [23], *e.g.* with an optimal pair-wise spacing of motifs

[64] that could reflect the cooperative binding of TFs. For trichome enhancers we did not detect any obvious bias in the number or respective arrangement of the cis-regulatory motifs they rely on (Fig. S2A). Likewise, recent results from the analysis of *Drosophila* cardiac enhancers support that similar expression patterns can be generated from divergent compositions and positioning of motifs [10, 65].

That several different inputs lead to similar enhancer outputs does not, however, formally rule out the existence of constraints, even though they are not detected by “horizontal” comparison of different enhancers within a same species. An independent way to evaluate this possibility is to look at the evolution of individual regulatory regions throughout species [15, 21]. Across *Drosophilidae*, trichome enhancers often display similar number and organization of cis-regulatory motifs (Fig. 8, S6). Furthermore, besides turnover of some motifs, svbF7, blue and yellow motifs are often imbedded within short-sized islands of high evolutionary conservation, when compared to neighboring sequences (Fig. 8). Similar strong evolutionary conservation was also noticed for the binding site of Twist [62] and its partner TFs [15], although these studies did not examine evolution of the detailed pattern of motif positioning. These data therefore suggest that despite diverse arrangements of motifs, patterns of evolutionary conservation likely represent the signature of functional constraints that locally shape the architecture of individual enhancers.

Materials and Methods

Fly strains and transgenic constructs

We used *btd, svb*^{1} or *svb*^{R9} [30, 31] and *pri*^{1} [34] stocks kept over GFP balancers. To delineate the epidermal enhancer of *sn* and *sha*, transgenic lines were initially generated using P-element mediated transformation (Fly Facility) and at least three independent insertions were analyzed for each construct. We then switched to the PhiC31 system (Bestgene) to quantify effects of mutations, with all constructs integrated at the same location (*zh-86F*), except for *sha1*, *sha3* and *snE1* for which mutant versions were assayed in P-elements for homogeneity (see supplemental methods). Genomic regions were amplified and cloned into pCasper or pAttB lacZ derivatives. QuickChangeXL site-directed mutagenesis (Stratagene) was used to introduce point mutations in enhancers, or CCGCCGGCGG stretches for linker scanning of *Emin*. All constructs were verified by sequencing.

Embryo staining

Dig- or biotin-labeled antisense RNA probes were used for *in situ* hybridization following standard protocols and embryos imaged using a Nikon Eclipse90i microscope. For immunodetection of lacZ reporter expression, 10-14h embryos were stained using anti-β–galactosidase (Cappel, 1/1000) and Alexafluor488 (Molecular Probes). Pictures were taken with a Leica SP2 confocal microscope, using the same settings to allow quantitative comparisons.

Microarrays

13-15h *svb*^{R9} or *pri*^{1} embryos were hand-selected using GFP balancers. 200 embryos were subjected to trizol (Invitrogen) extraction and RNA quality was monitored using Agilent Chip. Five independent samples of each genotype were used for microarrays (Affymetrix; IGBMC, Strasbourg). Data extraction and normalization were performed using Affymetrix software

and statistical analyses with R. A >2 fold difference in expression levels between mutant genotypes was the most efficient criteria to retrieve Svb downstream genes (with a false discovery rate (FDR) of 0.01 for *pri*). The top 150 genes downregulated both in *pri* and in *svb* mutants defined the set of Svb-regulated genes. 100 genes showing irrelevant variation of their expression (p-value>0.8, FDR>0.99) were used as a negative control set.

ChIP-seq

A *svb* rescue construct (RSQ8) {Kondo, 2010 #165} was used for ChIP-seq experiments. It expresses a Svb-GFP protein under the control of two *svb* enhancers (medial and proximal) driving specific in epidermal trichome cells. Stocks were expanded to fill three population cages. Adults were allowed to lay eggs for 2 hours on apple juice plates covered with yeast. Embryos deposited on the plates were aged for 12 hours at 25°C. Chromatin was collected from approximately 100 mg of whole embryos for each replicate chromatin collection. Chromatin immunoprecipitation was done with an anti-GFP antibody as described [8]. Data presented are from two independent replicates. Peaks were called for single replicates using MACS p<0.00001 for downstream computational analyses. MACS was used to call loose criteria peaks for two replicates of RSQ8 12-14 hr embryo. Those peaks were then used for an IDR analysis, IDR=0.02. DNA sequencing libraries were generated with Nextera DNA Sequencing Libarary Kits. Additional details are given in supplemental methods.

