

Decoding the transcriptional program of epidermal cell morphogenesis

Delphine MENORET^{1,2*}, Marc SANTOLINI^{3*}, Isabelle FERNANDES^{1,2,4}, Rebecca SPOKONY⁵, Jennifer ZANET^{1,2}, Ignacio GONZALEZ^{6,7}, Yvan LATAPIE^{1,2}, Pierre FERRER^{1,2}, Hervé ROUAULT^{3,9}, Kevin P. WHITE⁵, Philippe BESSE^{6,7}, Vincent HAKIM³, Stein AERTS⁸, 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, ENS, Paris, France

⁴ Present address: Department of Biology, McGill University, 1205 Dr.16 Penfield Avenue, Montreal, QC, Canada H3A 1B1.

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

⁶ University of Toulouse, INSA, Toulouse, France

⁷ Institut de Mathématiques, CNRS UMR5219, Toulouse, France

⁸ Laboratory of Computational Biology, KU Leuven, Belgium

⁹ present address: Institut Pasteur, Developmental Biology Department, Paris, France

* The two first authors contributed equally

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

Running title: transcriptional control in epidermis

Abstract

Developmental programs are implemented by regulatory interactions between Transcription Factors (TFs) and their target genes (TGs). How the Cis-Regulatory-Modules (CRMs) mediating these interactions are built and function remains yet poorly understood, especially during terminal differentiation. We addressed this question during late *Drosophila* embryogenesis when the finely tuned expression of a Transcription Factor, Ovo/Shavenbaby (Svb), triggers the morphological differentiation of epidermal trichomes. Here we find that Svb regulates a large set of terminal effectors of trichome formation, as deduced from microarray profiling and the experimental validation of 60 target genes. Combining genome-wide approaches, computational modelling and *in vivo* functional dissection, we investigated the nature and logic of CRMs directing the expression of Svb-dependent effectors. Challenging common views of CRM organization, we find that Svb-responsive CRMs display weak if any clustering of Svb binding sites. In addition, the *in vivo* function of each site relies on its intimate context, with a critical importance of adjacent nucleotides. Finally, Svb-responsive CRMs display various combinations of additional cis-regulatory elements, which contribute to different levels of activity. Together, these results show that trichome formation is underpinned by an unexpectedly flexible mode of regulation, shedding novel light on the functional organization of CRMs mediating terminal differentiation.

Introduction

Initially pioneered by genetic analysis of *Drosophila* embryogenesis, many studies have well established that transcriptional networks control developmental processes, through determining specific programs of genome expression [1]. These gene regulatory networks 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 (TGs) via their DNA binding properties [2], eukaryotic TFs recognize only short motifs, generally with a degenerated specificity. Consequently, thousands of putative Binding Sites (BS) are scattered throughout the genome hampering efficient prediction of cis-regulatory elements [3,4]. The fine structure of CRMs as well as the general rule(s) underlying their organization remain therefore poorly understood and rely on experimental cases.

Although higher eukaryotes encode hundreds to over thousand of TFs, only a few TFs have been studied in detail to elucidate the regulatory logic of their target CRMs [5,6]. In *Drosophila*, current models of CRM structure mainly come from works on early development, e.g. TFs controlling segmentation and mesoderm specification [7,8,9,10,11]. A general scheme emerging from these studies is that local enrichment for BS of a given TF (homotypic clustering) in evolutionary conserved regions is a relevant signature of CRMs [12,13,14]. It has been suggested that CRMs driving similar expression are built by a combinatorial code of homo- and heterotypic BS clustering, together defining a specific program of genome expression [10,15,16,17] [18]. Whether a functional CRM relies on a constrained organization of different BS (enhanceosome model) or can accommodate flexibility in the number and/or respective arrangement of BS (billboard model) remains a question under debate

[19]. While comparative studies between species have brought support to a possible flexibility of BS arrangement, e.g. for the eve stripe 2 enhancer [20,21], experimental manipulations show that even subtle variations in BS affinity can lead to strong modifications in the expression driven by a CRM involved in eye patterning [22,23,24]. It is difficult to predict to which extent the conclusions drawn from these dissections of individual CRM can be extrapolated to a broader range of cis-regulatory elements. On the other hand, the recent development of genome-wide ChIP-chip or ChIP-seq approaches has uncovered thousands of regions bound by TFs during early development [10,25], but which fraction of them represent active CRM and what is their regulatory code require additional functional studies. Moreover, it remains unclear whether the specific landscape of early segmentation genes, which are regulated by diffusible TFs in a syncitial environment, has influenced the functional organization of segmentation CRMs. Therefore a general question is to gain detailed insights into the structure and regulatory logic of CRM/TF interactions that occurs at later developmental stages, e.g. to govern terminal differentiation [26].

Here we focus on a *Drosophila* gene regulatory network that controls cell morphogenesis during terminal differentiation of the embryonic epidermis, a process governed by the TF Ovo/Shavenbaby (Svb) [27]. The subset of epidermal cells that express Svb undergo localized changes in cell shape leading to the formation of dorsal hairs and ventral denticles, collectively referred to as trichomes [28,29]. The epidermal expression of *svb* is finely regulated by a large array of at least 6 cis-regulatory regions (spanning more than 100kb) [30,31]. They collectively integrate upstream regulatory cascades to define the precise subset of trichome cells [27] and provide phenotypic robustness of the trichome pattern to variations in external

conditions and genetic backgrounds [31]. The activity of Svb is further regulated in a post-translational manner, in response to small peptides encoded by the atypical gene *polished-rice (pri)* [32]. Pri peptides trigger N-terminal truncation of the Svb TF, switching its activity from a repressor (full length) to an activator (cleaved) protein [32], therefore providing a temporal control to the program of trichome formation. Once activated, Svb triggers the expression of downstream genes encoding cellular effectors that directly participate in the remodelling of epidermal cells [33,34,35,36].

While evolutionary studies have further demonstrated the pivotal role of Svb in determining the trichome pattern [27,30,31,37,38,39], little is known concerning how this TF recognizes and selects specific cell effectors. Besides definition of DNA-binding specificity *in vitro* [40] and the identification of a couple of genes regulated by Ovo germline-specific isoforms [40,41], a single epidermal CRM dependent on Svb has been identified so far [35]. Combining transcriptome profiling to systematic *in vivo* functional assays, we identified an unbiased set of Svb downstream genes (60 of them validated *in situ*) and investigated the functional organization of CRMs driving their expression in trichome cells. Experimental identification and computational analyses of 14 Svb-responsive CRMs show that, in addition to varying number of Svb BS, these regulatory elements do not follow a similar combinatorial code. *De novo* motif discovery coupled to functional dissection establishes the importance of the intimate context of each Svb BS to be active *in vivo*, together with the presence of various combinations of additional cis-regulatory motifs. Models of CRM organization learned from these experiments further help identifying Svb-responsive regulatory modules, when used to analyze genome-wide data from ChIP-seq and transcriptome profiling experiments.

All together, these data thus show that the program of epidermal trichomes is hard-wired by CRMs that display a functional organization notably different from the CRMs underpinning A/P and D/V patterning, suggesting the existence of different regulatory logics between early-acting and terminal differentiation Gene Regulatory Networks during *Drosophila* embryogenesis.

Results

Global enrichment and evolutionary conservation of putative Svb binding sites in downstream genes

It has been previously shown that the role of Svb in trichome formation requires the transcriptional activation of 18 downstream genes directly involved in actin reorganization, extracellular matrix remodeling and cuticle sclerotization/pigmentation [33,34,35,36]. The whole register and molecular mode of action of this TF remained yet elusive and, as a first step, we sought to identify a larger set of Svb downstream genes.

We therefore systematically analyzed 57 additional candidate genes, selected because of their documented expression in subsets of epidermal cells (Berkeley Drosophila Genome Project database), using *in situ* hybridization in wild type versus *svb* mutant embryos. We found 21 novel Svb-dependent genes, *i.e.*, showing reduced expression in *svb* mutants and ectopic expression when *svb* was artificially expressed in supernumerary epidermal cells (Fig. 1A, S1A and Supplementary Information). Interestingly, most of these genes encode enzymes (eg redox), additional components of the extracellular matrix or of the cytoskeleton, broadening the conclusion that Svb regulates terminal effectors of cell differentiation. The 36 other genes we analyzed were expressed in epidermal cells in an Svb independent manner (Fig. S1B). We then used this extended set of 39 Svb downstream genes to examine whether they display an evolutionary conserved signature in their non-coding regions, when compared to random *Drosophila* genes or epidermal genes independent of *svb* activity.

The recently developed method *cis*TargetX aims at detecting significant enrichment of evolutionarily conserved DNA motifs within non protein coding regions, among a group of co-expressed genes, *e.g.* to predict direct targets of a TF [42,43]. This exploits a library of >3000 DNA motifs, including TF binding sites (from various species), as well as short DNA words that display strong conservation throughout the evolution of *Drosophila* species [44,45]. Each motif is individually ranked with a score representative both of clustering and evolutionary conservation, the latter relying on non-aligned comparison of orthologous regions [42,43]. We applied *cis*TargetX to the 39 Svb downstream genes and, as a control, to the 36 epidermal genes that are not regulated by Svb. In Svb downstream genes, 4 of the top 5 motifs match the consensus CnGTT (Fig. 1B and S1C) that characterizes the Ovo/Svb TF binding site as initially defined *in vitro* [40]. The Position Weight Matrix (PWM) OvoQ6 from TRANSFAC (ID M01101) [40,41] yielded an optimal set of 16 predicted direct targets out of the 39 Svb downstream genes (Fig. 1B and S1C). Among epidermal genes, enrichment of this motif appears specific to Svb-dependent genes since it was not detected in epidermal genes not regulated by Svb (Fig. S1D). In contrast, motifs matching the BS of TFs involved in general epidermis differentiation, such as Grainy head [46,47] or Vrille/c-EBP [48,49], were highly ranked in Svb non-regulated genes, whilst detected in Svb downstream genes only at low score (Fig. S1C,D). Hence, OvoQ6 appears as a specific signature of a subset of genes regulated by Svb, a result consistent with their direct regulation.

Distribution of Ovo/Svb binding sites clusters poorly correlates with CRM activity.

We then examined the genomic distribution of putative Svb/Ovo BS and found a significant enrichment of evolutionary conserved OvoQ6 motifs, in upstream and intronic regions of Svb TGs. However, these OvoQ6 motifs are generally scattered throughout the entire non-coding regions of a given gene (Fig. 2A,B) and only a few displayed clustering, even using relaxed conditions (*i.e.*, at least 2 occurrences by 1kb window). If these clusters were signatures of Svb-dependent CRMs, this suggests that Svb TGs may have multiple regulatory elements. To delineate the region(s) mediating *in vivo* epidermal expression, we undertook a systematic *in vivo* analysis of transgenic reporters, scanning the genomic region of two Svb downstream genes. *singed* encodes the fly Fascin, a well known regulator of actin reorganization [50], and *shavenoid* encodes a pioneer protein that displays a dramatic trichome phenotype when inactivated [51]. Although 8 regions containing OvoQ6 did not contribute to specific expression, we identified three small regions (<600bp), one in a large intron of *singed* and two in the *shavenoid* promoter that drove expression in trichome cells (Fig. 2A,B). Importantly, their expression was lost in the absence of *svb*, showing that these regions act as functional Svb-dependent CRM (Fig. 2A,B). *cisTargetX* also predicts the location of putative CRMs within each gene for the OvoQ6 PWM, with a score calculated from local clustering and evolutionary conservation in flies [42]. Two out of three *in vivo* defined CRMs matched *cisTargetX* predictions, in one case (sha3) to the highest ranked region within this gene. We therefore investigated whether these *in silico* analyses can predict additional CRMs and assayed 18 other regions (Fig. 2C), taken from the top 100 *cisTargetX* predictions. Transgenic reporter analyses identified 4 additional CRMs driving specific expression in trichome cells, under *svb* control (Fig. 2C). In summary, 6 out of 21 (28%) OvoQ6 predicted regions represent *bona fide* Svb-

dependent CRM, a result comparable to the discovery rate of active CRM from ChIP that is often around 20 % of the total number of peaks [10,11].

De novo motif discovery identifies specific signatures of Svb-dependent CRMs.

While evolutionarily conserved OvoQ6 clusters appear prognostic of some Svb-dependent CRM, they failed however to predict a number of active CRMs. This was for example the case of Emin, a CRM driving the epidermal expression of *miniature* [35]. Examination with cluster buster [52] or swan [53] did not detect multiple OvoQ6 motifs in Emin sequence, supporting that it contains a single putative Svb BS (even in *Drosophila melanogaster* only) and that BS clustering is not an absolute requisite for Svb regulation. Additional CRMs (4702B, 17058, 31559, tyn2, EminB and 32159) identified *in vivo* from previous transgenic assays (IF and SP, unpublished and Fig.S2A) were not either detected in the top 100 cisTargetX predictions likely because of weak clustering and/or evolutionary conservation of putative Svb BS. Reciprocally, 72% of OvoQ6-clustered regions we tested were devoid of activity during embryogenesis, suggesting that additional parameters are required to discriminate between CRM and inactive regions. To address this question, we compared the two sets of experimentally tested regions, *i.e.* 14 CRMs (positive) and 25 inactive regions (negative), using an algorithm designed for *de novo* motif discovery [54]. Briefly, we systematically searched, *ab initio*, for 10bp motifs that are evolutionarily conserved among *Drosophila* species and display a distribution within positive CRMs statistically different from background intergenic sequences. We then evaluated how well each motif discriminated between positive CRMs and negative regions (Fig. 3A) and ranked these *de novo* generated motifs accordingly. Unexpectedly, the most discriminative PWM clearly overlaps OvoQ6 (**CnGTTa**), with

a similar core consensus but extended to adjacent nucleotides (A**C**H**GTTAK**). We found that this motif, hereafter called svbF7, was sufficient when taken alone to detect 10 out of 14 active CRMs (Fig. 3B and S2B). In contrast, svbF7 was detectable in only 6/25 negative regions (Fig. 3B and S2B), even when lowering the threshold (data not shown). The proportion of CRMs containing the svbF7 motif reached 13/14, when relaxing the penalty imposed for weak evolutionary conservation (in this case dependent on aligned orthologous regions [54]). Hence, svbF7 performs better than ovoQ6 or any other related motifs [55], e.g. sorting out Emin, 32159, tyn2 and sha1 CRMs from negative regions (Fig. 3B and Fig. S2B). Similar results were observed when introducing SvbF7 in cisTargetX. In the set of 39 Svb downstream genes SvbF7 becomes the most significant PWM detected leading to better CRM predictions (less negative regions were predicted) (Fig.S2B). To evaluate whether this slight extension of the Svb PWM was relevant for CRM activity, we substituted two nucleotides flanking the core CnGTT in the unique Svb BS of Emin, *i.e.* altering the svbF7 motif without disrupting OvoQ6 consensus (Fig. 3C). When assayed *in vivo*, these two mutations strongly reduced Emin activity, when compared to wild type (Fig. 3C). These data therefore suggested that the svbF7 motif was a hallmark of Svb sites required *in vivo* for CRM activity.

Svb-dependent CRMs use different combinations of cis-regulatory elements.

Having shown the role of svbF7 in Emin activity, we further investigated its functional significance in CRMs containing unique (sha1, snE1, tyn2) or multiple (sha3, dyl2) occurrences of this motif. As observed for Emin, disruption of the single svbF7 site abolished sha1 activity (Fig. 4A) and decreased the activity of snE1 and tyn2, albeit only in ventral trichome cells (Fig. 4B,C). In the latter case, we found an

additional (non conserved) OvoQ6 putative BS that contributes to the activity of this element (Fig. 4C). For CRMs containing several svbF7 sites, simultaneous mutations also abrogated reporter expression (Fig. 4C-E). The individual disruption of each svbF7 led however to varying defects. The two svbF7 motifs of sha3 display partly redundant function, with a similar and limited impact on CRM activity when compared to the simultaneous KO (Fig. 4D). On the other hand, one of the three svbF7 of dyl2, plays a major role in CRM activity, whereas others contribute marginally to expression pattern or levels (Fig. 4E). Hence, disruption of svbF7 leads to a reduced activity for all five elements, confirming the importance of this motif for Svb-dependent CRMs. Nevertheless, the individual inactivation of svbF7 has different consequences on CRM activity, suggesting that additional elements are likely to modulate, locally, the *in vivo* function of each svbF7 motif.

We then searched for other putative cis-regulatory elements and evaluated their contribution to the activity of Svb-dependent CRMs. In a first approach, we performed a systematic mutagenesis of Emin by linker scanning (Fig. 5A). In addition to svbF7 whose inactivation abolished Emin activity, the alteration of three additional regions (8mt, 9mt and 10mt) impinged on epidermal expression (Fig. 5A). These results show that while Svb acts as a main switch for Emin activity, other motifs are required for complete expression. Consistently, *de novo* motif discovery identified a second PWM (WAGAAAGCSR), hereafter called the blue motif, enriched in positive regions and evolutionarily conserved in 7 out of 14 CRMs (Fig. 3 and 5B). Interestingly, mutations that disrupted the blue motif (9mt & 8mt) displayed the strongest effect on Emin activity, besides svbF7 KO (Fig. 5A). These unbiased data therefore show that the blue motif represents an element that, in addition to svbF7, is required for Emin activity. To further test the contribution of this element to the

activity of Svb-responsive CRMs, we mutated the blue motif in the two other enhancers that contain a single occurrence of it (Fig. 5B). As observed for Emin, disruption of the blue motif led to a limited reduction in snE1 activity, particularly in dorsal cells (Fig. 5C). Furthermore, the blue motif played a critical role in sha3, since its inactivation impaired reporter expression (Fig. 5C), similarly to the simultaneous inactivation of both svbF7 sites (Fig. 4).

In sum, genomic regions that behave *in vivo* as Svb-dependent CRMs are characterized by the presence of 1-3 Svb BS, showing a more constrained consensus (svbF7) than that detected *in vitro* (OvoQ6). A proportion of CRMs (9 out of 14) also comprises 1-5 blue motifs that contribute to their *in vivo* activity. Comparison of active CRMs did not detect significant constraints in the number, orientation or respective distribution of svbF7 and blue motifs, supporting the idea of a flexible organization of Svb-responsive CRMs.

Insights from genome-wide analyses into the functional organization of CRMs.

We then addressed whether and how svbF7, alone or in combination with the blue motif, can help predicting CRMs recognized by Svb within the whole genome. To extend in an unbiased way the repertoire of Svb downstream genes, we used microarray profiling of RNA samples from stage-15 whole embryos. However, we often detected only a modest reduction in the expression levels of previously identified downstream genes in *svb* mutant embryos (Fig. 6A and S3). We reasoned that in the absence of Pri peptides, Svb behaves as a sequence specific transcriptional repressor [32] and that dominant negative activity might further decrease residual expression of Svb TGs. Indeed, we measured such a reduction for known downstream genes, *e.g.*, *sha*, *m*, *CG32159*, *nyo*, *CG15889*, *dyl*, *CG4702*,

CG31559, *sn*, etc... (Fig. 6A and S3). Therefore we selected 150 genes down regulated in *svb* mutant embryos and showing a stronger reduction (2X) in *pri* mutants, as an unbiased set of candidate Svb-regulated genes. This set of genes encompasses 16 (out of 39) of the Svb downstream genes previously identified (Fig. S1), as well as 42 additional candidates showing expression in subsets of epidermal cells (Fig. S3). We examined the expression of 23 epidermal genes and *in situ* hybridization indicated that 21 of them require Svb activity to be expressed in trichome cells (Fig. 6A,B and S4). Consistent with the hypothesis of a direct regulation by Svb, this set of 150 genes displayed a 3 fold enrichment for putative Svb BS when compared to random distribution, or to a control set of 100 genes showing non significant variation of their expression in *svb* and *pri* mutants (Fig. 6C, S5B and Supplementary information). *svbF7* was one of the most highly enriched motif in the set of Svb-regulated genes, even after having masked the training set of sequences used for *de novo* motif discovery (Fig 6C and S5A). The blue motif was also retrieved in Svb regulated genes, but with a weak enrichment (1,5X) when compared to control (Fig. 6C). However, the combination of *svb7* and blue motifs was highly specific for Svb regulated genes, albeit reducing the sensitivity in *bona fide* TG prediction (Fig. 6C). These data support that, like observed for individual CRMs, a subset of Svb TGs are regulated independently of the blue motif. Within Svb TGs, the clustering is weak with an average value of 2 *svbF7* motif/gene, although a subset of Svb TGs exhibit more sites (Fig.6C). Whereas *svbF7* appears as a general feature of Svb targets, the blue motif may represent a wider spread cis-regulatory element providing efficient transcriptional output in a subset of Svb-responsive CRMs.

To test this hypothesis, we performed ChIP-seq making use of a Svb::GFP fusion to overcome the unavailability of suitable anti-Svb antibody. This transgenic

construct was expressed under the direct control of *svb* cis-regulatory regions, likely at levels comparable to endogenous since it allows robust phenotypical rescue of *svb* mutant embryos ([37] and data not shown). As deduced from ChIP-Seq, Svb bound to approximately 6000 sites throughout the *Drosophila* genome. This large number of binding events appears a feature shared by most *Drosophila* TFs that have been analyzed by ChIP studies [4,5,6,56,57]. Analysis of ChIP peaks with *i*-cistarget [58] clearly reveals that Ovo and Svb motifs are the most highly ranked detected motifs enriched in ChIP fragments. It has also been reported that multiple TFs display overlapping binding pattern, with common regions of TF co-localization, referred to as hotspots or HOT regions [5,6,59]. Consistently, 8% of Svb Chip peaks correspond to HOT regions as defined during early embryogenesis, suggesting that these regions remain accessible to TF binding in subsequent steps of development. Throughout the genome, we observed a strong cross correlation between svbF7 and ChIP-seq peaks (Fig. 7A), *i.e.* a high frequency of svbF7 motifs within a 1kb window from the peak center. The cross correlation between svbF7 and Svb Chip peaks was higher when HOT regions were removed from the analysis, consistent with the view that TF binding to HOT regions may be indirect [60]. We did not detect clustering of svbF7 motifs within ChIP-Seq peaks, since a second svbF7 is not systematically detected close to the first one (MS unpublished data and S5A). The genomic location of blue motifs confirmed a wider distribution, with a weaker but significant correlation with Svb ChIP-seq peaks (Fig. 7A). Focusing on the set of ChIP-seq peaks in Svb-regulated genes, we also observed clear association of svbF7 (Fig. S5A), low svbF7 clustering (Fig. S5A), and to a lesser extent of blue motifs, with ChIP peaks (Fig. S5A), further supporting that these two cis-regulatory elements represent hallmarks of Svb-responsive CRMs.