Motif detection and genome analysis

Detection of motifs enriched in Svb-dependent and Svb-independent epidermal genes was performed using *cisTargetX* [40]. For *de novo* motif discovery, genomic sequences of enhancer and negative regions were processed through a C++ program and statistical operations performed within the R software, as described [47]. To compute the cross-correlation between conserved motif instances and Svb ChIP-Seq data, we defined a 10 Kb window centered around each ChIP peak, collected distances of each motif to the peak center and plotted these values using a 500bp bin. In the cases of Svb-regulated and control

genes, each ChIP peak was associated with the nearest transcription start site. Additional details are given in supplemental information.

Acknowledgements

We are grateful to the Bloomington Drosophila Stock Center and Drosophila Genomic Resource Center for providing us with flies and molecular clones. We are indebted to B. Ronsin (Toulouse RIO Imaging), P. Valenti and O. Bohner for excellent technical assistance. We also thank C. Hermann, N. Negre, E. Preger-Ben Noon, A. Vincent and members of the SP/FP lab for critical reading of the manuscript. This work was supported by grants from ANR Blanc “Netoshape”, Association pour la Recherche contre le Cancer (n°3832, n°1111, SFI20101201669 and fellowships to IF & DM) and University Paul Sabatier (cisdecode). SA was supported by grants from FWO (G.0704.11N and G.0640.13) and HFSP (RGY0070/2011). SA, FP and SP designed the experiments, MS, HR and VH statistical analyses; DM, MS, IF, RS, JZ, IG, YL, PF, SA & SP performed the experiments and are listed according to their contributions. SP and FP wrote the paper and all authors commented on the manuscript. The authors declare no conflict of interest.

FIGURE LEGENDS

Figure 1: Enrichment in binding sites defines an evolutionarily conserved signature of *svb* downstream genes

A: Expression of *svb* mRNA determines the epidermal cells that form trichomes, visible on the larval cuticle. *In situ* hybridization shows mRNA expression of two *svb* downstream genes, *shavenoid (sha/koj)* and *CG15589*, in wild type (top) and *svb* mutant embryos (bottom). **B**: ROC curve showing significant enrichment in putative Svb binding sites (OvoQ6 position weight matrix) among the 39 Svb downstream genes (y axis) compared to a randomized set of 1000 *Drosophila* genes (x axis) using cisTargetX. The blue curve shows the detection of Svb downstream genes, the red one a random distribution, and the green curve shows a 2 sigma interval from random.

Figure 2: A subset of Svb binding sites correspond to functional enhancers

Svb-dependent trichome enhancers were identified by transgenic reporter gene assays, from a systematic scanning of the *singed (sn)* (A) and *shavenoid (sha)* (B) genes and regions predicted by cisTargetX (C). **A,B**. Vertical black lines represent evolutionarily conserved OvoQ6 clusters (at least two motifs in a 1kb window), as predicted by cistargetX. Horizontal boxes summarize regions tested by transgenic assays, using immunostaining of the lacZ reporter (green). Negative regions (pink) do not drive specific expression in the embryonic epidermis. Regions drawn in cyan display enhancer activity reproducing endogenous expression in trichome cells (as assayed by mRNA *in situ* hybridization, purple). The *snE1* and *sha3* enhancers are under the control of Svb, as demonstrated by a reduced expression in *svb* mutants. **C**. Putative enhancers (CRMs) predicted by cis-TargetX, from clustering and/or evolutionary conservation of OvoQ6 sites. Pictures show expression of positive enhancers (cyan) in wild type (top) and *svb* mutant (bottom) embryos. Additional regions (pink) showed no detectable activity during embryogenesis.

Figure 3: Computational analyses allow refining functional Svb binding sites

A. Statistical analysis of positive enhancers versus negative regions (intergenic genomic sequences used as background) was performed for *de novo* discovery of motifs, showing evolutionary conservation across *Drosophila* species and characteristic of active enhancers. **B.** The svbF7 and blue motifs perform best in discriminating between positive enhancers and negative regions as illustrated by the Pareto plot. **C**. While disruption of the core CnGTT OvoQ6 motif abolish *Emin* activity, point mutations that affect the 5' and 3' flanking nucleotides strongly reduce epidermal expression, as shown by anti-lacZ immuno-staining and quantification of fluorescence signals.

Figure 4: *In vivo* role of svbF7 motifs in Svb-dependent enhancers

Anti-LacZ staining (green) shows modifications of reporter gene expression resulting from individual and simultaneous inactivation of svbF7 motifs in *sha1* (**A**), *nyo1* (**B**), *tyn2* (**C**), *sha3* (**D**) and *dyl2* (**E**) enhancers. **F-H** Quantification of residual activity following individual disruption of svbF7 sites. Red boxes schematize evolutionarily conserved svbF7 motifs; the open black box, a site that does not appear conserved across *drosophilidae*; *** indicates a p-value <0,01, *<0,05.

Figure 5: Svb-dependent enhancers use various combinations of cis-regulatory elements

A. Linker-scanning mutagenesis of *Emin* identifies other 10bp regions required for full transcriptional activity, as deduced from altered pattern of lac-Z immuno-staining (green). Positions of SvbF7, blue and yellow motifs are indicated on top. **B**. Black, red, blue and yellow boxes schematize the distribution, number and orientation of OvoQ6, svbF7, blue and yellow motifs, respectively. Filled boxes represent motifs conserved across *Drosophila* species; open boxes those detected only in *D. melanogaster*. **C**. Point mutations that disrupt the blue motif in *Emin*, *snE1* and *sha3* reduce the activity of all three enhancers, to 40 +/-14

% (*), 44 +/- 6% (***), 8 +/- 4% (***) of wild-type levels, respectively. **D**. Point mutations that disrupt the yellow motif in *Emin*, *17058* and *nyo1* reduce the activity of all three enhancers, to 20 +/- 7 % (*), 16 +/- 8% (*), 6 +/- 3% (*) of wild-type levels, respectively. *** indicates a p-value <0,001, *<0,03.

Figure 6: Genome-wide profiling of embryonic genes regulated by Svb

A. Cross-correlation between conserved svbF7, blue or yellow motif instances and Svb ChIP-Seq peaks throughout the whole genome. Plots show the number of motifs found in a 10kb window on each side of the center of peaks. The p-value for correlation (Chi2 test) is <1E-46, <1E-9 and <1E-2, respectively. **B**. Modifications in mRNA levels as measured by microarrays between wild-type and *svb* (left) or *pri* (right) embryos in Svb-regulated (green) and control (blue) set of genes. Dark green dots represent known Svb targets (Fig. S1A), light green novel target genes as validated by *in situ* hybridization (Fig. S4) and open dots additional candidates. **C**. Whole mount *in situ* hybridization of *CG1273*, a Svb downstream target identified from microarray profiling, down regulated in trichome cells of *svb* mutants and showing a further reduced expression in *pri* mutant embryos.

Figure 7: Identification of Svb direct targets and their trichome enhancers using computational and *in vivo* experimental approaches

A. Chart flow diagram summarizing the pipeline used for enhancer prediction and validation. **B.** Motif distribution coupled to ChIP-seq allows predicting location of enhancers in Svb downstream targets. Graphs show ChIP intensity at the time of trichome formation (12-14h of embryogenesis). Active enhancers are drawn as cyan rectangles. Pictures show reporter gene expression driven by corresponding regions in wild type (*wt*) and *svb* mutant embryos, as revealed by anti-lacZ immunostaining (green). The composition, orientation and respective positioning of svbF7 (red), blue and yellow motifs is schematized by filled (evolutionarily conserved) and open (not traceable across species) boxes.

Figure 8. Pattern of evolutionary conservation of the three enhancers driving *miniature* expression in trichome cells

The position of epidermal enhancers is shown by cyan boxes and their respective architecture with respect to svbF7 (red), blue and yellow motifs are schematized across *Drosophila* species. Orthologous sequences were identified by BLAST and manually adjusted for optimized alignment. Motif search was performed in individual sequences taken independently, using a same threshold for each motif in all cases. Bottom histograms represent the pattern of evolutionary conservation across *Drosophila* species, focusing on individual regions harboring identified cis-regulatory motifs (color coded).