An independent way to evaluate this conclusion was to use *ab initio* analysis of ChiPed regions. PeakMotif [61] identified a PWM characteristic of peaks within Svb regulated genes that extensively matched svbF7 (Fig. S5A). A second PWM (TGAAAAGc and its reverse complement gCTTTTCA) that displays limited similarity with the blue motif was also detected in <50% of peaks, again only in Svb-regulated genes but not among the control set of genes (Fig. S5B). Hence, we interpret these results to imply that svbF7, and to a lesser extent blue motif, allow predicting CRM mediating regulation by Svb in trichome cells. Consistent with this conclusion, ChIP-seq regions that contain svbF7 motifs taken from *morpheyus* [36] or CG12017 (Fig. S4) behaved as Svb-dependent CRM when assayed *in vivo* (Fig. 7B). These experiments further illustrate that only a subset of active CRM relies on the blue motif, since the CRM12063-2 (*morpheyus*) contains a blue motif and CRM12017 does not.

Genome-wide analyses therefore indicate that the regulatory code learned from modeling and experimental dissection of a subset CRMs is likely relevant for understanding how the Svb TF selects its genomic set of TGs in trichome cells. Furthermore they fully support the functional importance of slight differences observed between Svb BS defined *in vivo* or *in vitro* and that Svb-dependent CRMs harbors a flexible use and arrangement of cis-regulatory elements.

Discussion

These data focusing on the gene regulatory network of epidermal cell morphogenesis provide novel insights into the nature and logic underlying cis-regulatory modules that mediate a program of terminal differentiation.

Consistent with the general view that TF recognize their targets through a clustering of BS [12];[18,42,62,63], we found statistical enrichment in motifs matching the Svb recognition sequence in its downstream genes (Fig. 6C). However, only a proportion of these putative sites mediate *in vivo* regulation. In addition, within the set of CRMs we experimentally manipulated, there is limited if any clustering of Svb BS, with 9 of out 16 CRM displaying a single recognizable site that has been conserved throughout *Drosophila* species. Therefore the local accumulation of BS is not a feature shared by CRMs mediating trichome formation, in contrast to gene regulatory networks underlying early development [7,8,10,11,12]. This conclusion is also supported by the weak autocorrelation of Svb BS observed within ChIP-seq fragments. Furthermore, *in silico* analyses predict that the *Drosophila* genome comprises only around 20 regions that present 3 evolutionary conserved Svb BS in a 1kb window (including dyl2 and 15589) and around 150 hits for 2 conserved sites (SP/MS unpublished). Even if some sites have been missed by our computational approaches, the presence of multiple BS within a short genomic region is therefore unlikely to be the critical parameter defining an active Svb-dependent CRM.

Indeed, we found that the nucleotide motif bound by Svb *in vivo* is more constrained than the consensus defined from *in vitro* [40] or bacterial one-hybrid approaches [55]. The presence of two A residues flanking the core CnGTT motif influence which putative BS are active *in vivo*, as inferred from CRM modeling and supported by *in vivo* results of site-directed mutagenesis. It is believed to be a

general feature of *in vivo* TF/DNA interaction as it was recently noticed for other TFs [18,64,65]. This shows that slight sequence differences, not detected by *in vitro* assays, can play a role in transcriptional regulation within genomic context, for instance, revealing the influence of co-factors as recently shown for Hox proteins [66].

Moreover, our findings highlight a paradoxical discrepancy between the enrichment of putative BS accumulated in non-coding regions of Svb downstream genes and the limited number of those acting as cis-regulatory elements. Is there a role for this evolutionarily conserved overrepresentation of Svb-like motifs in Svb targets? While isolated CRM drive transcription in trichome cells in the absence of multiple Svb-like motifs, it remains possible that these CRM do not perfectly reproduce the temporal or quantitative pattern of endogenous expression. We propose that the characteristic distributed accumulation of BS in Svb downstream genes may help reaching a precise regulation of their expression. For example, these sites presumably of weaker affinity (at least *in vivo*) can increase the local concentration in Svb facilitating their regulation through a few BS stably bound by Svb *in vivo*. Whether accumulated motifs close to an active BS may influence CRM activity remains currently hard to assay experimentally; it would be interesting to search for similar signatures in other genetic networks, in different contexts and species. It is also possible that these putative Svb BS may be used to drive expression in other developmental stages.

In addition, we found that the vicinity to additional cis-regulatory elements also contribute to define which Svb motifs act as regulatory elements. Statistical analyses of *in vivo* defined CRMs help predicting active cis-regulatory elements. In more than a half of Svb CRM we found 1-5 occurrences of a wider spread motif, we called the

blue motif, and show that its disruption generally impinges on CRM activity. Nevertheless, this orphan motif does not allow predicting which transcription or chromatin-associated factor(s) may mediate its activity, illustrating the critical need for large-scale BS discovery, ideally, from *in vivo* approaches. Furthermore, other epidermal CRMs do not rely on the blue motif, indicating that there are several ways to build a Svb-responsive element. Indeed, functional dissection of Emin by linker-scanning disclosed an additional region (mt10) contributing to CRM activity. This region contains a motif, TTATGCAA, previously proposed as a putative cis-regulatory element from its strong conservation throughout the evolution of *Drosophila* species [44]. In addition to Emin, this conserved motif (hereafter referred to as yellow motif) was detected in a half of Svb-dependent CRMs, with or without the blue motif (Fig. S2C). Again, whereas yellow motifs were detected within many genes, the combination of yellow and svbF7 motifs appeared highly specific for Svb downstream targets, expanding the diversity of signatures for Svb regulation (Fig.S2C). In summary, not only the number and respective arrangement of cis-regulatory elements can vary, but even the nature of those elements is different within the sample of Svb-dependent CRM we analyzed. Therefore, we find no evidence for a given “grammar”, or a unique combinational code, for the cis-regulatory elements of epidermis terminal differentiation, a feature that may actually be shared by other gene regulatory networks [16,19].

Our findings further reveal an unexpected level of modularity for the regulatory genome. For the *shavenoid* and *miniature* genes, we defined in each case two separable CRM (sha1, sha3 & Emin, EminB) that mediate Svb regulation. Recent works have established that apparently redundant, or shadow, CRMs ensure robust expression of key developmental regulators [31,67]. For example, the transcription of

svb itself involves separate CRMs that buffer the trichome pattern against genetic alterations or varying external conditions [31]. It has been proposed that shadow enhancers may be required to drive an acute expression of some developmentally important genes [68]. However, our data show that apparently redundant enhancers are not limited to regulatory factors operating at high positions in hierarchical gene networks. Instead, we provide evidence that several effector (“blue collar”) genes display a similar regulatory architecture, suggesting that multiple CRMs represent an overlooked feature of the successive tiers of gene networks.

It is interesting to note that a direct regulatory interaction occurring between a TF and a given downstream gene can rely on two separate elements [67]. It is possible that two apparently redundant CRMs within a gene arise from a local duplication [69] of a pre-existing regulatory module, especially if their activity relies on the same TF. Nevertheless, the two CRMs from *shavenoid*, and from *miniature*, display distinct organizations (with and without blue or yellow motifs), suggesting either convergent evolution of two ancient elements, or diversification of duplicated elements.

Collectively, these data show a surprising multiscale diversity of the cis-regulatory elements mediating a terminal differentiation program governed by a given TF. They also point towards the advantage of combining genome-wide experiments computational analyses and intimate functional dissection to help deciphering how developmental programs are encoded in the regulatory genome.

Materials and Methods

Fly strains

We used *btd*, *svb*¹ or *svb*^{R9} [35] and *pri*¹ [70] stocks over GFP balancers. Transgenic lines were generated (Bestgene www.thebestgene.com, or Fly Facility www.fly-facility.com) using P-element mediated transformation (at least three independent insertions were analysed), or phiC31 using the 86F landing platform, where the empty vector does not drive detectable expression during the embryonic stages examined.

Molecular procedures

Genomic regions were PCR amplified and cloned into pCasper-lacZ, pCasper-hs43-lacZ, pCbeta-lacZ or pATTb-lacZ vectors (cloning details available upon requests). QuikChange XL site-directed mutagenesis (Stratagene) was used to introduce point mutations in CRMs. Emin linker scanning was performed by replacing blocks of CRM sequence by CCGCCGGCGG. All constructs were verified by sequencing.

Antibody staining and *in situ* hybridization

10-14h embryos were collected, fixed and stained as described previously [35] using anti- β -galactosidase (Cappel, 1/1000) and streptavidin amplification kit (Vectastain) or Alexafluor488 (Molecular Probes). Dig- or biotin-labelled antisense RNA probes were used for *in situ* hybridization experiments, using standard protocols. Embryos were imaged with a Nikon Eclipse 90i microscope.

Microarrays

13-15h mutant embryos (*svb*^{R9} or *pri*¹) were selected under an epifluorescence Nikon SMZ1500 stereomicroscope using GFP balancers. 200 embryos were subjected to trizol (Invitrogen) mRNA extraction and RNA quality was monitored using Agilent Chip and RT

PCR. 5 independent RNA samples of each genotype were used for microarray analyses (Affymetrix; IGBMC, Strasbourg). Data extraction and normalization were performed using Affymetrix softwares and statistical analysis with the R program. A 2 fold differences in expression levels between both genotypes (see result) was found to be the most stringent criteria to retrieve Svb downstream genes (with a false discovery rate (FDR) of 0.01 for *pri*). The top 150 genes down regulated both in *pri* and *svb* mutants defined the set of “Svb-regulated” genes. The top 100 genes showing no significant (highest Pvalue) variation of their expression in the two mutant backgrounds (FDR>0,99) were used as a control set.

ChIP-seq experiments

8-10h and 12-14h staged embryos were collected and ChIP-seq data was generated as described [6]. Sequence data was analyzed using a virtual machine image on the Bionimbus cloud (<http://www.bionimbus.org/about-2/>) and aligned to the *Drosophila melanogaster* genome using Bowtie [71]. Sequence density along the genome was visualized using wig files generated with SPP [72] and sequence enrichment along the genome was defined by MACS with the following parameters: tag size=36, bandwidth=100, pvalue=1e-5 [73].

Motif detection and genome analysis

Detection of motifs enriched in Svb targets and Svb-independent epidermal genes was performed using *cisTargetX* (<http://med.kuleuven.be/cme-mg/lng/cisTargetX/>) [42]. For *de novo* motif discovery, genomic sequences of positive regions (active CRM) and negative regions (showing no transcriptional activity) were processed through a custom C++ program and statistical operations performed within the R software environment (<http://www.r-project.org>) as described [54]. We computed the cross-correlation between conserved motif instances and Svb ChIP-seq data as follows. In a 10 Kb centered around each ChIP peak, we 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 set of genes, each ChIP peak was

associated to the nearest transcription start site. Additional details of computational analyses are given in supplemental information.

Statistical analyses

To test for putative enrichment in a given motif between Svb-regulated and control set of genes (fig 6C), we used a Mann-Whitney U test using the transcribed region of each gene extended to 5 kb flanking sequences. A p-value was computed using the function `wilcox.test` from the R stats package. For Motif vs ChIPseq cross-correlations (fig 7 A), we performed a χ^2 -test to disentangle cross-correlation signals from small number fluctuations. Correlation data were binned in 500bp elements in a +/-10 kb region around the center of ChIPseq, resulting in k=40 bins. A χ^2 was computed as the sum over the bins of the standardized counts $\sum_i (O_i - E)^2 / E$, where O_i represents the observed count in bin i and E is the expected number of counts in a bin, taken to be uniform over the considered region. Finally, a p-value was computed as the probability that a χ^2 statistics with k-1 degrees of freedom takes at least the observed value.

Acknowledgements

We are grateful to the Bloomington Drosophila Stock Center, Flybase and Drosophila Genomic Resource Center for providing us with flies and molecular clones. We are indebted to B. Ronsin (Toulouse RIO Imaging) and to P. Valenti for excellent technical assistance. We also thank C. Hermann, N. Negre, E. Preger-Ben Noon and members of the SP/FP lab for critical reading of the manuscript. This work was supported by grants from ANR “Netoshape” and “Association pour la Recherche contre le Cancer” (n° 3832, n° 1111, and a fellowship to I.F.); DM was supported by a fellowship from University Paul Sabatier (cisdecode). All authors read and commented on the manuscript. The authors declare no conflict of interest. SA, FP and SP conceived and designed the experiments. MS, HR and VH contributed to the design of the bioinformatics analysis. DM, MS, IF, RS, JZ, IG, YL, PF performed the experiments and are listed according to their respective contributions. SP and FP wrote the paper.

FIGURE LEGENDS

Figure 1: Enrichment in Ovo/Svb binding sites defines an evolutionary conserved signature of *svb* downstream genes. **A:** Expression of *svb* (left top) defines the subset epidermal cells that form trichomes, visible on the larval cuticle (left bottom). Expression of *shavenoid/kojak* (*sha*) and *CG15589*, two genes transcribed in trichome cells under the control of Svb, as seen in wild type (top panels) versus *svb* mutant embryos (bottom panels). **B:** ROC curve showing significant enrichment in Svb binding sites, represented by the OvoQ6 position weight matrix, among 16/39 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 genomic regions containing clusters of putative Svb binding sites correspond to functional CRM. Vertical black lines represent evolutionary conserved OvoQ6 clusters (at least two conserved motifs in a 1kb window) in *singed* (**A**) and *shavenoid* (**B**) genes. Pink and cyan horizontal boxes summarize regions tested *in vivo* by transgenic assays using lacZ reporters. While pink constructs do not drive specific expression, blue constructs reproduced endogenous expression in trichome cells, under the control of Svb, as demonstrated by a reduced staining in *svb* mutants. **C.** Putative CRMs predicted by cis-TargetX on the basis of evolutionary conservation and clustering of putative Svb BSs. Pictures show expression of positive CRMs in wild type (top) and *svb* mutant (bottom) embryos.

Figure 3: Computational analyses of Svb-dependent CRM.

A. Statistical analysis of positive CRMs versus negative (intergenic genomic sequences used as background) was used for *de novo* discovery of nucleotide motifs that characterize active CRMs (left). The two matrixes svbF7 and blue motif perform best in discriminating between positive (CRMs) and negative fragments devoid of any CRM activity (Supplementary Informations) as illustrated by the Pareto plot (**B**). **C.** Functional evidence that positions flanking the CnGTT OvoQ6 core motif play a critical role in CRM activity. While disruption of the core motif abolished Emin activity, two point mutations that affect the 5' and 3' flanking nucleotides strongly reduce epidermal expression.

Figure 4: Functional influence of svbF7 motifs in the activity of Svb-dependent CRMs.

Pictures show modifications of reporter gene expression resulting from individual (A-E) and simultaneous (C-E) inactivation of svbF7 motifs in sha1 (A), snE1 (B), tyn2 (C), sha3 (D) and dyl2 (E) CRMs. Red boxes schematize svbF7 motifs; open black boxes represent OvoQ6 sites that have not been conserved throughout evolution of *Drosophila* species.

Figure 5: Svb-dependent CRMs use distinct combinatorial codes of cis-regulatory element. **A.** In addition to svbF7, linker-scanning mutagenesis of the Emin CRM identifies other motifs (8mt, 9mt and 10mt) required for full transcriptional activity, as deduced from altered pattern of lac-Z reporter expression (green). **B.** Within each epidermal CRMs, red and blue boxes schematize the distribution, number and orientation of svbF7 and blue motifs, respectively. Filled boxes represent motifs conserved throughout *Drosophila* species; open boxes motifs detected only in *D. melanogaster*; black boxes represent OvoQ6 motifs. **C.** Consequences of point mutations that disrupt the blue motifs in Emin, snE1 and sha3 CRMs.

Figure 6: microarray profiling of embryonic genes regulated by Svb.

A. Modifications in expression levels between wild-type and *svb* (left) or *pri* (right) mutant embryos observed in regulated (green) and control (light blue) set of genes. Dark green boxes represent previously identified Svb TGs (see Fig. S1 and sup. info); light green boxes novel downstream genes as validated by *in situ* hybridization (Fig. S4), and open boxes additional candidates. **B.** Whole mount *in situ* hybridization of CG1273, a Svb downstream target identified from microarray profiling, showing down regulation in *svb* mutant embryos and a further reduction in epidermal cells of *pri* mutants. C. Distribution of conserved svbF7 (red) and blue (blue) motifs within Svb-regulated (Svb reg) *versus* control set of genes. To avoid over-fitting, the positives sequences (CRMs) used in fig3 for the *de novo* motif discovery were masked prior analyses. The combination of svbF7 and blue motif exhibits higher selectivity (<5% FPR), albeit reducing sensitivity of TGs detection. The right panel plots number of conserved svbF7 and blue motifs detected in each set of genes. *** indicate a P value <0,001, **<0,01 and * <0,05.

Figure 7: genome-wide identification of Svb direct targets and their respective CRM using computational and *in vivo* experimental approaches.

A. Analysis of cross-correlation between conserved svbF7 or blue motif instances and Svb ChIP-Seq peaks, throughout the whole genome, among Svb-regulated (Fig. S5A top) and control set (Fig S5B bottom) of genes. Plots show numbers of svbF7 (red) or blue motif (blue) found in a 10kb window on each side of the center of ChIP-seq peaks. *** indicate a P value (Chi2 test) <0,001, **<0,01. B. Analyses of motif distribution coupled to ChIP-seq allow predicting location of active CRMs in Svb downstream targets. Graphs show intensity profiles of Svb ChIP at two time points of *Drosophila* embryogenesis (8-10h and 12-14h); red and blue vertical bars, the distribution of conserved svbF7 (red) and blue motif (blue).

Active CRM are drawn as cyan rectangles. Pictures show reporter gene expression driven by corresponding regions in wild type and *svb* mutant embryos.

REFERENCES

1. Stathopoulos A, Levine M (2005) Genomic regulatory networks and animal development. *Dev Cell* 9: 449-462.
2. Ptashne M (2005) Regulation of transcription: from lambda to eukaryotes. *Trends Biochem Sci* 30: 275-279.
3. Biggin MD (2010) MyoD, a lesson in widespread DNA binding. *Dev Cell* 18: 505-506.
4. Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, et al. (2008) Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol* 6: e27.
5. MacArthur S, Li XY, Li J, Brown JB, Chu HC, et al. (2009) 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* 10: R80.
6. Negre N, Brown CD, Ma L, Bristow CA, Miller SW, et al. (2011) A cis-regulatory map of the *Drosophila* genome. *Nature* 471: 527-531.
7. Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U, Gaul U (2008) Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature* 451: 535-540.
8. Jakobsen JS, Braun M, Astorga J, Gustafson EH, Sandmann T, et al. (2007) Temporal ChIP-on-chip reveals Biniou as a universal regulator of the visceral muscle transcriptional network. *Genes Dev* 21: 2448-2460.
9. Markstein M, Levine M (2002) Decoding cis-regulatory DNAs in the *Drosophila* genome. *Curr Opin Genet Dev* 12: 601-606.
10. Sandmann T, Girardot C, Brehme M, Tongprasit W, Stolc V, et al. (2007) A core transcriptional network for early mesoderm development in *Drosophila melanogaster*. *Genes Dev* 21: 436-449.
11. Zeitlinger J, Zinzen RP, Stark A, Kellis M, Zhang H, et al. (2007) Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the *Drosophila* embryo. *Genes Dev* 21: 385-390.
12. Markstein M, Markstein P, Markstein V, Levine MS (2002) Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the *Drosophila* embryo. *Proc Natl Acad Sci U S A* 99: 763-768.
13. Sinha S, Schroeder MD, Unnerstall U, Gaul U, Siggia ED (2004) Cross-species comparison significantly improves genome-wide prediction of cis-regulatory modules in *Drosophila*. *BMC Bioinformatics* 5: 129.
14. He Q, Bardet AF, Patton B, Purvis J, Johnston J, et al. (2011) High conservation of transcription factor binding and evidence for combinatorial regulation across six *Drosophila* species. *Nat Genet* 43: 414-420.
15. Markstein M, Zinzen R, Markstein P, Yee KP, Erives A, et al. (2004) A regulatory code for neurogenic gene expression in the *Drosophila* embryo. *Development* 131: 2387-2394.
16. Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EE (2009) Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* 462: 65-70.

17. Zinzen RP, Senger K, Levine M, Papatsenko D (2006) Computational models for neurogenic gene expression in the *Drosophila* embryo. *Curr Biol* 16: 1358-1365.
18. Khoueiry P, Rothbacher U, Ohtsuka Y, Daian F, Frangulian E, et al. (2010) A cis-regulatory signature in ascidians and flies, independent of transcription factor binding sites. *Curr Biol* 20: 792-802.
19. Junion G, Spivakov M, Girardot C, Braun M, Gustafson EH, et al. (2012) A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* 148: 473-486.
20. Hare EE, Peterson BK, Iyer VN, Meier R, Eisen MB (2008) Sepsid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. *PLoS Genet* 4: e1000106.
21. Arnosti DN, Kulkarni MM (2005) Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J Cell Biochem* 94: 890-898.
22. Rowan S, Siggers T, Lachke SA, Yue Y, Bulyk ML, et al. (2010) Precise temporal control of the eye regulatory gene Pax6 via enhancer-binding site affinity. *Genes Dev* 24: 980-985.
23. Swanson CI, Schwimmer DB, Barolo S (2011) Rapid evolutionary rewiring of a structurally constrained eye enhancer. *Curr Biol* 21: 1186-1196.
24. Shultzaberger RK, Malashock DS, Kirsch JF, Eisen MB (2010) The fitness landscapes of cis-acting binding sites in different promoter and environmental contexts. *PLoS Genet* 6: e1001042.
25. Harrison MM, Li XY, Kaplan T, Botchan MR, Eisen MB (2012) Zelda binding in the early *Drosophila melanogaster* embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet* 7: e1002266.
26. Rister J, Desplan C (2010) Deciphering the genome's regulatory code: the many languages of DNA. *Bioessays* 32: 381-384.
27. Payre F, Vincent A, Carreno S (1999) ovo/svb integrates Wingless and DER pathways to control epidermis differentiation. *Nature* 400: 271-275.
28. Delon I, Chanut-Delalande H, Payre F (2003) The Ovo/Shavenbaby transcription factor specifies actin remodelling during epidermal differentiation in *Drosophila*. *Mech Dev* 120: 747-758.
29. Payre F (2004) Genetic control of epidermis differentiation in *Drosophila*. *Int J Dev Biol* 48: 207-215.
30. McGregor AP, Orgogozo V, Delon I, Zanet J, Srinivasan DG, et al. (2007) Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature* 448: 587-590.
31. Frankel N, Davis GK, Vargas D, Wang S, Payre F, et al. (2010) Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* 466: 490-493.
32. Kondo T, Plaza S, Zanet J, Benrabah E, Valenti P, et al. (2010) Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Science* 329: 336-339.
33. Andrew DJ, Baker BS (2008) Expression of the *Drosophila* secreted cuticle protein 73 (dsc73) requires Shavenbaby. *Dev Dyn* 237: 1198-1206.
34. Bejsovec A, Chao AT (2012) crinkled reveals a new role for Wingless signaling in *Drosophila* denticle formation. *Development* 139: 690-698.
35. Chanut-Delalande H, Fernandes I, Roch F, Payre F, Plaza S (2006) Shavenbaby couples patterning to epidermal cell shape control. *PLoS Biol* 4: e290.