REFERENCES

1. Stathopoulos A, Levine M: **Genomic regulatory networks and animal development.** *Dev Cell* 2005, **9:**449-462.

2. Ptashne M: **Regulation of transcription: from lambda to eukaryotes.** *Trends Biochem Sci* 2005, **30:**275-279.

3. Rister J, Desplan C: **Deciphering the genome's regulatory code: the many languages of DNA.** *Bioessays* 2010, **32:**381-384.

4. Yanez-Cuna JO, Kvon EZ, Stark A: **Deciphering the transcriptional cis-regulatory code.** *Trends Genet* 2013, **29:**11-22.

5. Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Luengo Hendriks CL, et al: **Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm.** *PLoS Biol* 2008, **6:**e27.

6. Slattery M, Negre N, White KP: **Interpreting the regulatory genome: the genomics of transcription factor function in Drosophila melanogaster.** *Brief Funct Genomics* 2012, **11:**336-346.

7. MacArthur S, Li XY, Li J, Brown JB, Chu HC, Zeng L, Grondona BP, Hechmer A, Simirenko L, Keranen SV, et al: **Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions.** *Genome Biol* 2009, **10:**R80.

8. Negre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, Kheradpour P, Eaton ML, Loriaux P, Sealfon R, et al: **A cis-regulatory map of the Drosophila genome.** *Nature* 2011, **471:**527-531.

9. Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U, Gaul U: **Predicting expression patterns from regulatory sequence in Drosophila segmentation.** *Nature* 2008, **451:**535-540.

10. Junion G, Spivakov M, Girardot C, Braun M, Gustafson EH, Birney E, Furlong EE: **A transcription factor collective defines cardiac cell fate and reflects lineage history.** *Cell* 2012, **148:**473-486.

11. Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EE: **Combinatorial binding predicts spatio-temporal cis-regulatory activity.** *Nature* 2009, **462:**65-70.

12. Zeitlinger J, Zinzen RP, Stark A, Kellis M, Zhang H, Young RA, Levine M: **Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo.** *Genes Dev* 2007, **21:**385-390.

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

14. Sinha S, Schroeder MD, Unnerstall U, Gaul U, Siggia ED: **Cross-species comparison significantly improves genome-wide prediction of cis-regulatory modules in Drosophila.** *BMC Bioinformatics* 2004, **5:**129.

15. He Q, Bardet AF, Patton B, Purvis J, Johnston J, Paulson A, Gogol M, Stark A, Zeitlinger J: **High conservation of transcription factor binding and evidence for combinatorial regulation across six Drosophila species.** *Nat Genet* 2011, **43:**414-420.

16. Markstein M, Zinzen R, Markstein P, Yee KP, Erives A, Stathopoulos A, Levine M: **A regulatory code for neurogenic gene expression in the Drosophila embryo.** *Development* 2004, **131:**2387-2394.

17. Khoueiry P, Rothbacher U, Ohtsuka Y, Daian F, Frangulian E, Roure A, Dubchak I, Lemaire P: **A cis-regulatory signature in ascidians and flies, independent of transcription factor binding sites.** *Curr Biol* 2010, **20:**792-802.

18. Erives A, Levine M: **Coordinate enhancers share common organizational features in the Drosophila genome.** *Proc Natl Acad Sci U S A* 2004, **101:**3851-3856.

19. Levine M: **Transcriptional enhancers in animal development and evolution.** *Curr Biol* 2010, **20:**R754-763.

20. Spitz F, Furlong EE: **Transcription factors: from enhancer binding to developmental control.** *Nat Rev Genet* 2012, **13:**613-626.

21. Crocker J, Erives A: **A closer look at the eve stripe 2 enhancers of Drosophila and Themira.** *PLoS Genet* 2008, **4:**e1000276.

22. Papatsenko D, Goltsev Y, Levine M: **Organization of developmental enhancers in the Drosophila embryo.** *Nucleic Acids Res* 2009, **37:**5665-5677.

23. 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:**359-370.

24. Rowan S, Siggers T, Lachke SA, Yue Y, Bulyk ML, Maas RL: **Precise temporal control of the eye regulatory gene Pax6 via enhancer-binding site affinity.** *Genes Dev* 2010, **24:**980-985.