36. Fernandes I, Chanut-Delalande H, Ferrer P, Latapie Y, Waltzer L, et al. (2010) Zona pellucida domain proteins remodel the apical compartment for localized cell shape changes. *Dev Cell* 18: 64-76.
37. Frankel N, Ereyilmaz DF, McGregor AP, Wang S, Payre F, et al. (2011) Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature* 474: 598-603.
38. Sucena E, Delon I, Jones I, Payre F, Stern DL (2003) Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. *Nature* 424: 935-938.
39. Khila A, El Haidani A, Vincent A, Payre F, Souda SI (2003) The dual function of ovo/shavenbaby in germline and epidermis differentiation is conserved between *Drosophila melanogaster* and the olive fruit fly *Bactrocera oleae*. *Insect Biochem Mol Biol* 33: 691-699.
40. Lee S, Garfinkel MD (2000) Characterization of *Drosophila* OVO protein DNA binding specificity using random DNA oligomer selection suggests zinc finger degeneration. *Nucleic Acids Res* 28: 826-834.
41. Lu J, Oliver B (2001) *Drosophila* OVO regulates ovarian tumor transcription by binding unusually near the transcription start site. *Development* 128: 1671-1686.
42. Aerts S, Quan XJ, Claeys A, Naval Sanchez M, Tate P, et al. (2010) Robust target gene discovery through transcriptome perturbations and genome-wide enhancer predictions in *Drosophila* uncovers a regulatory basis for sensory specification. *PLoS Biol* 8: e1000435.
43. Aerts S, van Helden J, Sand O, Hassan BA (2007) Fine-tuning enhancer models to predict transcriptional targets across multiple genomes. *PLoS One* 2: e1115.
44. Elemento O, Tavazoie S (2005) Fast and systematic genome-wide discovery of conserved regulatory elements using a non-alignment based approach. *Genome Biol* 6: R18.
45. Stark A, Lin MF, Kheradpour P, Pedersen JS, Parts L, et al. (2007) Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* 450: 219-232.
46. Mace KA, Pearson JC, McGinnis W (2005) An epidermal barrier wound repair pathway in *Drosophila* is mediated by grainy head. *Science* 308: 381-385.
47. Ting SB, Caddy J, Hislop N, Wilanowski T, Auden A, et al. (2005) A homolog of *Drosophila* grainy head is essential for epidermal integrity in mice. *Science* 308: 411-413.
48. Szuplewski S, Fraisse-Veron I, George H, Terracol R (2010) vril is required to ensure tracheal integrity in *Drosophila* embryo. *Dev Growth Differ* 52: 409-418.
49. Szuplewski S, Kottler B, Terracol R (2003) The *Drosophila* bZIP transcription factor Vrille is involved in hair and cell growth. *Development* 130: 3651-3662.
50. Jayo A, Parsons M (2012) Fascin: a key regulator of cytoskeletal dynamics. *Int J Biochem Cell Biol* 42: 1614-1617.
51. Ren N, He B, Stone D, Kirakodu S, Adler PN (2006) The shavenoid gene of *Drosophila* encodes a novel actin cytoskeleton interacting protein that promotes wing hair morphogenesis. *Genetics* 172: 1643-1653.
52. Frith MC, Li MC, Weng Z (2003) Cluster-Buster: Finding dense clusters of motifs in DNA sequences. *Nucleic Acids Res* 31: 3666-3668.
53. Kim J, Cunningham R, James B, Wyder S, Gibson JD, et al. (2010) Functional characterization of transcription factor motifs using cross-species comparison across large evolutionary distances. *PLoS Comput Biol* 6: e1000652.

54. Rouault H, Mazouni K, Couturier L, Hakim V, Schweisguth F (2010) Genome-wide identification of cis-regulatory motifs and modules underlying gene coregulation using statistics and phylogeny. *Proc Natl Acad Sci U S A* 107: 14615-14620.
55. Zhu LJ, Christensen RG, Kazemian M, Hull CJ, Enuameh MS, et al. (2011) FlyFactorSurvey: a database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res* 39: D111-117.
56. Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, et al. (2010) Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* 330: 1787-1797.
57. Wunderlich Z, Mirny LA (2009) Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet* 25: 434-440.
58. Hermann C, Van de Sande B, Potier D, Aerts S (2012) i-cisTarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules. *Nucleic Acids Res* in press.
59. Moorman C, Sun LV, Wang J, de Wit E, Talhout W, et al. (2006) Hotspots of transcription factor colocalization in the genome of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 103: 12027-12032.
60. Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, et al. (2010) Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* 330: 1775-1787.
61. Thomas-Chollier M, Herrmann C, Defrance M, Sand O, Thieffry D, et al. (2011) RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. *Nucleic Acids Res*.
62. Li L, Zhu Q, He X, Sinha S, Halfon MS (2007) Large-scale analysis of transcriptional cis-regulatory modules reveals both common features and distinct subclasses. *Genome Biol* 8: R101.
63. Berman BP, Nibu Y, Pfeiffer BD, Tomancak P, Celtniker SE, et al. (2002) 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* 99: 757-762.
64. Ozdemir A, Fisher-Aylor KI, Pepke S, Samanta M, Dunipace L, et al. (2011) High resolution mapping of Twist to DNA in *Drosophila* embryos: Efficient functional analysis and evolutionary conservation. *Genome Res* 21: 566-577.
65. Guertin MJ, Lis JT (2010) Chromatin landscape dictates HSF binding to target DNA elements. *PLoS Genet* 6.
66. Slattery M, Riley T, Liu P, Abe N, Gomez-Alcala P, et al. (2011) Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* 147: 1270-1282.
67. Perry MW, Boettiger AN, Bothma JP, Levine M (2010) Shadow enhancers foster robustness of *Drosophila* gastrulation. *Curr Biol* 20: 1562-1567.
68. Hong JW, Hendrix DA, Levine MS (2008) Shadow enhancers as a source of evolutionary novelty. *Science* 321: 1314.
69. Nourmohammad A, Lassig M (2011) Formation of regulatory modules by local sequence duplication. *PLoS Comput Biol* 7: e1002167.
70. Kondo T, Hashimoto Y, Kato K, Inagaki S, Hayashi S, et al. (2007) Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat Cell Biol* 9: 660-665.
71. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: R25.

72. Kharchenko PV, Tolstorukov MY, Park PJ (2008) Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* 26: 1351-1359.

73. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9: R137.

Figure1

[Click here to download high resolution image](#)

Menoret_Fig1

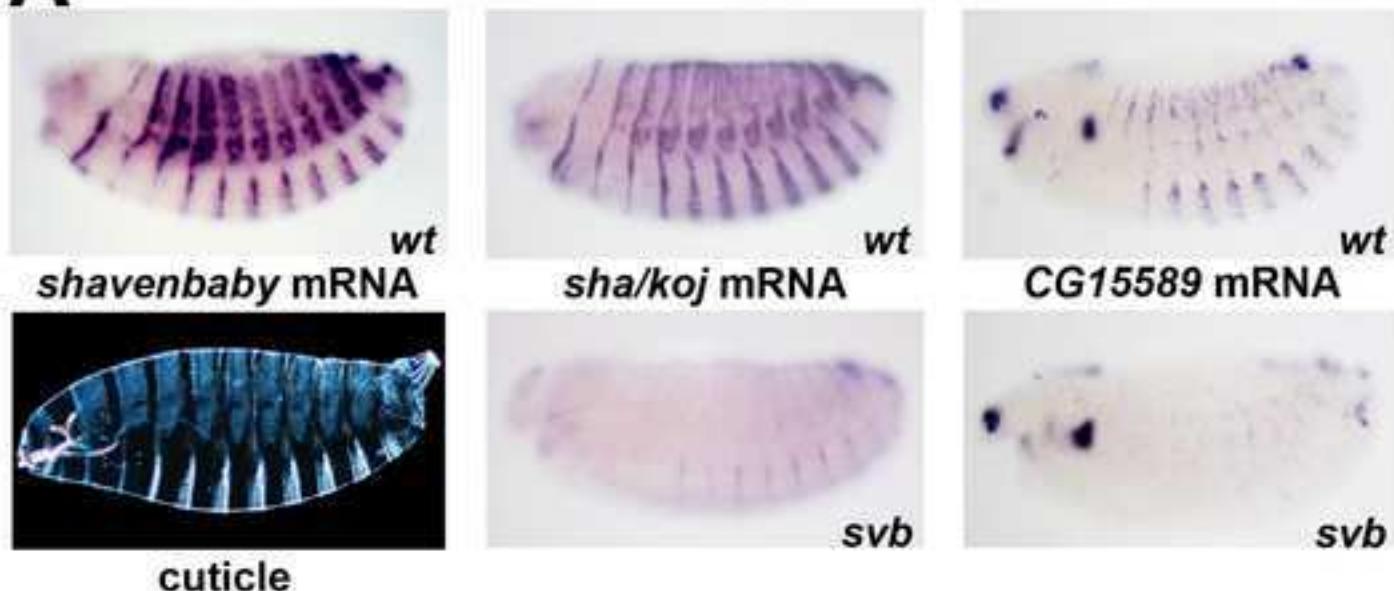
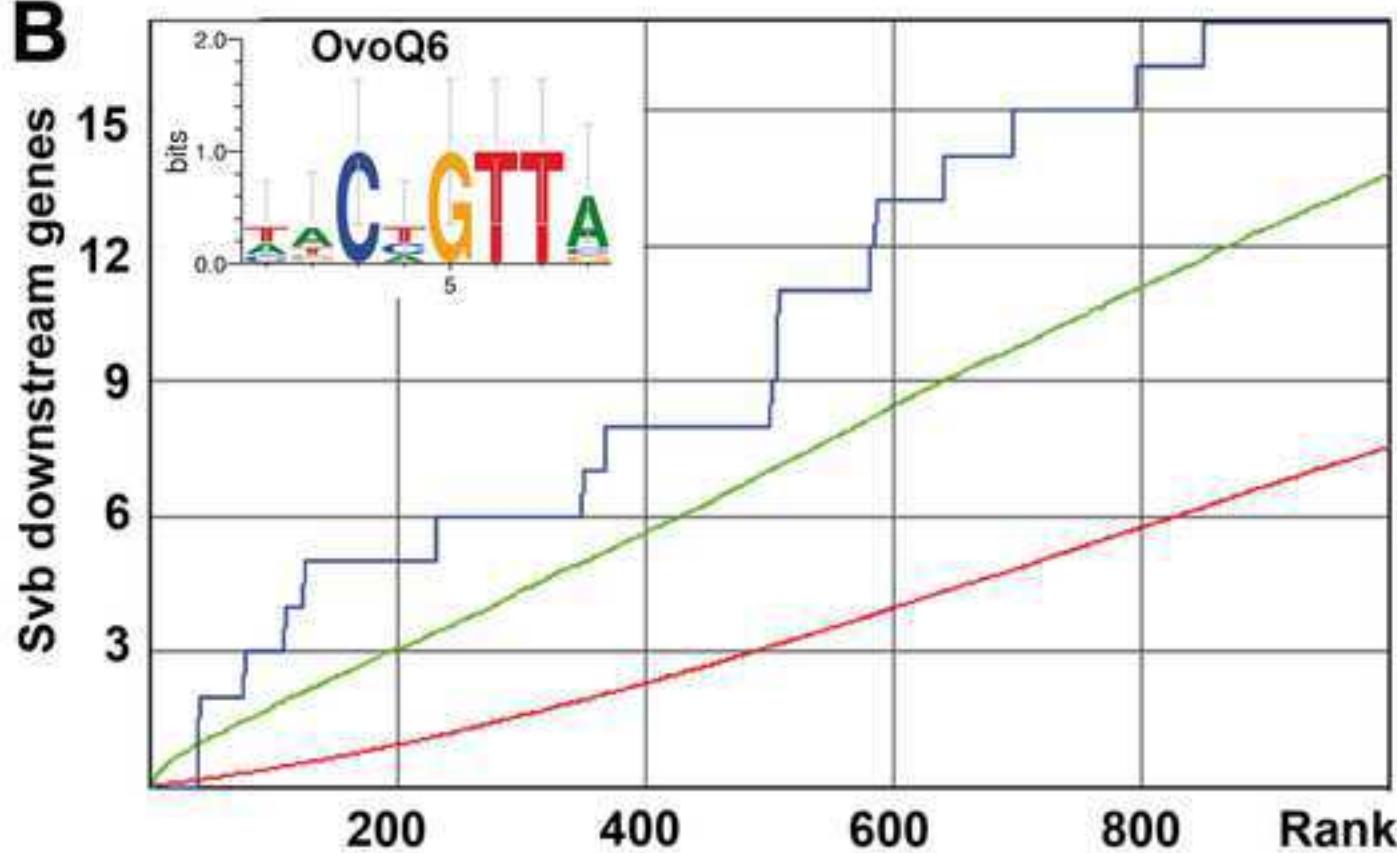
A**B**

Figure2

[Click here to download high resolution image](#)

Menoret_Fig2

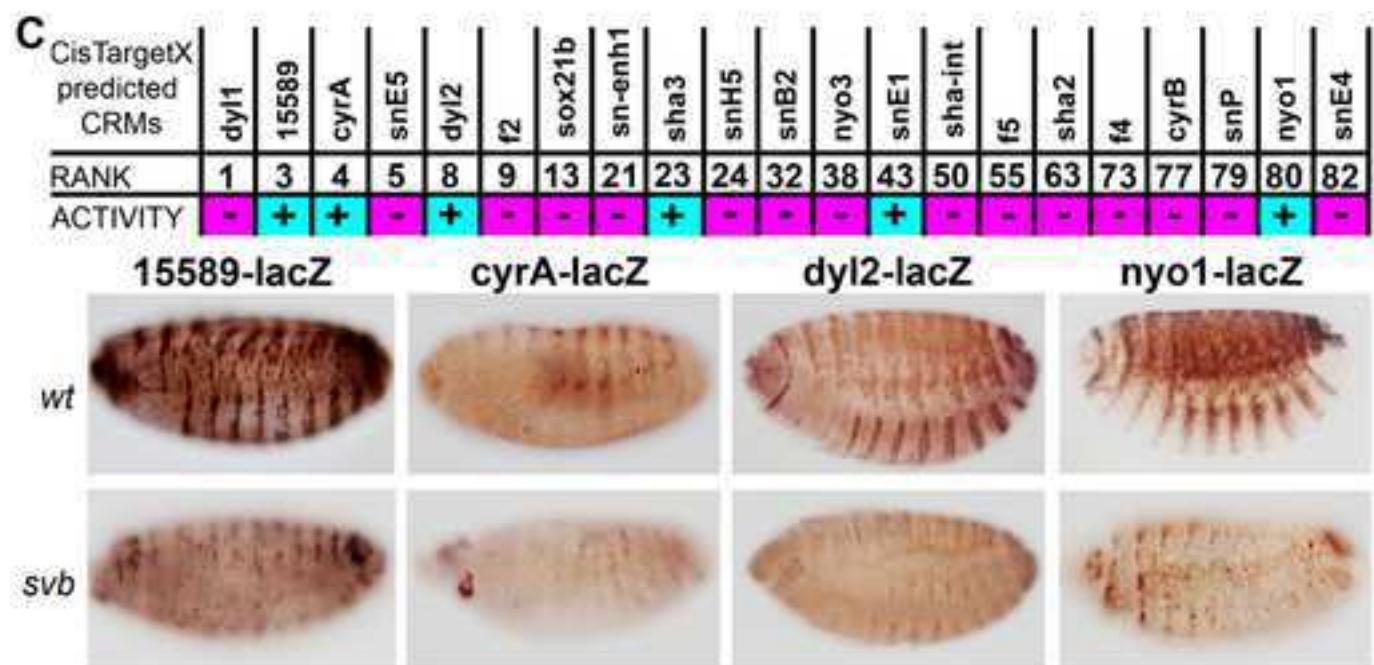
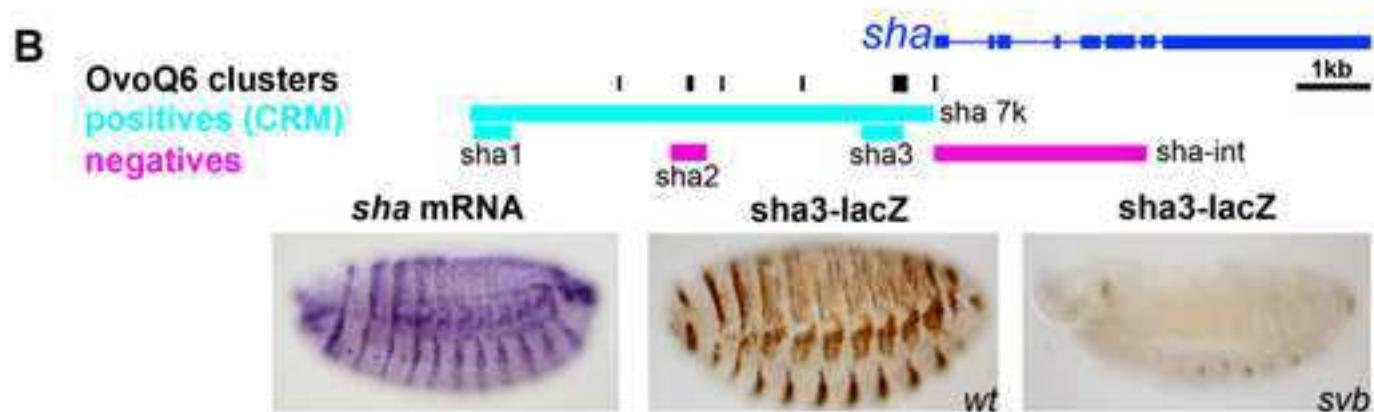
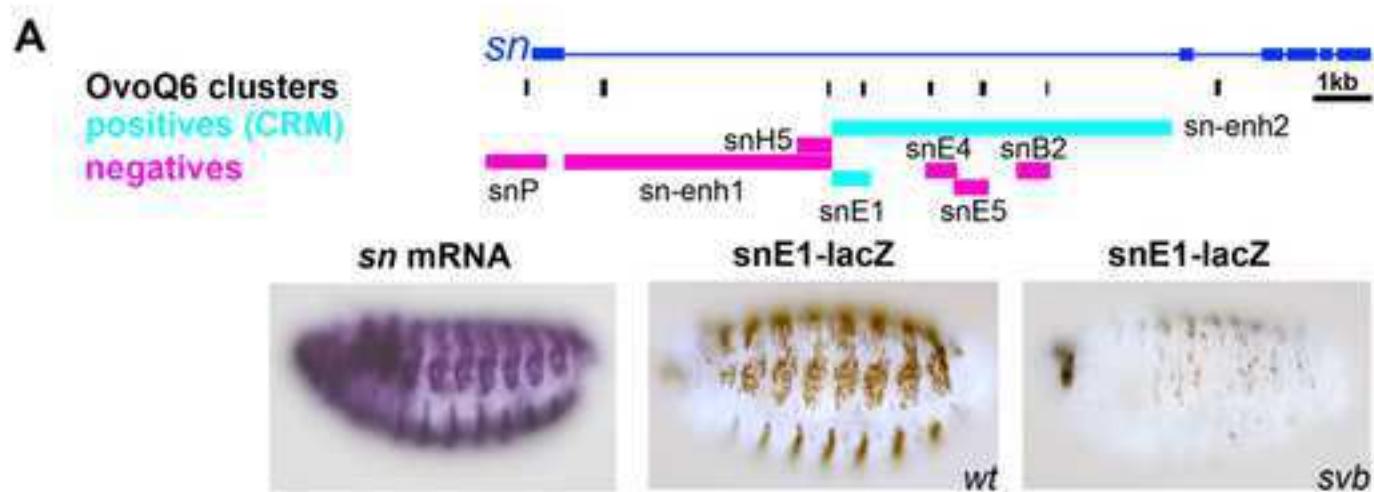


Figure3

[Click here to download high resolution image](#)

Menoret_Fig3

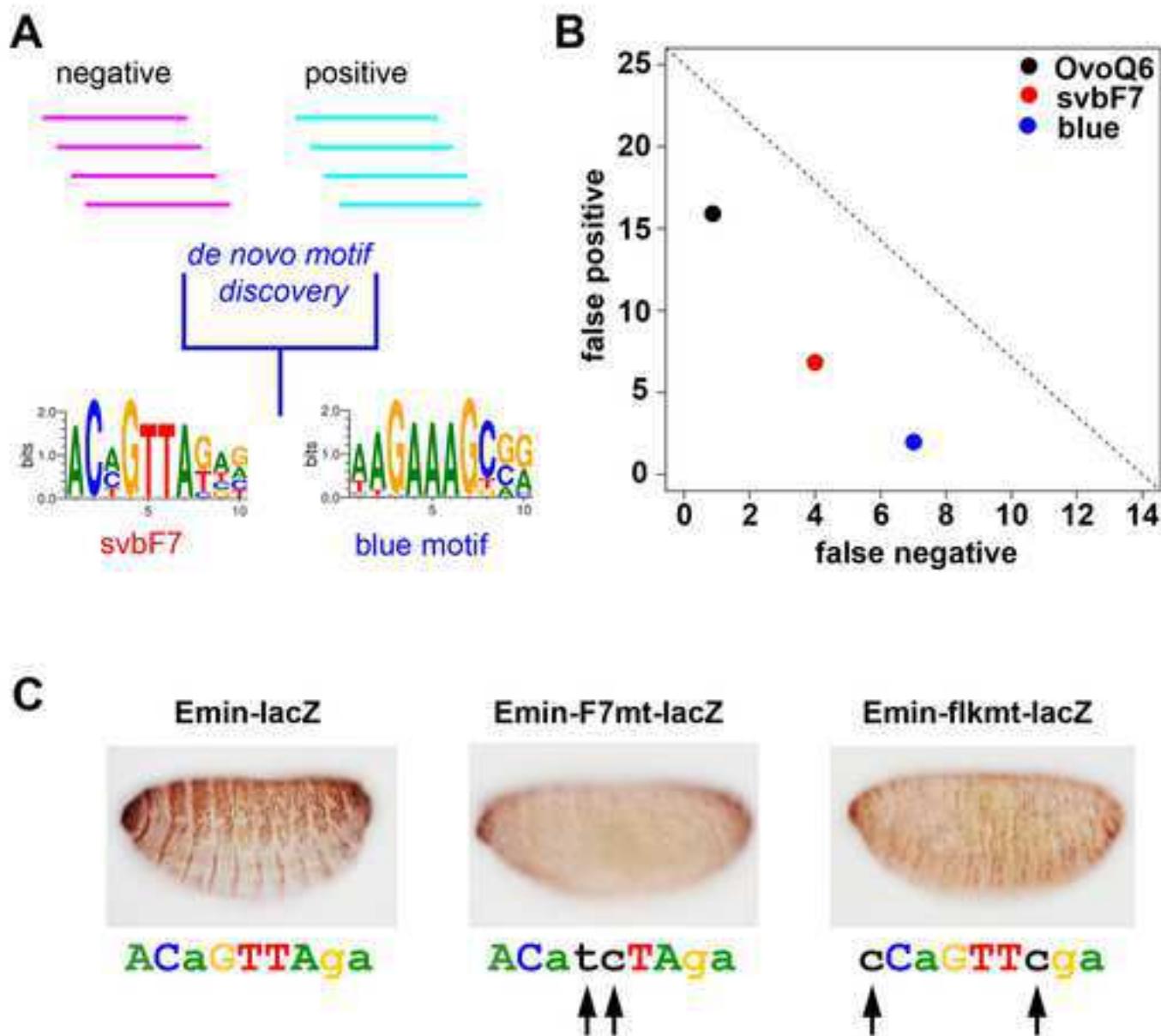
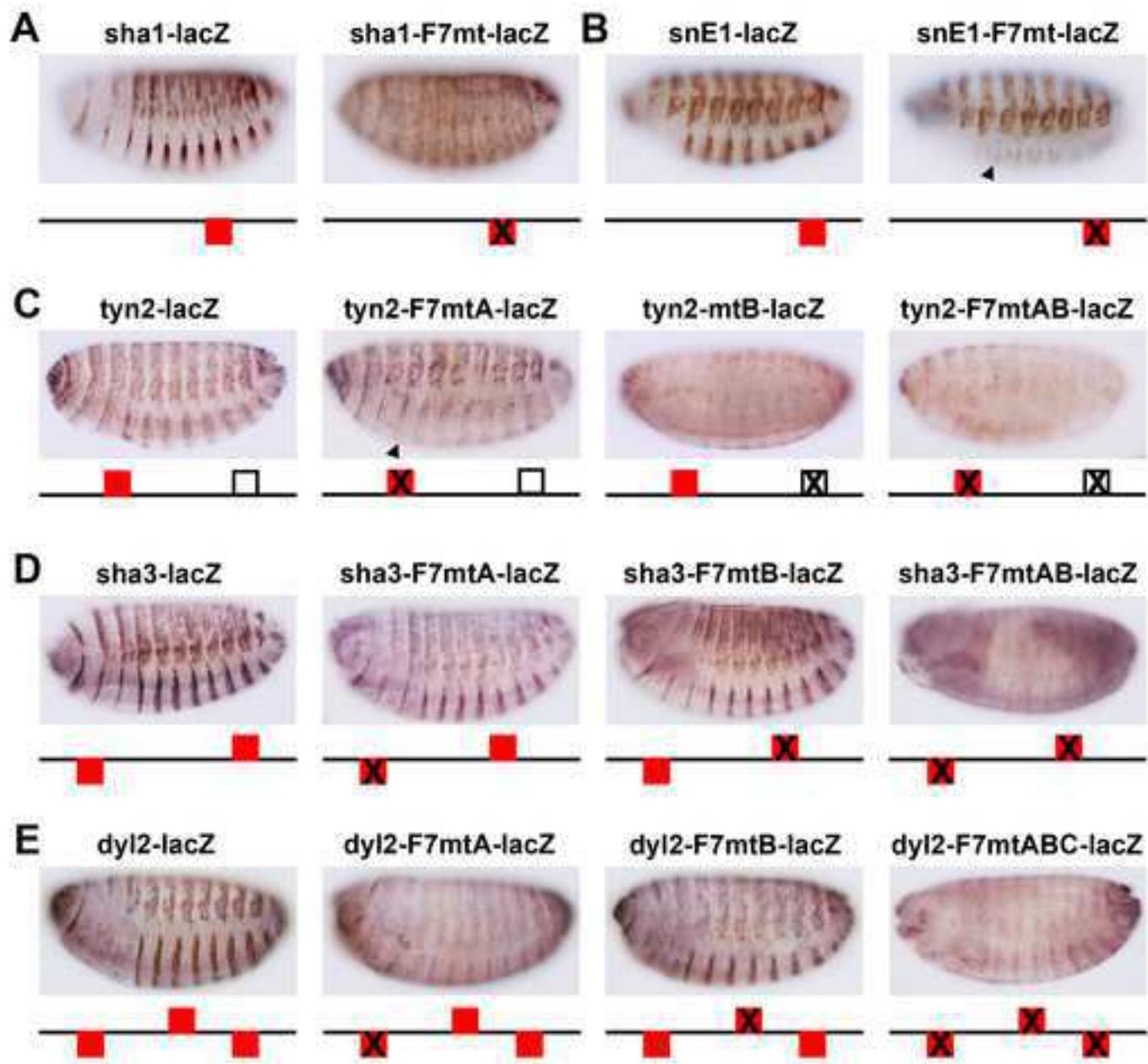


Figure4

[Click here to download high resolution image](#)

Menoret_Fig4



Fig

[Click here to download high resolution image](#)

Menoret_Fig5

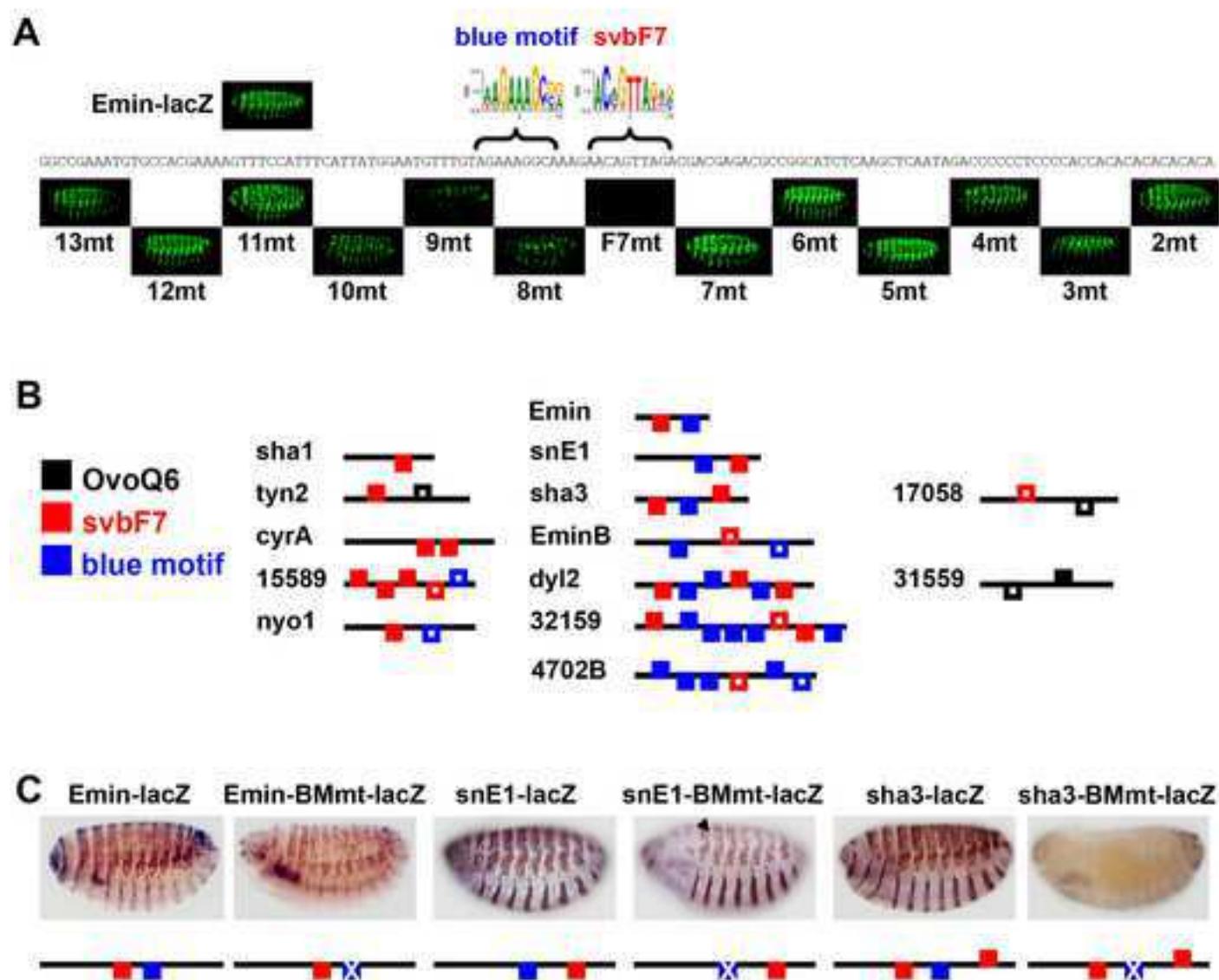


Figure6

[Click here to download high resolution image](#)

Menoret_Fig6

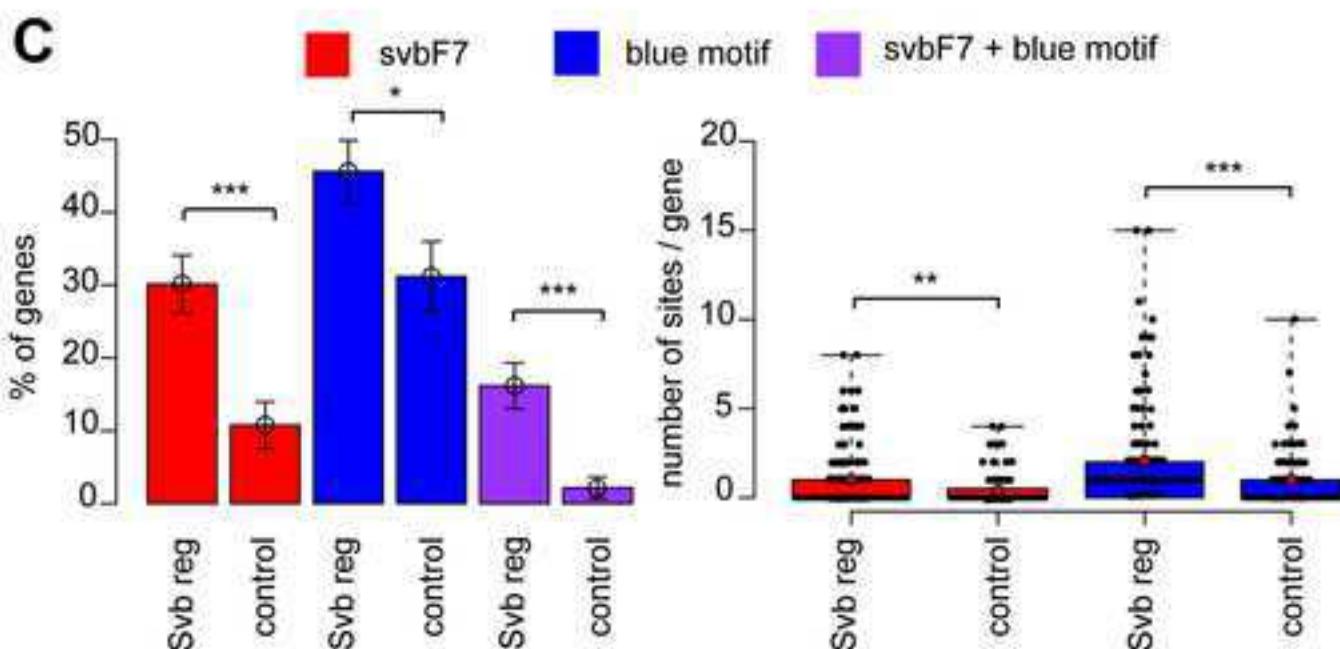
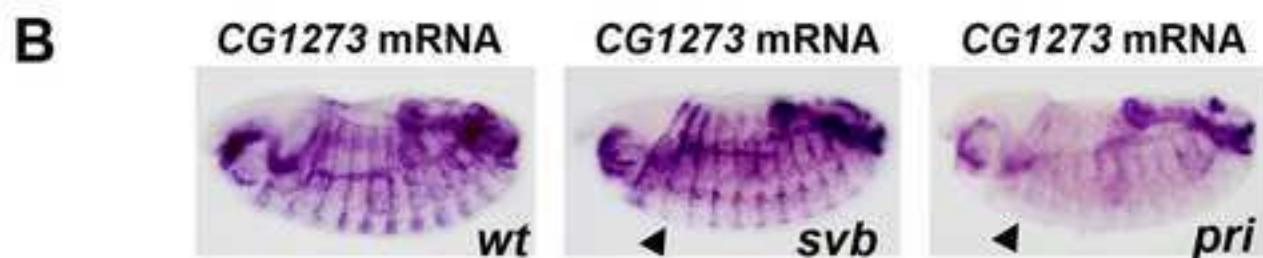
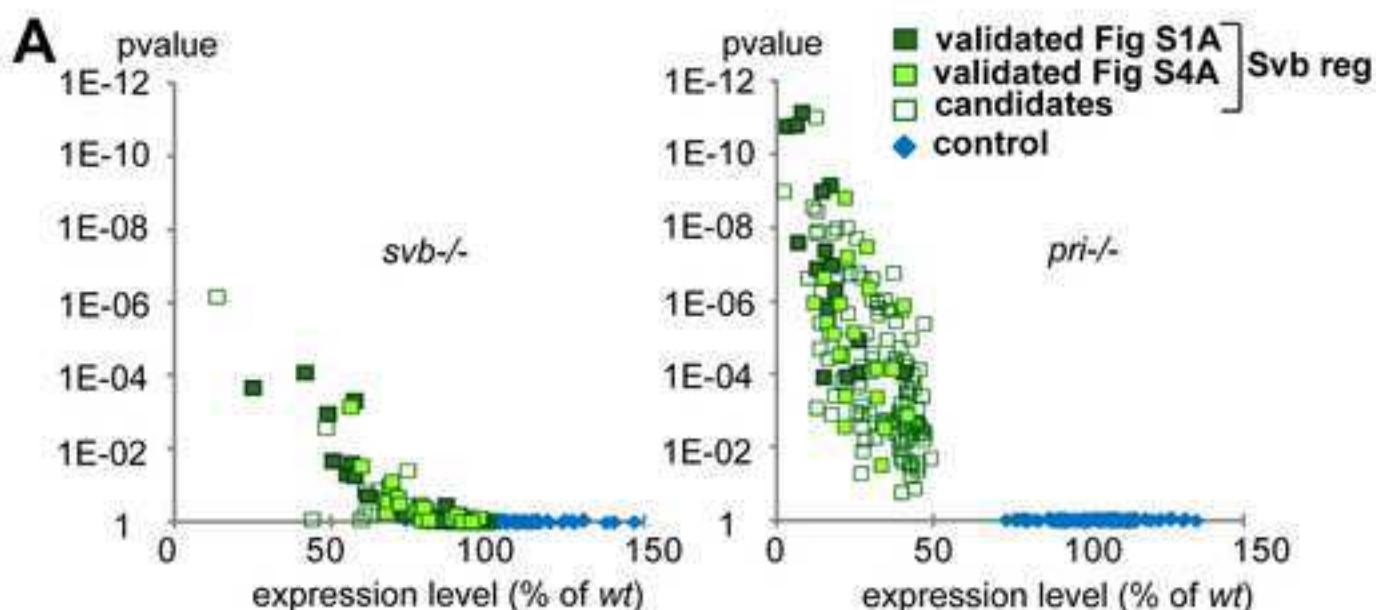
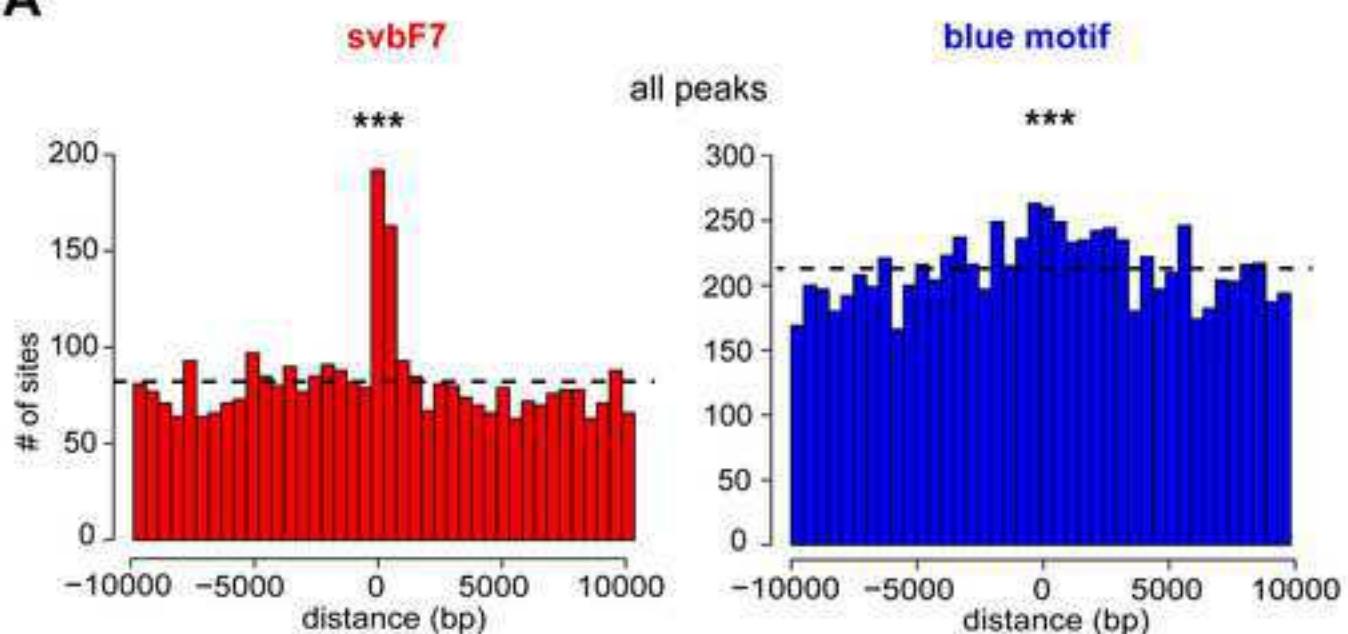
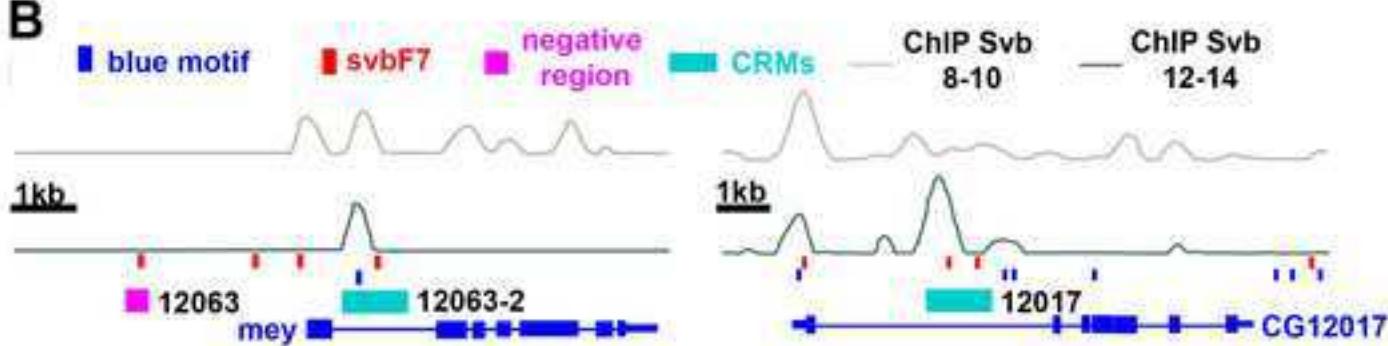


Figure7

[Click here to download high resolution image](#)