25. Papatsenko D, Levine M: **A rationale for the enhanceosome and other evolutionarily constrained enhancers.** *Curr Biol* 2007, **17:**R955-957.

26. Crocker J, Tamori Y, Erives A: **Evolution acts on enhancer organization to fine-tune gradient threshold readouts.** *PLoS Biol* 2008, **6:**e263.

27. Etchberger JF, Lorch A, Sleumer MC, Zapf R, Jones SJ, Marra MA, Holt RA, Moerman DG, Hobert O: **The molecular signature and cis-regulatory architecture of a C. elegans gustatory neuron.** *Genes Dev* 2007, **21:**1653-1674.

28. Laurencon A, Dubruille R, Efimenko E, Grenier G, Bissett R, Cortier E, Rolland V, Swoboda P, Durand B: **Identification of novel regulatory factor X (RFX) target genes by comparative genomics in Drosophila species.** *Genome Biol* 2007, **8:**R195.

29. Payre F, Vincent A, Carreno S: **ovo/svb integrates Wingless and DER pathways to control epidermis differentiation.** *Nature* 1999, **400:**271-275.

30. Sucena E, Delon I, Jones I, Payre F, Stern DL: **Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism.** *Nature* 2003, **424:**935-938.

31. Chanut-Delalande H, Fernandes I, Roch F, Payre F, Plaza S: **Shavenbaby couples patterning to epidermal cell shape control.** *PLoS Biol* 2006, **4:**e290.

32. Chanut-Delalande H, Ferrer P, Payre F, Plaza S: **Effectors of tridimensional cell morphogenesis and their evolution.** *Semin Cell Dev Biol* 2012, **23:**341-349.

33. Fernandes I, Chanut-Delalande H, Ferrer P, Latapie Y, Waltzer L, Affolter M, Payre F, Plaza S: **Zona pellucida domain proteins remodel the apical compartment for localized cell shape changes.** *Dev Cell* 2010, **18:**64-76.

34. 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:**336-339.

35. Lee S, Garfinkel MD: **Characterization of Drosophila OVO protein DNA binding specificity using random DNA oligomer selection suggests zinc finger degeneration.** *Nucleic Acids Res* 2000, **28:**826-834.

36. Lu J, Oliver B: **Drosophila OVO regulates ovarian tumor transcription by binding unusually near the transcription start site.** *Development* 2001, **128:**1671-1686.

37. Brown CD, Johnson DS, Sidow A: **Functional architecture and evolution of transcriptional elements that drive gene coexpression.** *Science* 2007, **317:**1557-1560.

38. Andrew DJ, Baker BS: **Expression of the Drosophila secreted cuticle protein 73 (dsc73) requires Shavenbaby.** *Dev Dyn* 2008, **237:**1198-1206.

39. Bejsovec A, Chao AT: **crinkled reveals a new role for Wingless signaling in Drosophila denticle formation.** *Development* 2012, **139:**690-698.

40. Aerts S, Quan XJ, Claeys A, Naval Sanchez M, Tate P, Yan J, Hassan BA: **Robust target gene discovery through transcriptome perturbations and genome-wide enhancer predictions in Drosophila uncovers a regulatory basis for sensory specification.** *PLoS Biol* 2010, **8:**e1000435.

41. Elemento O, Tavazoie S: **Fast and systematic genome-wide discovery of conserved regulatory elements using a non-alignment based approach.** *Genome Biol* 2005, **6:**R18.

42. Mace KA, Pearson JC, McGinnis W: **An epidermal barrier wound repair pathway in Drosophila is mediated by grainy head.** *Science* 2005, **308:**381-385.

43. Szuplewski S, Kottler B, Terracol R: **The Drosophila bZIP transcription factor Vrille is involved in hair and cell growth.** *Development* 2003, **130:**3651-3662.

44. Jayo A, Parsons M: **Fascin: a key regulator of cytoskeletal dynamics.** *Int J Biochem Cell Biol* 2012, **42:**1614-1617.

45. Frith MC, Li MC, Weng Z: **Cluster-Buster: Finding dense clusters of motifs in DNA sequences.** *Nucleic Acids Res* 2003, **31:**3666-3668.