Menoret_Fig7

A**B**

12063-2-lacZ



12063-2-lacZ



12017-lacZ

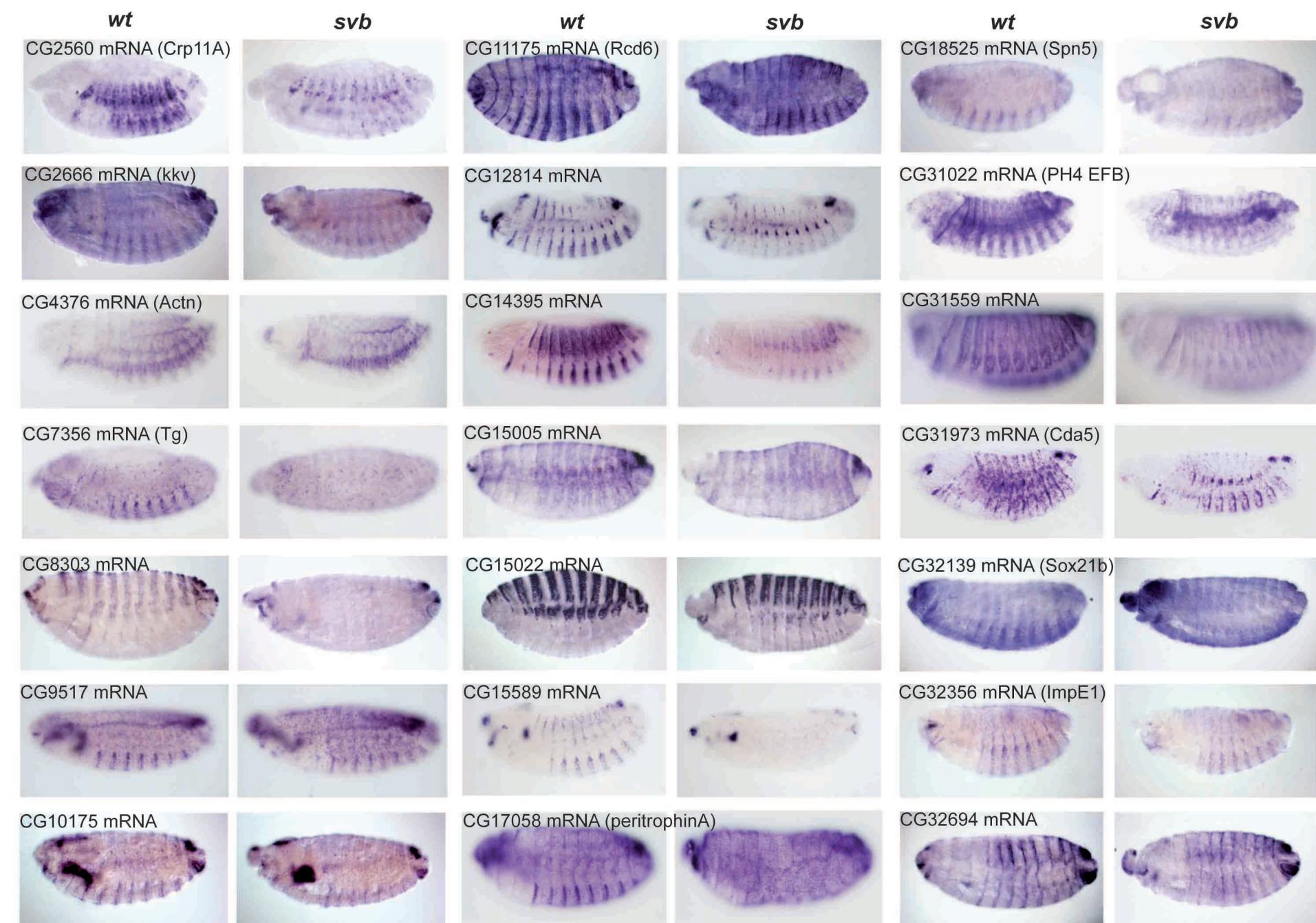


12017-lacZ

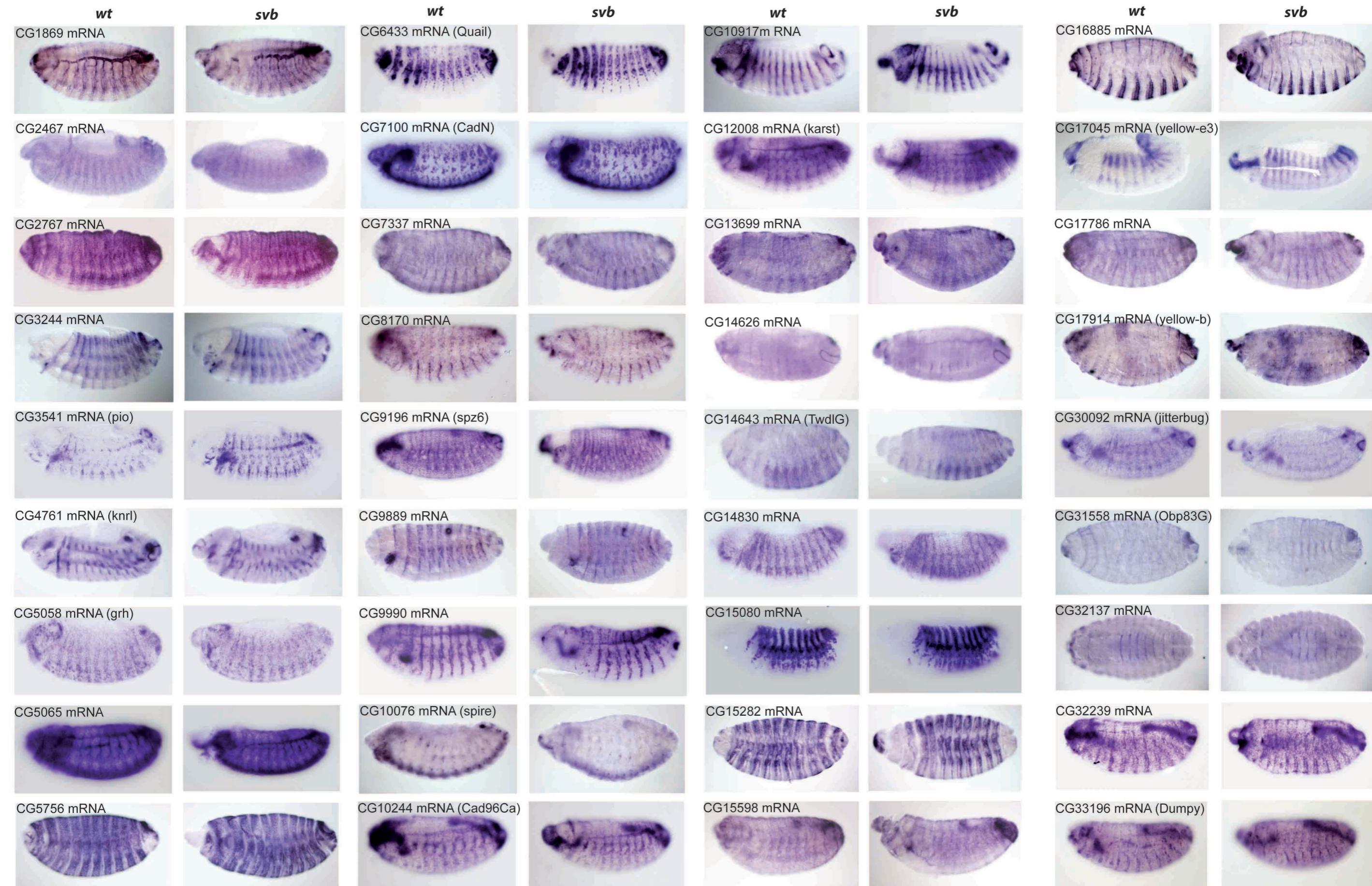


Supporting Information

[Click here to download Supporting Information: Menoret et al_SupMaterial.pdf](#)

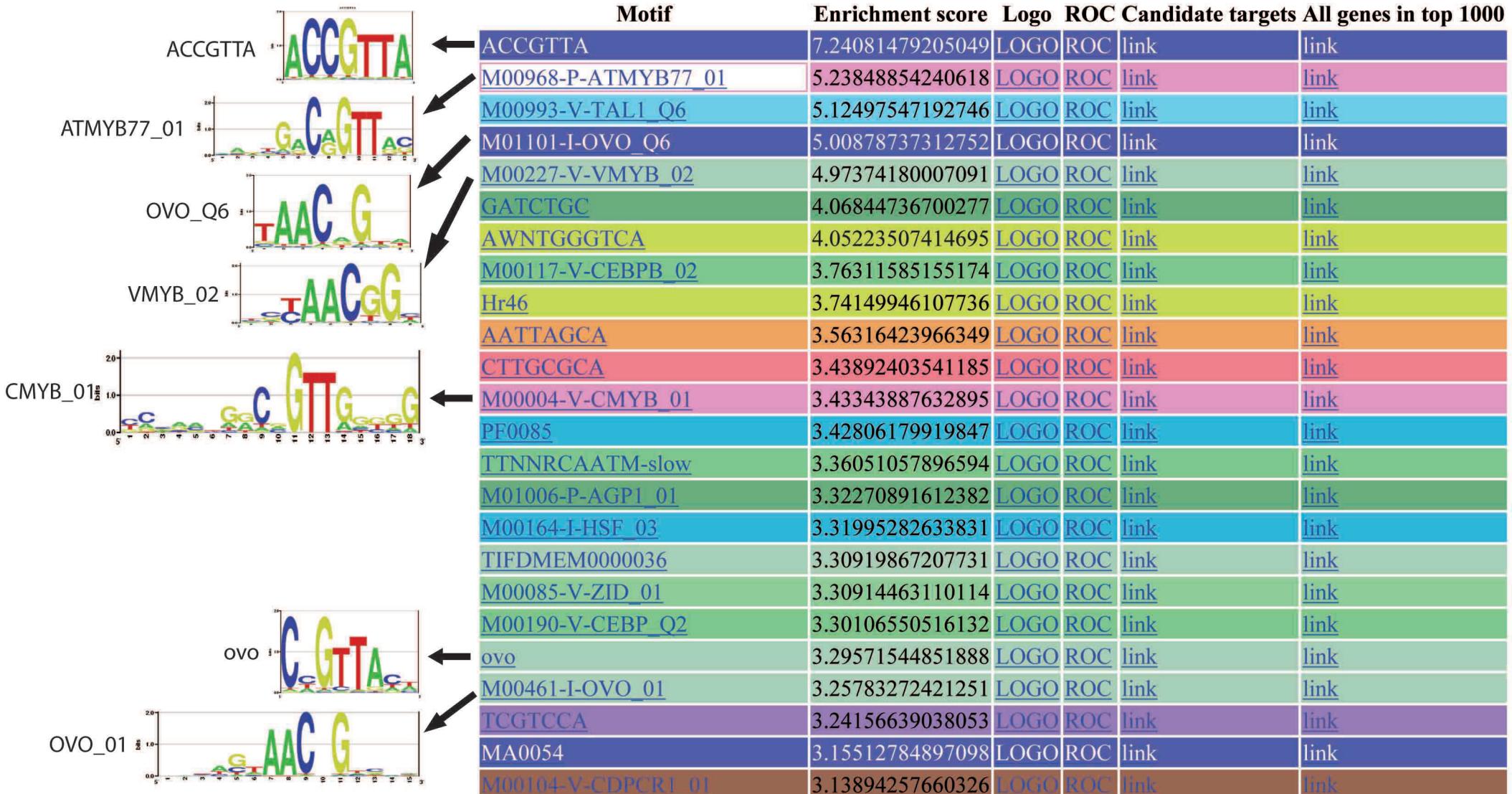


Menoret et al.; Figure S1A



Menoret et al.; Figure S1B

CisTargetX results for 39 Svb downstream genes

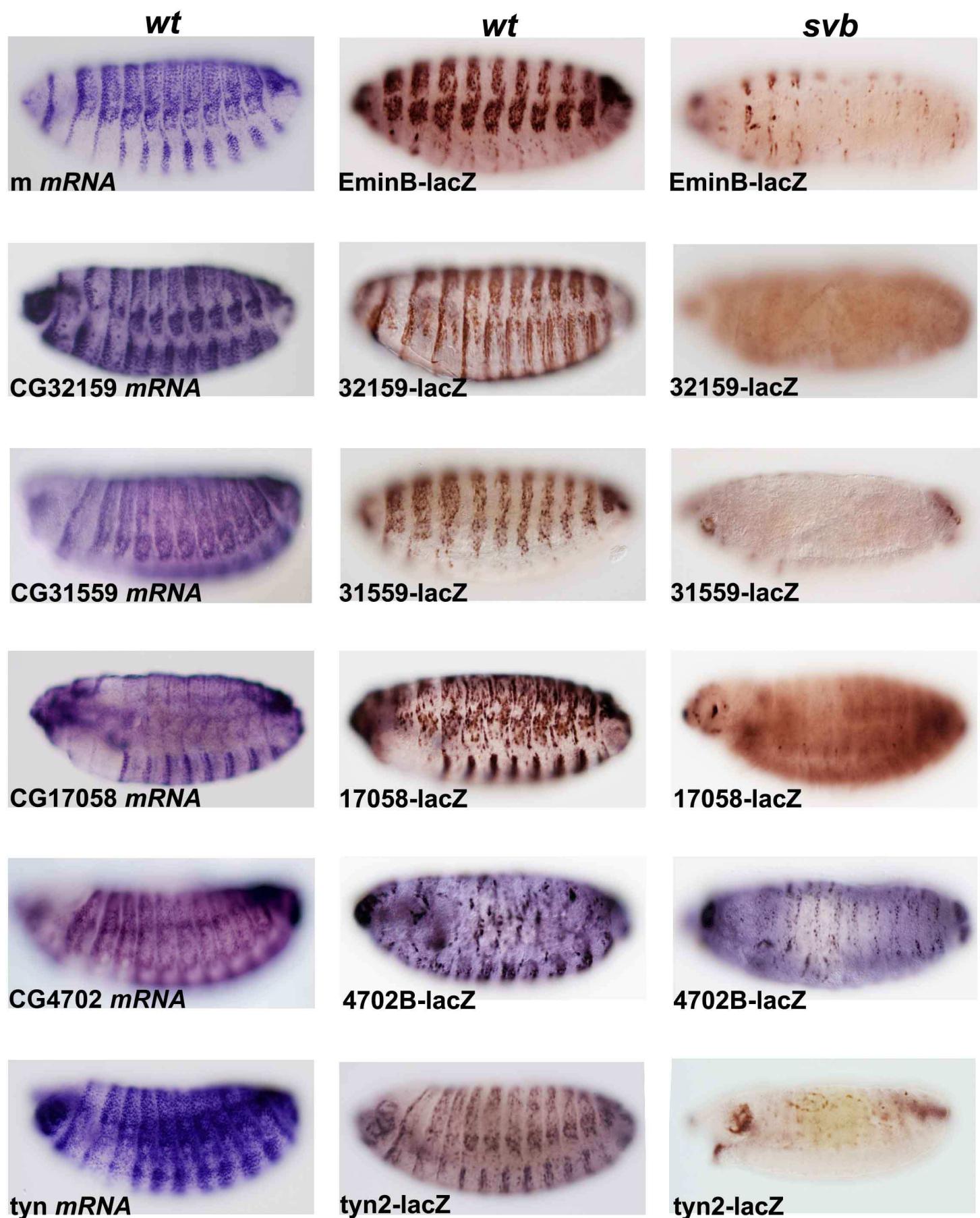


Menoret et al.; Figure S1C

CisTargetX results for 36 epidermal Svb independent genes

Motif	Enrichment score	Logo	ROC	Candidate targets	All genes in top 1000
TIFDMEM0000099	4.51171012170964	LOGO	ROC	link	link
SelexConsensus_grh	4.32098633351667	LOGO	ROC	link	link
M00117-V-CEPB_02	4.26834490169103	LOGO	ROC	link	link
badis_Isgf3g_primary	4.20474107256725	LOGO	ROC	link	link
M00109-V-CEPB_01	4.11706381897675	LOGO	ROC	link	link
MA0050	3.98553697619372	LOGO	ROC	link	link
AATTWNAYGCR	3.9350880238277	LOGO	ROC	link	link
GCANTTGYNWYAATT	3.38916063838056	LOGO	ROC	link	link
badis_Sry_secondary	3.33871168601454	LOGO	ROC	link	link
PF0001	3.30801697357978	LOGO	ROC	link	link
badis_Irf4_primary	3.26638178864145	LOGO	ROC	link	link
M00276-F-MAT1MC_02	3.18304564438102	LOGO	ROC	link	link
pros	3.14138853464811	LOGO	ROC	link	link
ATAATTGC	3.09975334970978	LOGO	ROC	link	link
badis_Irf6_primary	3.04489751362992	LOGO	ROC	link	link
MA0051	3.00324040389701	LOGO	ROC	link	link
PF0085	2.87609852003322	LOGO	ROC	link	link
RWWNTNRCACYT-brachyenteron	2.85193739638815	LOGO	ROC	link	link
CTTGACC	2.82567149246181	LOGO	ROC	link	link
AAGCGCA	2.79938366374089	LOGO	ROC	link	link
PF0146	2.75777040359717	LOGO	ROC	link	link
AGGTGAA	2.74018671833095	LOGO	ROC	link	link
Hr46	2.73135102610866	LOGO	ROC	link	link
badis_Irf5_primary	2.72920239623823	LOGO	ROC	link	link
badis_Irf3_primary	2.72700991677858	LOGO	ROC	link	link
PF0032	2.69410080008959	LOGO	ROC	link	link
PF0168	2.67004930041748	LOGO	ROC	link	link
MA0054	2.65246561515126	LOGO	ROC	link	link
AAAAGCT	2.53404979953681	LOGO	ROC	link	link
badis_Esrra_secondary	2.5252579569037	LOGO	ROC	link	link
ACACGTCA	2.52313125182788	LOGO	ROC	link	link
ATTATGCAA	2.52091684757365	LOGO	ROC	link	link
TTATGCAA	2.5099544502755	LOGO	ROC	link	link
YAATTWNRYGC	2.50559141615085	LOGO	ROC	link	link

Menoret et al.; Figure S1D



Menoret et al.; Figure S2A

Name	Pos/ Neg	svbF7	blue motif	TTATGGAA
dyl2	+	3	3	0
15589	+	3	0	0
32159	+	2	5	1
sha3	+	2	1	1
cyrA	+	2	0	0
Emin	+	1	1	0
sha1	+	1	0	0
nyo1	+	1	0	1
snE1	+	1	1	0
tyn2	+	1	0	0
4702B	+	0	4	0
EminB	+	0	1	0
17058	+	0	0	1
31559	+	0	0	0
nyo3	-	2	0	1
32356	-	1	0	1
12063	-	1	0	0
f4	-	1	0	0
snB2	-	1	0	0
1499-1	-	1	0	0
sn-enh1	-	0	2	0
dyl1	-	0	1	0
f5	-	0	0	0
f3	-	0	0	0
sha2	-	0	0	0
nyo2	-	0	0	2
tyn1	-	0	0	0
sox21b	-	0	0	0
snP	-	0	0	0
4702	-	0	0	0
f2	-	0	0	0
f1	-	0	0	0
dyl3	-	0	0	0
cyrB	-	0	0	0
snE5	-	0	0	0
snH5	-	0	0	0
15013-2	-	0	0	0
15013-1	-	0	0	0
snE4	-	0	0	0

conserved motifs

Name	Pos/Neg	svbF7	blue motif	TTATGGAA
dyl2	+	3	3	0
15589	+	4	1	0
32159	+	3	5	1
sha3	+	2	1	1
cyrA	+	2	0	0
Emin	+	1	1	1
sha1	+	1	0	0
nyo1	+	1	1	1
snE1	+	1	1	1
tyn2	+	1	0	1
4702B	+	1	5	0
EminB	+	1	2	0
17058	+	1	0	1
31559	+	0	0	0
nyo3	-	2	0	1
32356	-	2	0	1
12063	-	1	0	0
f4	-	1	0	1
snB2	-	1	0	1
1499-1	-	1	0	0
sn-enh1	-	1	3	0
dyl1	-	1	2	0
f5	-	0	0	0
f3	-	0	0	1
sha2	-	0	0	0
nyo2	-	0	0	2
tyn1	-	0	0	0
sox21b	-	0	0	0
snP	-	1	0	0
4702	-	0	0	0
f2	-	0	0	0
f1	-	0	0	0
dyl3	-	0	0	0
cyrB	-	0	0	1
snE5	-	0	0	1
snH5	-	0	0	0
15013-2	-	0	0	0
15013-1	-	0	0	0
snE4	-	1	1	0

D. mel motifs

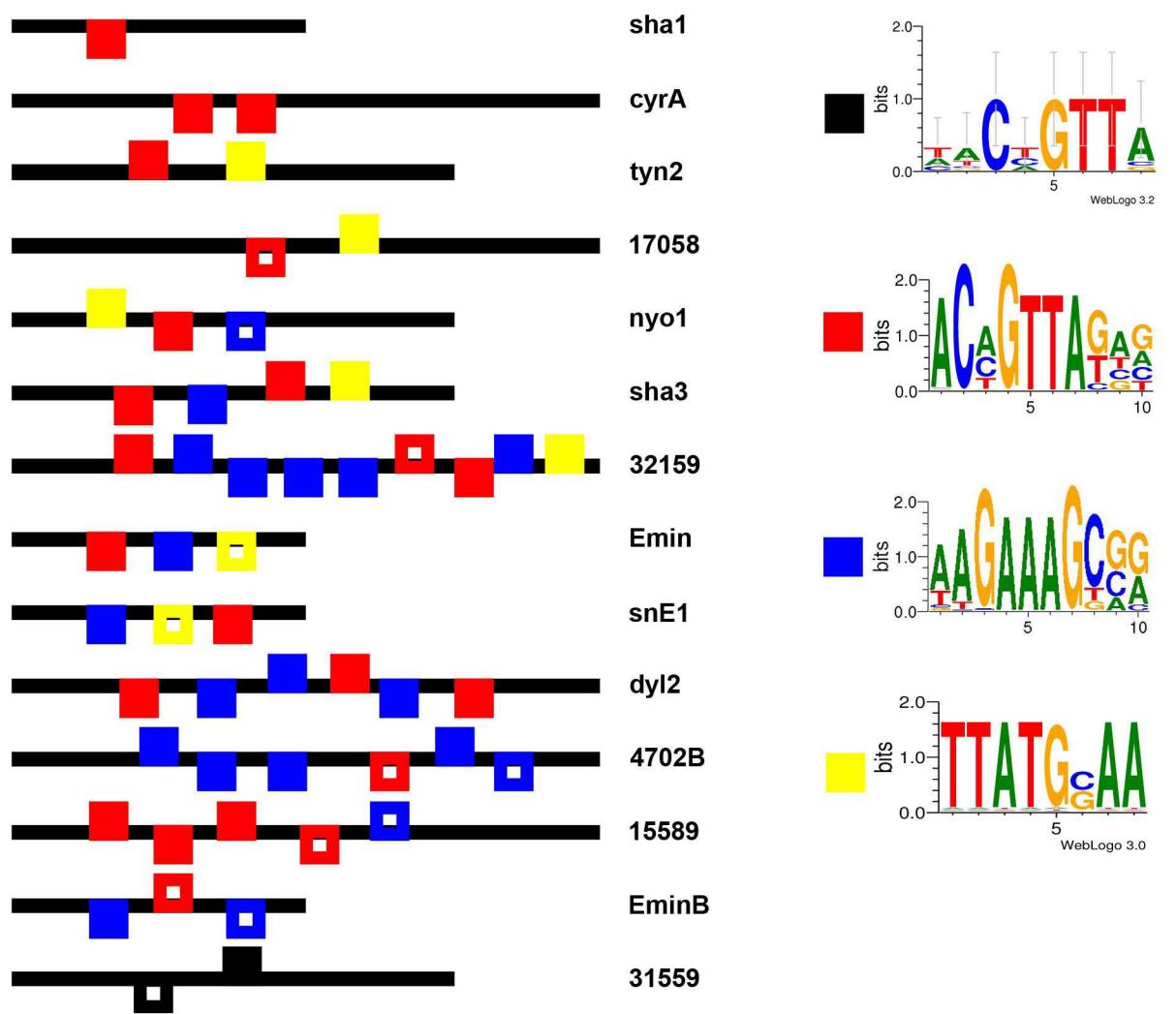
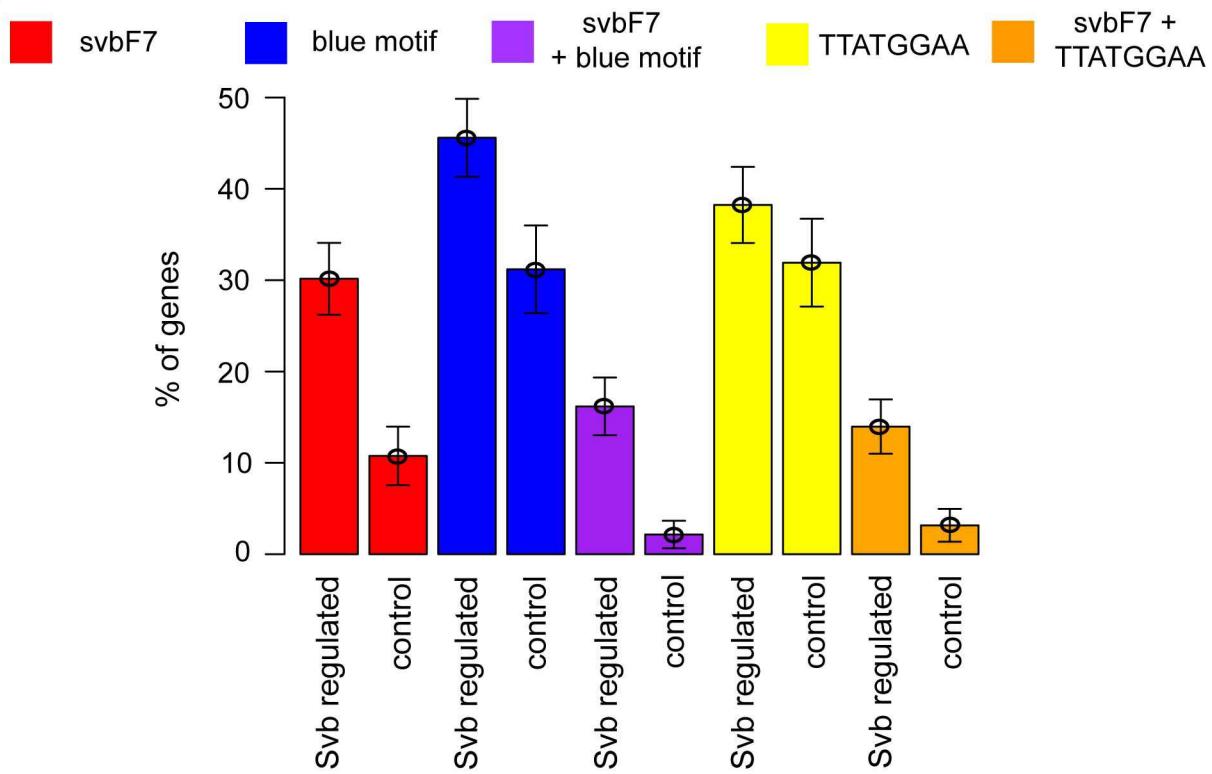
CisTargetX results on the 39 Svb downstream genes with SvbF7

Motif	Enrichment score	Logo	ROC	Candidate targets	All genes in top 1000
SvbF7	7.00488224690521	LOGO	ROC	link	link
ACCGTTA	6.61159001238553	LOGO	ROC	link	link
Bleu motif	4.82608364297727	LOGO	ROC	link	link
nAACwGTw	4.70815627341058	LOGO	ROC	link	link
AATTAGCA	4.52720147022047	LOGO	ROC	link	link
MA0054	4.17331449058271	LOGO	ROC	link	link
M00968-P-ATMYB77_01	4.07885199335089	LOGO	ROC	link	link
M01101-I-OVO_Q6	4.01595564839969	LOGO	ROC	link	link
AWNTGGGTCA	3.99760323428303	LOGO	ROC	link	link
ovo	3.81405287538217	LOGO	ROC	link	link
M00117-V-CEBPB_02	3.75113031269652	LOGO	ROC	link	link
M00227-V-VMYB_02	3.39459534187904	LOGO	ROC	link	link
M00218-P-MYBPH3_01	3.38670380380889	LOGO	ROC	link	link
PF0139	3.26342801641395	LOGO	ROC	link	link
PF0105	3.17174459903407	LOGO	ROC	link	link
M00630-V-FOXM1_01	3.13239177962111	LOGO	ROC	link	link
Hr46	3.11663492121525	LOGO	ROC	link	link
CAGNNGCA	3.05636034970929	LOGO	ROC	link	link
ATCGATC	3.05379101173296	LOGO	ROC	link	link
CGCAAGC	3.04852124710804	LOGO	ROC	link	link
PF0083	2.98289825777381	LOGO	ROC	link	link
GATCTGC	2.93838054467373	LOGO	ROC	link	link
CAATTAG	2.88859306694872	LOGO	ROC	link	link
CGCATGCGC	2.8465660386216	LOGO	ROC	link	link
SelexConsensus_pros	2.80207454325597	LOGO	ROC	link	link
CTTGGCCA	2.76799148846791	LOGO	ROC	link	link
PF0146	2.74182618948447	LOGO	ROC	link	link
M00104-V-CDPCR1_01	2.72604311334416	LOGO	ROC	link	link

CisTargetX CRMs predictions using SvbF7

	15589	cyrA	nyo3	sha2	dy11	dy12	pmin	32159	Emin	sha3	sne1	sne5	EminB	snB2	nyo1	sn_enh1
RANK	1	2	7	10	11	15	21	24	26	34	38	52	55	58	97	104
ACTIVITY	+	+	-	-	-	+	-	+	+	+	+	-	+	-	+	-

Menoret et al., Fig.S2B

A**B**

Menoret et al.; Figure S2C

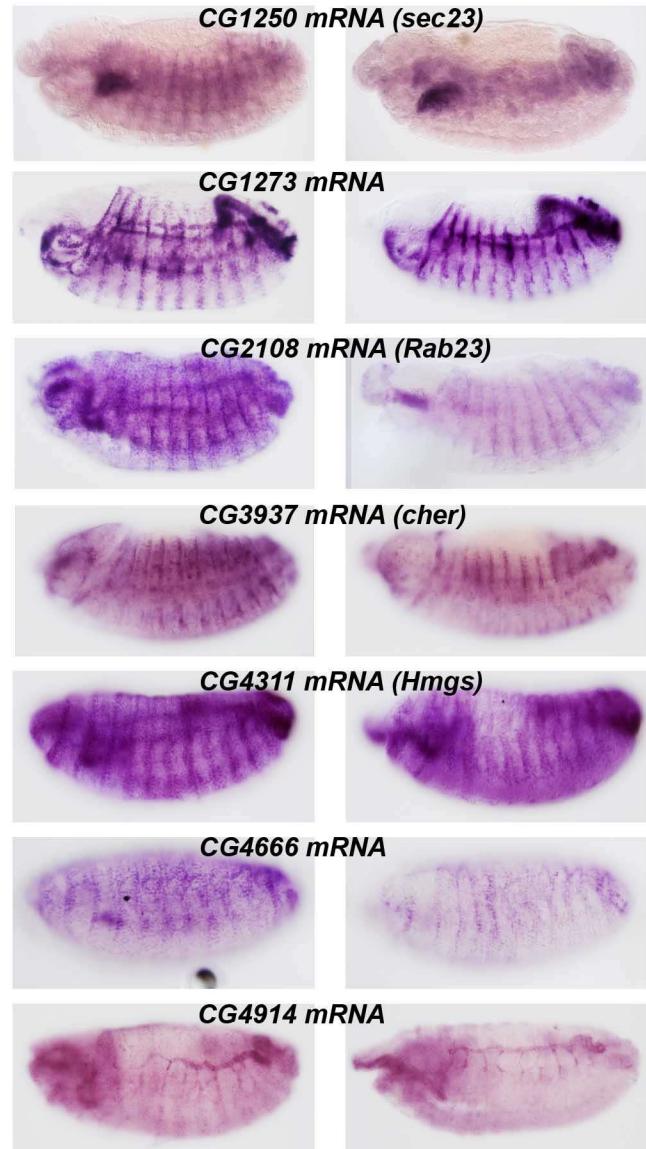
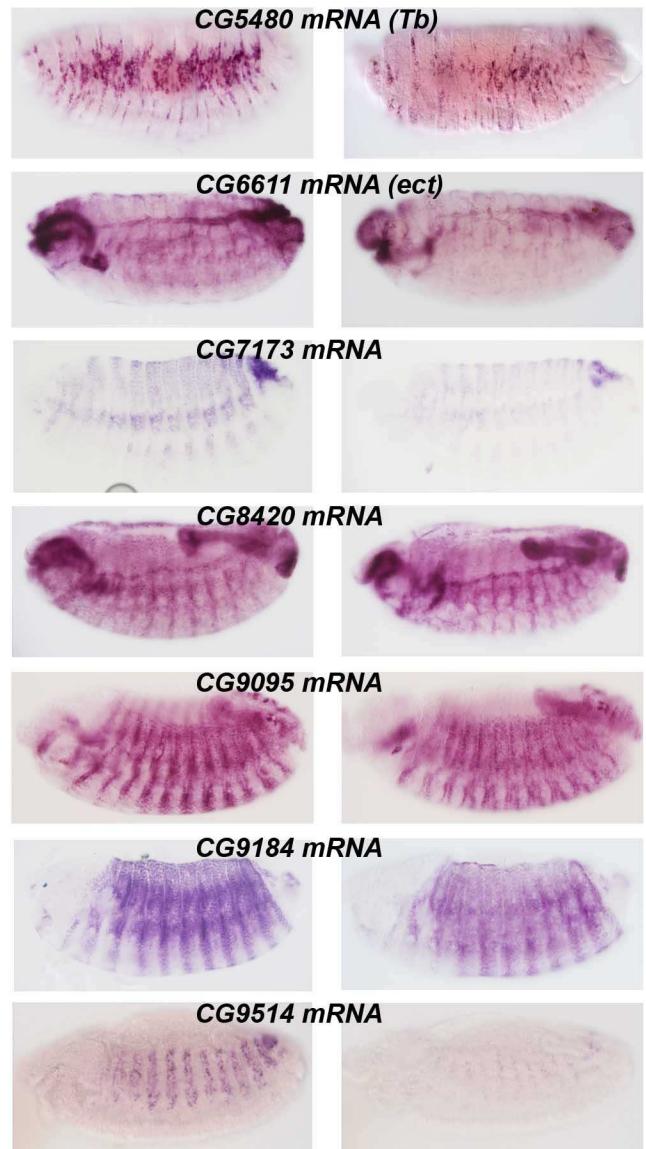
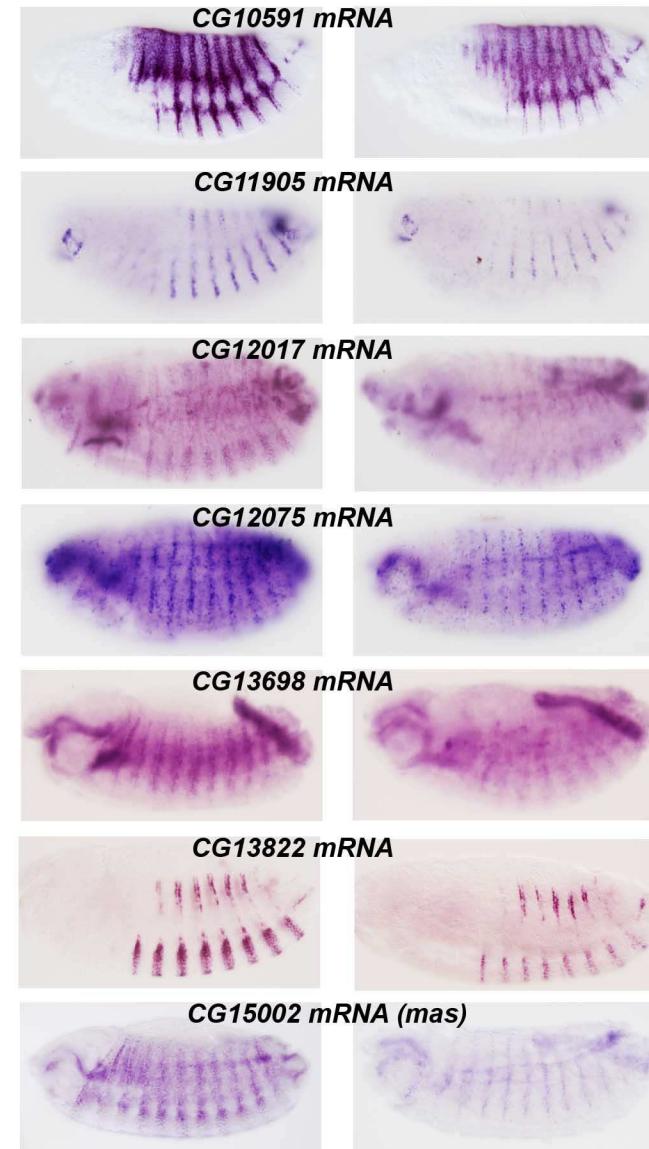
Gene symbol	Representative Public ID	putative function / features	epidermal expression	additional expression #	svb mutant microarrays (% of wt)	pri mutant microarrays (% of wt)	validated in svb mutant (in situ)	validated in svb ectopic (in situ)	svbF7	blue motif	ChIP peaks in 5kb (12-14h)	ChIP peaks in 5kb (8-10h)	reference
CG15370	CG15370	unknown	no	ubiquitous	13,1	2,1	no	ND	no	yes	yes	yes	this work
sha	CG13209	cuticle pattern formation & epidermal cell differentiation	stripes		24,6	6,6	yes	yes	yes	yes	yes	yes	ref 1
CG14395	CG14395	unknown	stripes		41,2	3,0	yes	yes	yes	yes	yes	yes	this work, FigS1
CG4386	CG4386	proteolysis	no	trachea	43,5	11,3	no	ND	no	yes	no	no	this work
CG15818	CG15818	unknown	no	midgut	47,9	17,8	no	ND	no	yes	no	yes	this work
m	CG9369	actin filament organization & epidermal cell differentiation	stripes		48,5	6,2	yes	yes	yes	yes	yes	yes	ref 1
CG32159	CG32159	chitin-based cuticle development	stripes		50,1	12,6	yes	yes	yes	yes	yes	yes	ref 2
nyo	CG1499	regulation of embryonic cell shape	stripes	all gut	54,7	15,3	yes	yes	yes	yes	yes	yes	ref 1
CG4914	CG4914	proteolysis	stripes	hindgut	55,9	21,7	yes	ND	yes	yes	yes	yes	this work, FigS4
neyo	CG7802	regulation of embryonic cell shape	stripes	all gut	56,1	17,3	yes	yes	no	yes	yes	yes	ref 1
mey	CG12063	regulation of embryonic cell shape	stripes		57,1	7,8	yes	yes	yes	yes	yes	yes	ref 1
CG16798	CG16798	Extracellular matrix component	stripes		57,1	18,4	yes	yes	no	yes	yes	yes	ref 1
CHOp24	CG3564	protein secretion	stripes	salivary gland	59,0	27,0	ND	ND	no	yes	no	yes	
CG11905	CG11905	Microtubule associated complex	stripes	trachea	59,3	20,1	yes	no	yes	yes	no	yes	this work, FigS4
CG14356	CG14356	unknown	no	foregut	60,3	27,1	ND	ND	yes	no	no	no	
CG17211	CG17211	unknown	ND	ND	60,5	26,4	ND	ND	no	yes	yes	yes	
CG17780	CG17781	unknown	no	hindgut & anal pad	61,0	27,1	no	ND	no	yes	no	no	this work
mwh	CG13913	actin filament polarization & planar polarity	stripes		61,1	25,6	yes	yes	no	no	yes	yes	ref 1
CG30283	CG30283	proteolysis	stripes	atrium & adult eve PR	62,0	18,3	ND	ND	no	yes	no	yes	
CG4500	CG4500	mesoderm development	ND	ND	65,2	30,3	ND	ND	no	no	yes	yes	
spz6	CG9196	Toll signalling pathway	stripes		66,9	16,5	no	no	no	yes	no	yes	this work, FigS1
CG1273	CG1273	unknown	stripes	all gut & trachea	67,2	20,8	yes	yes	yes	yes	yes	yes	this work, FigS4
CG6785	CG6785	unknown	ND	ND	67,7	21,4	ND	ND	no	no	yes	yes	
CG10591	CG10591	unknown	stripes		67,8	21,7	yes	yes	yes	no	yes	yes	this work, FigS4
CG13698	CG13698	unknown	stripes	gut & salivary gland	68,4	29,3	yes	yes	yes	yes	yes	yes	this work, FigS4
CG9514	CG9514	oxidation reduction process	stripes		69,0	22,4	yes	yes	yes	yes	yes	no	this work, FigS4
CG17562	CG17562	oxidation-reduction process	no	oenocytes	69,2	25,3	no	ND	no	no	no	no	this work
f	CG5424	actin filament assembly & epidermal cell differentiation	stripes		70,3	26,0	yes	yes	yes	yes	yes	yes	ref 1
CG9184	CG9184	unknown	stripes	corpus cardiacum	70,9	15,7	yes	yes	yes	yes	yes	yes	this work, FigS4
CG12075	CG12075	Pleckstrin Homology-like domain	stripes	gut & salivary gland	71,5	33,4	yes	yes	yes	yes	yes	yes	this work, FigS4
tyn	CG17131	actin filament organization & cell matrix adhesion	stripes		72,9	22,2	yes	yes	yes	yes	yes	yes	ref 1
CG4678	CG4678	proteolysis	no	foregut & anal pad	73,4	36,3	ND	ND	yes	no	yes	yes	
CG13616	CG13616	unknown	no	intestine & anal pad	73,4	12,7	ND	ND	no	yes	no	no	
CG14756	CG14756	unknown	no	salivary gland	74,0	12,3	ND	ND	no	yes	no	no	
CG4666	CG4666	unknown	stripes	post spiracle	76,5	24,5	yes	yes	no	yes	yes	yes	this work, FigS4
CG4686	CG4686	unknown	ND	ND	77,0	33,7	ND	ND	no	no	no	no	
CG13082	CG13082	unknown	stripes	all gut	77,7	20,7	ND	ND	yes	yes	no	yes	
CG7173	CG7173	unknown	stripes		77,8	11,6	yes	yes	no	no	yes	yes	this work, FigS4
CG1632	CG1632	proteolysis	ND	ND	78,0	18,9	ND	ND	no	no	yes	yes	
Orct	CG6331	transmembrane transport	no	midgut & fat body	78,1	35,5	ND	ND	no	no	yes	no	
mas	CG15002	proteolysis	stripes	foregut & hindgut & trachea	78,1	31,7	yes	yes	no	yes	yes	yes	this work, FigS4
CG5039	CG5039	unknown	ND	ND	78,6	12,6	ND	ND	no	yes	yes	yes	
CG12017	CG12017	putative thioesterase	stripes		78,6	18,0	yes	yes	yes	yes	yes	yes	this work, FigS4
CG5873	CG5873	oxidation-reduction process	stripes	all gut & trachea	78,7	36,1	ND	ND	yes	yes	yes	yes	
CG8420	CG8420	endoplasmic reticulum	stripes		79,3	15,0	yes	no	yes	yes	yes	yes	this work, FigS4
CG11200	CG11200	metabolic process	stripes	trachea	79,8	32,0	ND	ND	no	no	yes	yes	
Cyp301a1	CG8587	oxidation-reduction process	stripes	trachea	80,3	34,3	ND	ND	yes	no	no	no	
CG42331	CG6879	oxidation-reduction process	ND	ND	80,4	9,8	ND	ND	yes	yes	yes	yes	
CG13822	CG13822	unknown	stripes	lymph gland	80,8	34,3	yes	yes	yes	no	yes	yes	this work, FigS4
CG12009	CG12009	chitin metabolic process	no	trachea	82,4	17,3	ND	ND	yes	yes	yes	yes	