46. Kim J, Cunningham R, James B, Wyder S, Gibson JD, Niehuis O, Zdobnov EM, Robertson HM, Robinson GE, Werren JH, Sinha S: **Functional characterization of transcription factor motifs using cross-species comparison across large evolutionary distances.** *PLoS Comput Biol* 2010, **6:**e1000652.

47. Rouault H, Mazouni K, Couturier L, Hakim V, Schweisguth F: **Genome-wide identification of cis-regulatory motifs and modules underlying gene coregulation using statistics and phylogeny.** *Proc Natl Acad Sci U S A* 2010, **107:**14615-14620.

48. Zhu LJ, Christensen RG, Kazemian M, Hull CJ, Enuameh MS, Basciotta MD, Brasefield JA, Zhu C, Asriyan Y, Lapointe DS, et al: **FlyFactorSurvey: a database of Drosophila transcription factor binding specificities determined using the bacterial one-hybrid system.** *Nucleic Acids Res* 2011, **39:**D111-117.

49. 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:**598-603.

50. Herrmann C, Van de Sande B, Potier D, Aerts S: **i-cisTarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules.** *Nucleic acids research* 2012, **40:**e114.

51. Thomas-Chollier M, Defrance M, Medina-Rivera A, Sand O, Herrmann C, Thieffry D, van Helden J: **RSAT 2011: regulatory sequence analysis tools.** *Nucleic acids research* 2011, **39:**W86-91.

52. Payre F: **Genetic control of epidermis differentiation in Drosophila.** *Int J Dev Biol* 2004, **48:**207-215.

53. Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL: **Phenotypic robustness conferred by apparently redundant transcriptional enhancers.** *Nature* 2010, **466:**490-493.

54. Perry MW, Boettiger AN, Bothma JP, Levine M: **Shadow enhancers foster robustness of Drosophila gastrulation.** *Curr Biol* 2010, **20:**1562-1567.

55. Hong JW, Hendrix DA, Levine MS: **Shadow enhancers as a source of evolutionary novelty.** *Science* 2008, **321:**1314.

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

57. Schroeder MD, Pearce M, Fak J, Fan H, Unnerstall U, Emberly E, Rajewsky N, Siggia ED, Gaul U: **Transcriptional control in the segmentation gene network of Drosophila.** *PLoS Biol* 2004, **2:**E271.

58. Chen H, Xu Z, Mei C, Yu D, Small S: **A system of repressor gradients spatially organizes the boundaries of Bicoid-dependent target genes.** *Cell* 2012, **149:**618-629.

59. Aerts S: **Computational strategies for the genome-wide identification of cis-regulatory elements and transcriptional targets.** *Curr Top Dev Biol* 2012, **98:**121-145.

60. Halfon MS, Zhu Q, Brennan ER, Zhou Y: **Erroneous attribution of relevant transcription factor binding sites despite successful prediction of cis-regulatory modules.** *BMC Genomics* 2011, **12:**578.

61. Walter J, Biggin MD: **DNA binding specificity of two homeodomain proteins in vitro and in Drosophila embryos.** *Proc Natl Acad Sci U S A* 1996, **93:**2680-2685.

62. Ozdemir A, Fisher-Aylor KI, Pepke S, Samanta M, Dunipace L, McCue K, Zeng L, Ogawa N, Wold BJ, Stathopoulos A: **High resolution mapping of Twist to DNA in Drosophila embryos: Efficient functional analysis and evolutionary conservation.** *Genome Res* 2011, **21:**566-577.

63. Slattery M, Riley T, Liu P, Abe N, Gomez-Alcala P, Dror I, Zhou T, Rohs R, Honig B, Bussemaker HJ, Mann RS: **Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins.** *Cell* 2011, **147:**1270-1282.

64. Sorge S, Ha N, Polychronidou M, Friedrich J, Bezdan D, Kaspar P, Schaefer MH, Ossowski S, Henz SR, Mundorf J, et al: **The cis-regulatory code of Hox function in Drosophila.** *EMBO J* 2012, **31:**3323-3333.

65. Jin H, Stojnic R, Adryan B, Ozdemir A, Stathopoulos A, Frasch M: **Genome-wide screens for in vivo tinman binding sites identify cardiac enhancers with diverse functional architectures.** *PLoS Genet* 2013, **9:**e1003195.