CG8239	CG8239	isoprenoid biosynthesis	stripes	hindgut & anal pad	82,9	12,7	no	ND	no	no	no	no	this work
snRNP-U1	CG5454	mRNA splicing	no	ubiquitous	83,2	38,9	ND	ND	no	yes	no	yes	
CG1140	CG1140	ketone body catabolism	no	midgut & fat body	83,4	39,7	ND	ND	no	no	yes	no	
dyl	CG15013	actin filament organization & cell matrix adhesion	stripes	foregut & hindgut	85,3	16,1	yes	yes	yes	yes	yes	yes	ref 1
Hr46	CG33183	metamorphosis	no	ubiquitous	85,5	39,9	ND	ND	yes	yes	yes	yes	
vri	CG14029	tracheal system development	stripes	hindgut	85,6	30,5	ND	ND	yes	yes	yes	yes	
wus	CG9089	tracheal system development & extracellular matrix organization	stripes	all gut & trachea	86,1	12,6	no	ND	no	no	yes	yes	this work
ImpE1	CG32356	intrinsic to membrane, imaginal disc eversion	stripes	all gut	86,3	14,8	yes	yes	yes	yes	yes	yes	this work, FigS1
Plip	CG10371	protein dephosphorylation	ubiquitous	midgut & hindgut	86,7	32,5	ND	ND	no	no	yes	yes	
CG17672	CG11271	translation	no	ubiquitous	86,7	22,3	ND	ND	no	no	yes	yes	
CG32354	CG32354	unknown	stripes	all gut	86,8	42,3	ND	ND	no	yes	no	yes	
CG8306	CG8306	oxidation-reduction process	stripes	foregut & hindgut	86,8	41,3	ND	ND	no	no	no	yes	
CG10585	CG10585	unknown	stripes	all gut	87,1	27,3	ND	ND	no	yes	no	yes	
fw	CG1500	wing imaginal disc morphogenesis & cell adhesion	no		87,2	30,5	ND	ND	no	yes	yes	yes	
Obp99c	CG7584	sensory perception of chemical stimulus	no	fat body & amnioserosa	87,3	27,4	ND	ND	no	yes	no	yes	
CG5525	CG5525	mitotic spindle organization & microtubule organization	no	muscle system & hindgut	87,4	31,5	ND	ND	no	yes	no	no	
CG34007	Dm.2L.4959.0	unknown	ND	ND	87,5	25,6	ND	ND	no	no	yes	yes	
CG14470	CG14470	unknown	no	hindgut	87,5	39,1	ND	ND	no	no	no	no	
CG18249	CG18249	unknown	no	midgut & amnioserosa	87,6	28,4	ND	ND	no	no	yes	yes	
CG8386	CG8386	lateral inhibition	ubiquitous	salivary gland	87,9	41,5	ND	ND	no	no	no	no	
pwn	CG11101	EGF-like calcium binding, chaeta morphogenesis	ND	ND	88,0	29,8	ND	ND	no	yes	yes	yes	
PH4alpha SG1	CG31014	oxidation-reduction process & salivary gland development	no	salivary gland	88,1	39,7	ND	ND	no	no	no	no	
CG6415	CG6415	glycine catabolic process	no	fat body	88,2	37,1	ND	ND	no	yes	yes	yes	
cher	CG3937	actin binding	stripes	muscle system	88,3	40,5	yes	no	yes	yes	no	yes	this work, FigS4
CG2016	CG2016	unknown	stripes	all gut & trachea	88,4	22,9	ND	ND	no	yes	no	no	
Tb	CG5480	chitin-based cuticle development	stripes		88,6	21,5	yes	yes	no	yes	yes	no	this work, FigS4
CG10932	CG10932	mitotic spindle organization	no	midgut	88,8	34,6	ND	ND	no	no	no	yes	
scu	CG7113	ecdysone metabolism	no	midgut	89,1	39,8	ND	ND	no	no	no	no	
CG7860	CG7860	salivary gland autophagic cell death	no	midgut & crystal cells	89,7	15,8	ND	ND	no	yes	no	yes	
CG9356	CG9356	unknown	no		89,7	39,8	ND	ND	no	no	no	yes	
Hmgs	CG4311	hydroxymethylglutaryl-CoA synthase	stripes	foregut & hindgut	90,0	28,7	yes	no	no	no	yes	yes	this work, FigS4
CG11836	CG11836	proteolysis	ubiquitous		90,3	32,3	ND	ND	no	no	yes	no	
CG9503	CG9503	oxidation reduction process	no	dorsal trunk	90,4	22,4	ND	ND	no	no	yes	yes	
CG1837	CG1837	apoptotic cell clearance	no	ubiquitous	90,4	39,5	ND	ND	yes	no	yes	yes	
tw	CG12311	somatic muscle development	ND	ND	90,5	38,0	ND	ND	no	no	no	yes	
PH4alpha EFB	CG31022	oxidation reduction process	stripes	muscle system & plasmacytocytes	90,6	40,5	yes	yes	yes	yes	yes	yes	this work, FigS1
amd	CG10501	chitin-based cuticle development	ND	ND	90,6	38,1	ND	ND	yes	no	yes	yes	
Rab23	CG2108	GTPase, morphogenesis of a polarized epithelium	stripes	foregut	91,5	31,9	yes	yes	no	yes	yes	yes	this work, FigS4
Smn	CG16725	neuromuscular junction development	no	gonad	91,5	37,6	ND	ND	no	no	no	yes	
Pros28.1	CG3422	ubiquitin dependent proteolysis	no	ubiquitous	91,6	28,1	ND	ND	yes	no	yes	yes	
CG9175	CG9175	cell morphogenesis & protein secretion	no	midgut & hindgut& salivary gland	91,6	40,9	ND	ND	no	no	no	yes	
Rlc1	CG9378	translation	ND	ND	91,8	45,2	ND	ND	no	yes	no	yes	
Lip4	CG6113	lipide metabolism	no	amnioserosa	91,9	44,5	ND	ND	no	no	yes	no	
CG7840	CG7840	lipide metabolism	no		91,9	35,2	ND	ND	yes	no	no	no	
Gtp-bp	CG2522	protein secretion	stripes	midgut & hindgut& salivary gland	92,1	39,0	ND	ND	no	no	no	yes	
CG32250	CG32250	transport	ND	ND	92,2	41,7	ND	ND	no	no	no	no	
CG15506	CG15506	unknown	ND	ND	92,2	13,6	ND	ND	yes	yes	yes	yes	
TRAM	CG11642	protein targeting to membrane	stripes	salivary gland	92,3	36,0	ND	ND	yes	no	no	yes	
CG6704	CG6704	unknown	no	yolk nuclei	92,5	16,3	ND	ND	no	yes	no	no	
CG17218	CG17218	tracheal system development	stripes	all gut & anal pad	92,5	40,8	ND	ND	no	no	no	no	
CG4065	CG4065	unknown	no	midgut & muscle system	92,6	46,3	ND	ND	no	no	yes	yes	
mRpL46	CG13922	unknown	no	midgut & muscle system	92,7	42,8	ND	ND	no	no	no	no	

CG6180	CG6180	unknown	no	midgut	92,7	44,0	ND	ND	no	yes	no	yes	
T-cp1	CG5374	protein folding & mitotic spindle organization	no	ubiquitous	92,9	41,1	ND	ND	yes	no	yes	no	
CG13585	CG13585	unknown	no	garland cell	93,0	45,3	ND	ND	no	no	no	no	
nimb3	Dm.2L.8912.0	unknown	ND	ND	93,6	29,5	ND	ND	no	no	no	no	
CG2663	CG2663	transport	head	post spiracle	93,7	41,8	ND	ND	no	yes	yes	yes	
CG11786	CG11786	unknown	no	dorsal trunk	93,8	27,0	ND	ND	no	no	yes	no	
rt	CG6097	somatic muscle development & synaptic activity	ND	ND	94,0	25,7	ND	ND	no	yes	yes	yes	
CG13627	CG13627	unknown	no	trachea	94,2	16,2	ND	ND	no	yes	no	no	
Gmap	CG33206	protein targeting to Golgi	ND	ND	94,3	41,8	ND	ND	no	no	no	yes	
Rlc1	CG9378	translation	ND	ND	94,3	40,5	ND	ND	no	yes	no	yes	
CG6672	CG6672	transmembrane transport	ND	ND	94,4	42,0	ND	ND	no	yes	no	yes	
CG4702	CG4702	unknown	stripes	all gut	94,5	16,9	yes	yes	no	yes	yes	yes	ref 1
CG3831	CG3831	unknown	no	corpus allatum	94,5	37,3	ND	ND	yes	no	no	yes	
sec23	CG1250	chitin-based cuticle development, cell polarity & transport	stripes	salivary gland	94,5	41,8	yes	ND	no	no	yes	yes	this work, FigS4
CG31559	CG31559	redox homeostasis	stripes	foregut	94,8	14,3	yes	yes	no	yes	yes	yes	this work, FigS1
CG11771	CG11771	proteolysis	no	all gut & muscle system	95,0	45,4	ND	ND	yes	no	no	yes	
CG9095	CG9095	cell adhesion	stripes		95,4	36,9	yes	yes	yes	yes	yes	yes	this work, FigS4
--	GM01028	unknown	ND	ND	95,6	42,4	ND	ND	no	no	yes	no	
CG32039	CG32039	unknown	no		95,7	26,7	ND	ND	no	no	yes	yes	
CG1753	CG1753	cystein biosynthesis	no	midgut	96,0	26,3	ND	ND	no	yes	yes	yes	
CG4822	CG4822	unknown	ND	ND	96,3	43,6	ND	ND	yes	yes	yes	no	
CG8112	CG8112	unknown	no	yolk nuclei	96,8	30,2	ND	ND	no	yes	yes	yes	
ect	CG6611	tube development	stripes	oesophagus & anal pad	96,9	19,7	yes	no	yes	yes	yes	yes	this work, FigS4
CG15239	CG15239	unknown	stripes	salivary gland	97,0	19,4	ND	ND	yes	yes	yes	yes	
CG9689	CG9689	unknown	stripes	oesophagus & post spiracle	97,1	43,7	ND	ND	no	yes	no	yes	
mRpl45	CG6949	translation & transport	no	ubiquitous	97,2	22,6	ND	ND	no	no	no	yes	
CG8213	CG8213	proteolysis	ND	ND	97,3	13,7	ND	ND	yes	yes	no	yes	
CG2263	CG2263	phenylalanyl-tRNA aminoacylation	no	ubiquitous	97,4	46,9	ND	ND	no	yes	yes	yes	
Rpb8	CG11246	transcription	no	midgut	97,8	45,1	ND	ND	no	no	yes	yes	
qua	CG6433	actin filament organization	stripes	all gut & lymph gland	97,8	43,1	no	no	yes	yes	no	yes	this work, FigS1
CG11227	CG13630	proteolysis	ND	ND	97,8	47,0	ND	ND	yes	no	no	no	
CG9205	CG9205	unknown	ND	ND	98,0	44,9	ND	ND	no	yes	no	yes	
Nf-YA	CG3891	transcription & phagocytosis	no	ubiquitous	98,3	39,6	ND	ND	no	yes	yes	yes	
kar	CG12286	transmembrane transport	no	midgut & amnioserosa	98,3	42,8	ND	ND	no	no	no	yes	
CG31717	CG31717	unknown	ND	ND	98,4	45,2	ND	ND	yes	yes	no	no	
Past1	CG6148	endocytosis	no	midgut & salivary gland	98,4	46,6	ND	ND	no	yes	yes	yes	
bw	CG17632	transport & eye pigment biosynthesis	no	malpighian tubule & fat body	98,4	37,4	ND	ND	no	yes	no	yes	
mRpL51	CG13098	translation	no	midgut & muscle system	98,5	39,3	ND	ND	no	no	no	yes	
pk	CG11084	planar polarity	no	ubiquitous	98,6	45,5	ND	ND	yes	yes	yes	yes	
CG5171	CG5171	trehalose biosynthesis	no	amnioserosa & yolk nuclei	98,9	37,8	ND	ND	no	yes	yes	yes	
CG13365	CG13365	unknown	no		98,9	43,2	ND	ND	no	yes	yes	yes	
CG5742	CG5742	neurogenesis	no	ubiquitous	98,9	43,8	ND	ND	no	no	no	yes	
KD	CG7125	protein phosphorylation & intracellular signal transduction	ND	ND	99,0	46,5	ND	ND	no	yes	yes	yes	
CG11127	CG11127	unknown	no		99,1	47,2	ND	ND	no	yes	no	yes	
Fib	CG9888	centrosome organization & RNA processing	no	ubiquitous	99,1	49,2	ND	ND	no	no	no	yes	
CG30423	CG30423	unknown	ND	ND	99,1	43,0	ND	ND	yes	yes	no	no	
CG3842	CG3842	oxidation reduction process	stripes	foregut & anal pad	99,2	14,1	ND	ND	no	no	yes	yes	
CG15743	CG15743	phosphatidylinositol phosphorylation	no	salivary gland	99,9	44,9	ND	ND	no	no	no	yes	

: from BDGP and experimental approaches

Menoret et al.; Figure S3

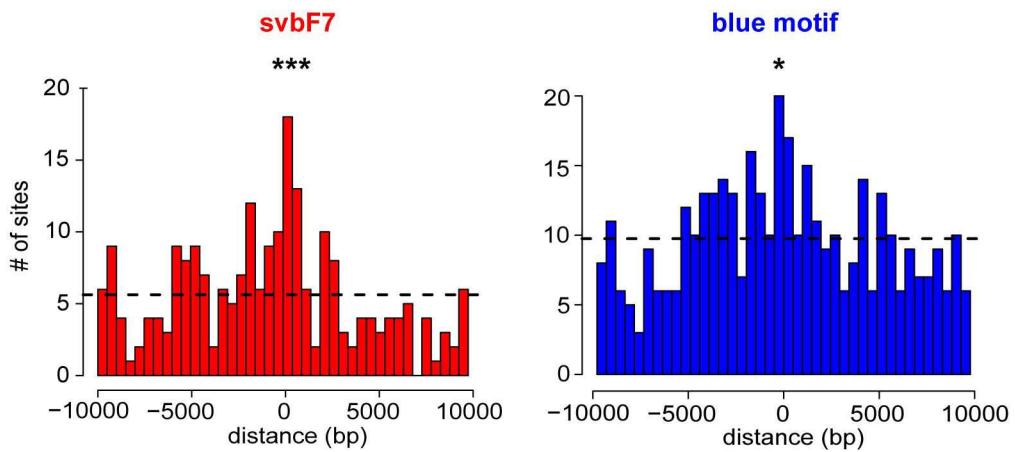
*wt**svb**wt**svb**wt**svb*

Menoret et al.; Figure S4A

CisTargetX results for 150 Svb regulated genes

Motif	Enrichment score	Logo	ROC	Candidate targets	All genes in top 1000
ken_SANGER_10_FBgn0011236	7.37531372161641	LOGO	ROC	link	link
ACCGTTA	6.84383162504572	LOGO	ROC	link	link
ovo_SOLEXA_5_FBgn0003028	6.63292929690881	LOGO	ROC	link	link
ken_SOLEXA_5_FBgn0011236	6.42715336468385	LOGO	ROC	link	link
→ blue motif	6.25223867875159	LOGO	ROC	link	link
→ svbF7	5.65047866631769	LOGO	ROC	link	link
M01101-I-OVO_Q6	5.06408069648283	LOGO	ROC	link	link
M00972-V-IRF_Q6_01	5.05374217897479	LOGO	ROC	link	link
AAAGTGANA	5.0160914584977	LOGO	ROC	link	link
Hr46_FlyReg_FBgn0000448	4.85317836834593	LOGO	ROC	link	link
Hr46	4.85317836834593	LOGO	ROC	link	link
Motif_1_f55t8-1x2	4.49654237716018	LOGO	ROC	link	link
M01138-V-RORA_Q4	4.46741278968724	LOGO	ROC	link	link
MA0389.1-SRD1	4.46228639377529	LOGO	ROC	link	link
M01126-P-BPC1_Q2	4.43823176680381	LOGO	ROC	link	link
M00028-I-HSF_01	4.37308024685257	LOGO	ROC	link	link
M00157-V-RORA2_01	4.32682266768719	LOGO	ROC	link	link
AWNTGGGTCA	4.31307226795011	LOGO	ROC	link	link
M00063-V-IRF2_01	4.27540440233619	LOGO	ROC	link	link
MA0050.1-IRF1	4.24622337945276	LOGO	ROC	link	link
MA0072.1-RORA_2	4.17594546358957	LOGO	ROC	link	link
RAGTGAAGT	4.12277839428199	LOGO	ROC	link	link
MA0302.1-GAT4	3.91703675233069	LOGO	ROC	link	link
M00029-F-HSF_01	3.87588842394043	LOGO	ROC	link	link
kni_NAR_FBgn0001320	3.87240796116408	LOGO	ROC	link	link
M00767-V-FXR_IR1_Q6	3.8106854685787	LOGO	ROC	link	link
nAACwGTw	3.75928434836453	LOGO	ROC	link	link

Cross-correlation between conserved svbF7 or blue motif instances and ChIP peaks within Svb-regulated genes

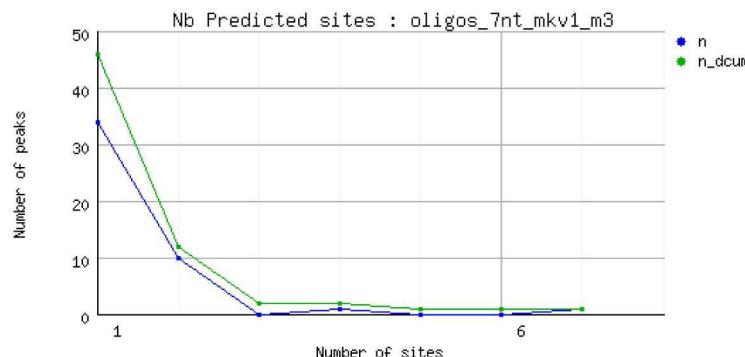
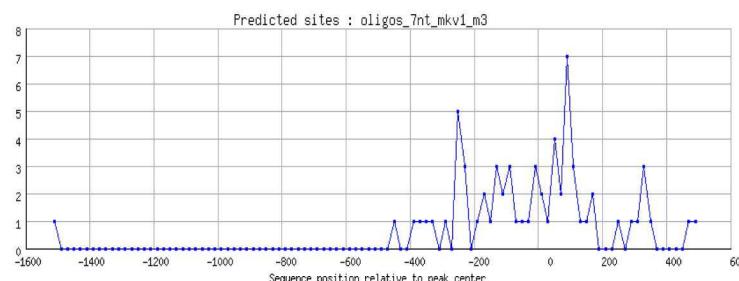
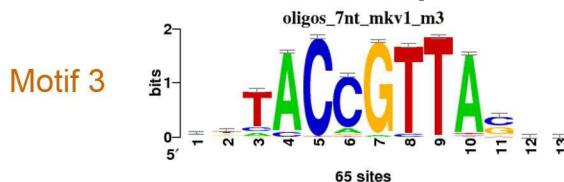


RSA tools "peak-motifs" results from microarrays Svb regulated genes

All discovered motifs

Motif 1	aCrCACACaCaC	Motif 6	wrCCGGwTAhv
Motif 2	atAwatAhATATdTatwtat	Motif 7	yaaATATAwATATAtat
Motif 3	rktACC GTTAsck	Motif 8	caYaCACACACACaC
Motif 4	srAGGCATGrg	Motif 9	wgCTTTTCasy
Motif 5	rgGGACTACwa		

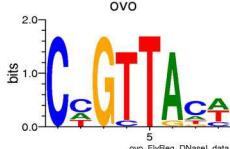
Discovered motifs with motif comparison



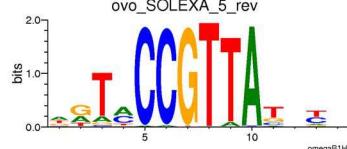
► jaspar_core_vertebrates : no matches

► FlyFactorSurvey : 3 matches

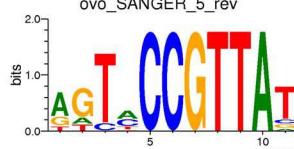
ovo_FlyReg
(aligned 6/9 bases)



ovo_SOLEXA_5
(aligned 8/9 bases)



ovo_SANGER_5
(aligned 8/9 bases)

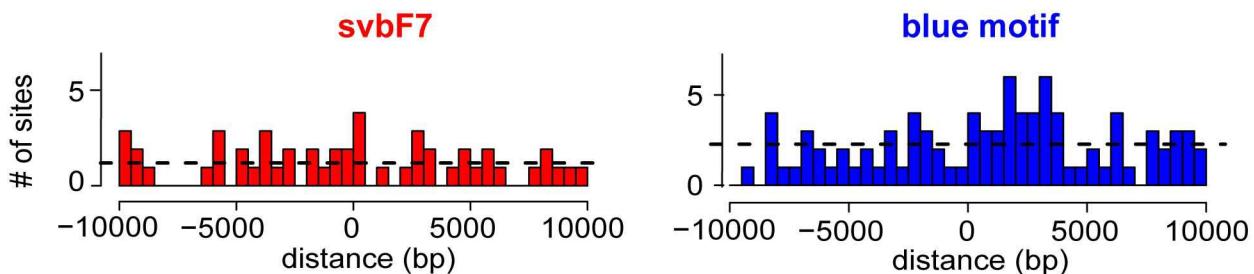


► DMMPMM_drosophila : 1 match

ovo
(aligned 5/9 bases)



Cross-correlation between conserved svbF7 or blue motif instances and ChIP peaks within control genes

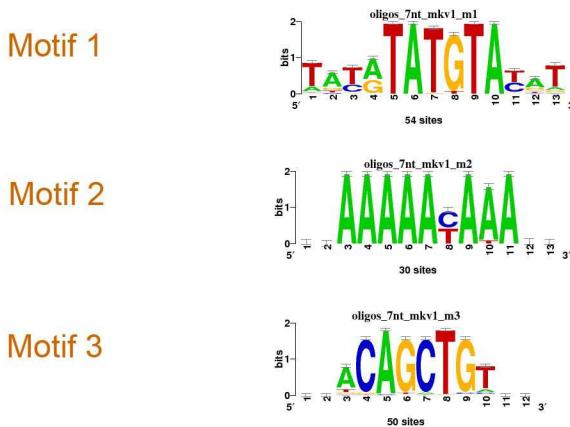


RSA tools "peak-motifs" results from microarrays control genes

All discovered motifs

Motif 1	tayrTATGTAyrt	Motif 6	raCAGCTGTy
Motif 2	rsAAAAAYAAAak	Motif 7	bataTACATAACata
Motif 3	sraCAGCTGtys	Motif 8	mmACAAmAACAAaAcrcA
Motif 4	wwkCTAGTrCCbt		
Motif 5	rcGCGCGCGCgy		

Discovered motifs with motif comparison



Menoret et al.; Figure S5B

Figure S1 : In situ Hybridization of epidermal genes (from BDGP) and motifs predictions candidate genes.

A: identification of 21 additional Svb-dependent genes. Genes having an epidermal expression pattern were chosen from the BDGP expression pattern project and tested in *svb* mutant embryos. Their expression is decreased in *svb* mutants.

B: 36 epidermal genes showing no significant decrease of their expression in *svb* mutant embryos.

C: Motifs predictions in the set of 39 Svb downstream genes using cisTargetX: predicted motifs are ranked according to their enrichment and conservation. Related motifs are shown using the same color code. Ovo binding sites or motifs related to (CnGTT) are detected on top.

D: Motifs Predictions in epidermal control genes using cisTargetX. Motifs enriched in the set of 36 epidermal genes (not Svb-dependent) are ranked. Motifs associated to the epidermal transcription factors (grh, cEBP (vri in drosophila)...) are found on top of the detected motifs, but not OvoQ6.

Figure S2 : CRM activity and *de novo* motif discovery.

Several fragments were selected through different parameters (single, multiple Ovo BS, evolutionary conserved or not) and assayed for *in vivo* CRM activity in transgenic assays.

A: expression pattern of 5 CRMs not predicted by cisTargetX on the basis of conserved Ovo clusters. These CRMs contain only a single Ovo binding site (sometimes not conserved) and are able to sustain *svb* dependent epidermal activity

B: number of conserved and non-conserved motifs (Dmel) in the various genomic fragment assayed for CRM activity *in vivo*. The numbers of conserved binding sites are listed. Note that svbF7 is more discriminative between positives and negatives, the blue and TTATGCAA motifs being less discriminant. Note that blue and TTATGCAA motifs are frequently located in a mutually exclusive manner. CisTarget prediction using SvbF7: In the set of 39 Svb downstream, SvbF7 PWM becomes the most significant detected motif. CRM prediction leads to the detection of less negative (in comparison with Fig2: not detected are not shown)) and prediction of few new positives (Emin, EminB, 32159)

C: on top, schematic representation of CRM architecture (binding sites location and orientation). Open boxes indicate non-conserved sites. Note that several grammars are found (variation in the type, number, orientation of the binding sites). Bottom: svbF7+blue or svbF7+TTATGCAA motifs combinations perform equally to predict TGs.

Figure S3 : High confidence list of genes regulated by Svb from microarrays experiments.

From microarrays analysis, genes are selected according to a significant level of expression in wild type condition. From this set of 5000 genes, 150 of them display a decrease of expression in *svb* mutant and *pri* mutant embryos. Genes are ranked according to their respective level of expression in *svb* mutant embryos compared to the wild type. % of residual expression relative to wt is indicated in *pri* and *svb* mutants. Putative gene function is indicated. Validation of downregulated genes by In Situ Hybridization is performed in *svb* mutants and in ectopic *svb* expression. For each gene, expression pattern, predictions for svbF7 and blue motifs, and ChIP peak in a 5kb window are mentioned. Expression pattern is indicated when existing. Co-occurrence of in silico prediction and ChIP are shown, giving rise to high confident *svb* direct target genes. In green are summarized the *svb*-regulated genes. In grey color bacground, genes tested but not *svb*)dependent.

Figure S4 : In situ Hybridization and motifs predictions in svb downstream set defined from genome wide approach.

A: validation of new Svb downstream genes by In situ Hybridization experiments of microarrays. 21 genes downregulated in *svb* and *pri* mutants from microarrays were assayed in wild type embryos (left panels) and tested for residual expression in *svb* mutant embryos (right panels).

B: cisTargetX predicted motifs. svbF7 and blue motif are highly ranked using this set of downstream gene.

Figure S5 : Motif analysis of svb ChIP peaks associated of Svb regulated (A) and non-regulated genes (microarrays) (B).

On top : cross-correlation between conserved svbF7 or blue motif instances and Svb ChIP-Seq peaks among Svb-regulated (A) and control set (B) of genes

Bottom: Sequences of peaks associated to Svb regulated genes and non-regulated genes (5kb upstream and downstream + introns) are subjected to a de novo motif discovery approach using the RSAT “peak motif” software. Discovered motifs are listed and subjected to motif comparison.

A: For peaks associated to *svb* regulated genes, Motif 3 is clearly related to svbF7 and to the Ovo PWM previously reported. The distributions of the predicted sites respective to the peak of the ChIP fragments are shown (left graph). The number of sites (clusters) found in ChIP fragments is also indicated (right graph). Motif9 appears related to the blue motif. Bottom: Logos of PWM for known TFs. Note that all are Ovo “signature”

B: For peaks associated to *svb* non regulated genes, we cannot find a motif that corresponds to svbF7 motif, indicating that svbF7 is specific of Svb bound fragments.

SUPPLEMENTARY METHOD

De novo motif generation

We used the phylogeny-based *de novo* motif generation algorithm described in [Rouault et al. 2011], available on the website (<https://github.com/hrouault/Imogene/>). The 14 positive CRM (Fig.S2) were used as the training set for the algorithm and scanned for conserved motifs as described in [Rouault et al, 2010]. The score threshold for motif generation, which sets the searched PWM information content, was varied from 7 to 13 bits in different runs of the algorithm, with a motif width set to 10 bp. In each run, the 5 highest scoring motifs were kept. This resulted in a large number of different motifs. In order to find the most discriminative ones, the 27 negative CRM (Fig. S2) were used as a negative set.

The positive and negative CRM sets were used to evaluate the False Negative Rate (FNR) and False Positive Rate (FPR) of the motifs generated by the algorithm. For each created motif, the two sets were scanned for conserved instances with a scanning threshold varied between 7 and 13 bits. For each threshold, FPR and FNR were computed as the proportion of Positive (resp. Negative) CRMs with at least one conserved instance for the motif with a score higher than the threshold. The best motifs, shown as red and blue dots, were selected based on the minimization of both FPR and FNR in a Pareto plot, as shown in Fig. 4B. These motifs were generated with a threshold of 10.1 bits and were scanned with optimal thresholds of 10.1 and 8.7 bits respectively.

Genome wide ranking of enhancers and genes.

In order to rank enhancers genome wide, we followed the method presented in [Rouault et al, 2010]. Coding sequences as well as the training set used for motif generation were masked . Conserved instances of *de novo* svbf7 and blue motifs at optimal threshold were then determined genome wide. Genomic fragment of 1Kbp were scored according to the additive Poisson score introduced in [Rouault et al., 2011] using the negative enhancers as a background set of intergenic fragment. Around each of the determined motif instances, the optimal scoring 1Kbp genomic fragment was defined as a putative enhancer. Each putative enhancer was associated to the nearest gene transcription start site. Each gene was attributed the highest score among its associated enhancers, or 0 if it had no associated enhancers. In order to assess the predictive power of different motif combinations as shown in Fig. 5B, the first 100 genes from the svb affymetrix data were used as True Positives, and a set of 96 negative genes were chosen for False Positive evaluation. A ROC plot was built using results from svbf7 alone (red curve) and the svbf7 and blue motif combination (blue curve). The same procedure was applied to ovoQ6 alone and is shown as the green curve.

Microarrays procedure.

Biotinylated cRNA targets were prepared, starting from 200 ng of total RNA, using the MessageAmp™ Premier RNA Amplification Kit (Ambion CAT# AM1792), according to the manufacturer recommendations. Following fragmentation, 6.5 µg of cRNAs were hybridized for 16 hours at 45°C on GeneChip® Drosophila Genome 2.0 Array interrogating over 18,500 transcripts (Affymetrix, Santa Clara, CA). The chips

were washed and stained using the GeneChip® Fluidics Station 450 and scanned using the GeneChip® Scanner 3000 7G according to Affymetrix recommendations. Raw data (.CEL Intensity files) were extracted from the scanned images using the Affymetrix GeneChip® Command Console (AGCC) version 3.2. CEL files were further processed with Affymetrix Expression Console software version 1.1 to calculate probeset signal intensities using the statistics-based Affymetrix algorithms MAS-5.0 with default settings and global scaling as normalization method. The trimmed mean target intensity of each chip was arbitrarily set to 100.

Supplementary information to Fig1

List of all genes analyzed by in situ hybridization and their expression pattern:

<u>CG number</u>	<u>symbol</u>	<u>expression in epidermis</u>	<u>other expression</u>	<u>svb dependent</u>	<u>ectopic</u>
CG10175	CG10175	Ventral & Lateral & Dorsal	salivary glands	yes	yes
CG11175	CG11175	Ventral	ubiquitous epidermis	yes	yes
CG12063	mey	Ventral & Lateral & Dorsal	no	yes	yes
CG12814	CG12814	Ventral & Lateral & Dorsal	ubiquitous epidermis	yes	yes
CG13209	sha	Ventral & Lateral & Dorsal	no	yes	yes
CG13913	mwh	Ventral & Lateral & Dorsal	no	yes	yes
CG14395	CG14395	Ventral & Lateral & Dorsal	no	yes	yes
CG1499	nyo	Ventral & Lateral & Dorsal	no	yes	yes
CG15005	CG15005	Lateral	ubiquitous epidermis	yes	yes
CG15013	dyl	Ventral & Lateral & Dorsal	no	yes	yes
CG15022	CG15022	Lateral & Dorsal	no	yes	yes
CG1520	wsp	Ventral & Dorsal	no	yes	yes
CG15335	cyr	Lateral & Dorsal	no	yes	yes
CG15589	CG15589	Ventral & Lateral & Dorsal	trachea	yes	yes
CG16798	CG16798	Ventral & Lateral & Dorsal	no	yes	yes
CG17058	Peritrophin-A	Ventral	ubiquitous epidermis	yes	yes
CG17131	try	Ventral & Lateral & Dorsal	no	yes	yes
CG17905	CG17905	Ventral	no	yes	yes
CG18525	Spn5	Ventral	no	yes	yes
CG2560	crp11A	Ventral & Lateral & Dorsal	no	yes	yes
CG2666	kkv	Ventral & Lateral	ubiquitous epidermis	yes	yes
CG31022	PH4aEFB	Ventral & Lateral & Dorsal	fat body	yes	yes
CG31559	CG31559	Ventral & Lateral & Dorsal	no	yes	yes
CG31973	CG31973	Ventral & Lateral & Dorsal	ubiquitous epidermis	yes	yes
CG32139	Sox21b	Ventral	nervous system	yes	yes
CG32159	dsc73	Ventral & Lateral & Dorsal	no	yes	yes
CG32356	ImpE1	Ventral	trachea	yes	yes
CG32694	CG32694	Ventral & Lateral & Dorsal	no	yes	yes
CG32858	sn	Ventral & Lateral & Dorsal	nervous system	yes	yes
CG3757	y	Ventral	no	yes	yes
CG4376	Actn	Ventral	trachea	yes	yes
CG4702	CG4702	Ventral & Lateral & Dorsal	ubiquitous epidermis	yes	yes
CG5424	f	Ventral & Dorsal	no	yes	yes
CG5847	zye	Ventral	no	yes	yes
CG7356	Tg	Ventral	no	yes	yes
CG7802	neo	Ventral & Lateral & Dorsal	no	yes	yes
CG8303	CG8303	Ventral & Lateral & Dorsal	no	yes	yes
CG9369	m	Ventral & Lateral & Dorsal	no	yes	yes
CG9517	CG9517	Ventral	no	yes	yes
CG10076	spir	Ventral	nervous system	no	ND
CG10244	Cad96Ca	Ventral & Lateral	no	no	ND
CG10917	fj	Ventral & Lateral & Dorsal	no	no*	no
CG12008	kst	Ventral & Lateral & Dorsal	trachea	no*	no
CG13699	CG13699	Ventral & Lateral & Dorsal	ubiquitous epidermis	no	ND
CG14626	CG14626	Ventral	ubiquitous epidermis	no	ND
CG14643	CG14643	Ventral	ubiquitous epidermis	no	ND
CG14830	CG14830	Ventral & Lateral & Dorsal	no	no	no
CG15080	CG15080	Ventral & Lateral & Dorsal	ubiquitous epidermis	no	ND
CG15282	CG15282	Ventral & Lateral & Dorsal	no	no	ND

CG15598	<i>osi17</i>	Ventral & Lateral & Dorsal	ubiquitous epidermis	no	ND
CG16885	<i>CG16885</i>	Ventral	trachea	no	no
CG17786	<i>CG17786</i>	Ventral & Lateral & Dorsal	no	no	ND
CG1869	<i>Cht7</i>	Ventral & Lateral & Dorsal	trachea	no	ND
CG2467	<i>CG2467</i>	Ventral & Lateral & Dorsal	ubiquitous endoderm	no	ND
CG2767	<i>CG2767</i>	Ventral & Lateral & Dorsal	ubiquitous endoderm	no*	no
CG30092	<i>jbug</i>	Ventral & Lateral & Dorsal	trachea	no*	no
CG32137	<i>CG32137</i>	Ventral	no	no	ND
CG32239	<i>Gef64C</i>	Ventral & Lateral & Dorsal	no	no*	ND
CG3244	<i>CG3244</i>	Ventral & Lateral & Dorsal	no	no	ND
CG33196	<i>dp</i>	Ventral & Lateral & Dorsal	no	no*	no
CG3541	<i>pio</i>	Ventral & Lateral & Dorsal	no	no*	no
CG4761	<i>knrl</i>	Ventral & Lateral	trachea	no	ND
CG5058	<i>grh</i>	Ventral & Lateral & Dorsal	hémocytes	no	ND
CG5065	<i>CG5065</i>	Ventral & Lateral & Dorsal	ubiquitous epidermis	no	ND
CG5756	<i>CG5756</i>	Ventral & Lateral & Dorsal	ubiquitous epidermis	no	ND
CG6433	<i>qua</i>	Ventral & Lateral & Dorsal	no	no*	no
CG7100	<i>CadN</i>	Ventral & Lateral & Dorsal	nervous system	no*	no
CG7337	<i>CG7337</i>	Ventral & Lateral & Dorsal	ubiquitous epidermis	no	ND
CG9196	<i>CG9196</i>	Ventral & Lateral & Dorsal	no	no*	no
CG9889	<i>yellow-d</i>	Lateral & Dorsal	no	no*	no
CG9990	<i>CG9990</i>	Ventral & Lateral & Dorsal	trachea	no*	no
CG8170	<i>CG8170</i>	Ventral & Lateral & Dorsal	ubiquitous epidermis	no*	no
CG17045	<i>ye -e3</i>	Lateral & Dorsal	no	no	ND
CG17914	<i>CG17914</i>	Ventral	gut	no*	no
CG31558	<i>CG31558</i>	Ventral & Lateral & Dorsal	ubiquitous epidermis	no*	no

* : it was sometimes difficult to conclude whether there was or not a significant decrease in svb mutants; in this case, a lack of responsiveness to svb ectopic expression was analyzed
ectopic svb expression was performed as described in Chanut et al., 2006 by
ectopically expressing svb in naked cells (wgGAL4 > uasSvb^{ACT})

ND: not determined

expression pattern are listed from BDGP and this
study

Supplementary Information to Figure 2 and 3

CRM name	related gene	genomic position (2006)				activity
EminB	miniature	chrX	11654097	11655097		positive
Emin	miniature	chrX	11650982	11651144		positive
155589	CG15589	chr3R	2027050	2028050		positive
17058	Peritrophin-A	chrX	20113835	20115016		positive
32159	dsx-c73A	chr3L	16435341	16436341		positive
CyrA	cypher	chrX	8019397	8020397		positive
dyl2	dusky like	chr3L	4298268	4299468		positive
sha1	shavenoid	chr2R	7209659	7210257		positive
sha3	shavenoid	chr2R	7215630	7216294		positive
tyn2	tryunity	chrX	77484	78384		positive
4702B	CG4702	chr3R	7951250	7952250		positive
31559	CG31559	chr3R	1977200	1978200		positive
nyo1	nyobe	chr3R	27384231	27384921		positive
snE1	singed	chrX	7869678	7870390		positive
32356	ImpE1	chr3L	8370153	8371153		negative
nyo2	nyobe	chr3R	27381479	27382574		negative
sox21b	sox21b	chr3L	14121641	14122852		negative
f4	forked	chrX	17162096	17163096		negative
f5	forked	chrX	17158996	17160096		negative
4702	CG4702	chr3R	7950428	7950586		negative
1499-1	nyobe	chr3R	27370295	27370415		negative
12063	morpheyus	chr3R	27320842	27321073		negative
15013-1	dusky like	chr3L	4299768	4299927		negative
15013-2	dusky like	chr3L	4300420	4300715		negative
CyrB	cypher	chrX	8017597	8018797		negative
dyl1	dusky like	chr3L	4303968	4304768		negative
dyl3	dusky like	chr3L	4305468	4306468		negative
f1	forked	chrX	17153478	17154756		negative
f2	forked	chrX	17159378	17160246		negative
neyo	neyo	chr3R	25647300	25648300		negative
tyn1	tryunity	chrX	86343	87613		negative
snE5	singed	chrX	7871996	7872659		negative
sn-enh1	singed	chrX	7864407	7869657		negative
snH5	singed	chrX	7868528	7868978		negative
snB2	singed	chrX	7869989	7870390		negative
nyo3	nyobe	chr3R	27377275	27378274		negative
sha-int	shavenoid	chr2R	7216771	7220065		negative
snP	singed	chrX	7862910	7864103		negative
snE4	singed	chrX	7871432	7872096		negative

Supplementary information to Fig6

A: Genes used as control from microarray experiments: these genes show no variation of their expression in svb and pri mutants.

Using microarrays analysis, genes with significant level of expression in wild type condition were used. Among these 5000 genes, 100 of them with a $0.8 \leq \text{pvalue} \leq 1$ associated to *svb* and *pri* mutant are selected. Gray line indicates a gene that is validated as a true *Svb* non target gene by *in situ* hybridization. For each gene, expression pattern, predictions for *svbF7* and blue motifs, and ChIP peak in a 5kb window are mentioned.

name	CG Number	expression pattern	svb*	pri*	chip in 5kb
robo	CG13521	ND	91,22	73,45	no
CG8090	CG8090	ND	85,52	115,10	yes
blow	CG1363	not epidermal	101,79	104,82	no
ush	CG2762	ep stripes	100,46	98,55	yes
RpS30	CG15697	ND	115,96	77,01	no
Bap55	CG6546	not epidermal	108,19	109,38	yes
lack	CG4943	ND	109,29	111,22	no
ttk	CG1856	ep stripes	97,50	80,48	yes
CG12164	CG12164	ND	83,20	93,45	no
CG3305	CG3305	ND	88,29	79,96	yes
CG11877	CG11877	not epidermal	113,94	118,58	no
CG14229	CG14229	not epidermal	105,43	109,87	no
Rga	CG2161	not epidermal	93,53	108,09	yes
Rpb5	CG11979	not epidermal	81,38	91,59	no
bsf	CG10302	ND	98,81	94,84	no
hkl	CG10473	not epidermal	97,76	103,84	yes
CG15099	CG15099	ND	88,57	94,68	yes
gry	CG17569	ND	97,01	89,65	no
slik	CG4527	ND	90,39	88,72	yes
trr	CG3848	ND	80,15	106,49	yes
CG14442	CG14442	ND	111,93	120,36	no
CG12375	CG12375	not epidermal	103,42	122,64	no
RpS27	CG10423	not epidermal	106,16	76,13	no
cenG1A	CG31811	not epidermal	108,36	102,03	no
Rgl	CG8865	not epidermal	137,42	100,20	yes
MBD-R2	CG10042	not epidermal	99,57	106,58	no
CG6230	CG6230	ND	95,47	123,11	no
kis	CG3696	ND	112,50	96,17	yes
CG4210	CG4210	ND	91,21	118,02	yes
betaTub56D	CG9277	ep stripes	105,75	105,22	no
RhoGAP1A	CG40494	ND	73,46	113,54	yes
CG9293	CG9293	ND	99,22	92,58	yes
CG10731	CG10731	ND	97,43	96,07	yes
CG2249	CG2249	ND	97,42	83,94	no
CG18549	CG18549	not epidermal	98,92	90,20	yes
CG9776	CG9776	ND	107,81	92,47	no
Bre1	CG10542	ND	108,66	134,50	no
vsg	CG16707	not epidermal	113,76	107,51	yes
CBP	CG1435	not epidermal	102,78	91,76	yes

Ufd1-like	CG6233	not epidermal	112,22	110,15	no
CG8928	CG8928	ND	91,78	113,34	no
Ald	CG7643	not epidermal	84,52	91,17	no
Atg18	CG7986	not epidermal	100,61	98,82	no
Chd1	CG3733	ND	97,28	94,33	no
mRpS11	CG5184	not epidermal	114,70	108,13	no
CG6852	CG6852	ND	111,03	109,18	yes
Krn	CG32179	not epidermal	102,68	108,51	yes
CG7852	CG7852	ND	106,54	113,74	no
CG8931	CG8931	not epidermal	110,33	111,09	yes
mRpL33	CG3712	ND	105,01	112,98	yes
RanBPM	CG42236	ND	124,37	103,44	no
CG8289	CG8289	not epidermal	101,82	109,37	no
CG5931	CG5931	not epidermal	123,35	92,97	no
CG32267	CG32267	ND	95,97	115,07	no
CG32164	CG32164	not epidermal	98,31	125,93	no
CG10365	CG10366	not epidermal	100,39	101,34	yes
kst	CG12008	ep stripes	127,31	103,42	yes
mRpS11	CG5184	not epidermal	114,70	108,13	no
CG2249	CG2249	ND	97,42	83,94	no
Ssdp	CG7187	not epidermal	111,89	87,21	no
CG17082	CG17082	ND	91,18	94,01	yes
CG4841	CG4841	ND	110,54	95,91	no
Stlk	CG40293	not epidermal	105,66	87,23	no
CG9715	CG9715	ND	119,34	97,89	yes
Rap2I	CG3204	ND	108,16	107,35	yes
drl	CG17348	ep stripes	115,07	78,43	yes
CG8878	CG8878	not epidermal	103,04	96,89	yes
Trim9	CG31721	not epidermal	96,29	91,91	yes
CG9917	CG9917	not epidermal	126,61	102,54	no
CG12006	CG12006	not epidermal	114,23	113,87	yes
crp	CG7664	not epidermal	82,88	111,58	yes
Cad87A	CG6977	ND	109,50	112,09	no
CG7028	CG7028	not epidermal	101,44	109,71	yes
zormin	CG33484	ND	130,13	114,51	yes
bbx /// waw	CG1414	ND	101,98	109,42	yes
CG1965	CG1965	ND	90,48	108,38	yes
CG1371	CG1371	not epidermal	115,46	100,11	yes
RpS30	CG15697	not epidermal	115,96	77,01	no
CG32676	CG32676	ND	90,95	93,01	yes
msn	CG16973	ND	114,08	115,00	yes
CG31108	CG31108	not epidermal	103,94	105,91	no
CG2918	CG2918	epidermal ubiquitous	103,01	88,58	no
mmy	CG9535	not epidermal	95,52	97,92	no
Nat1	CG3845	not epidermal	108,49	82,47	no
Taf1	CG17603	not epidermal	107,83	109,83	no
CG5869	CG5869	ND	89,39	108,39	yes
CG13284	CG13284	not epidermal	100,89	92,66	yes
CG12404	CG12404	ND	104,22	90,36	yes
tra2	CG10128	not epidermal	95,99	79,72	yes
CG6406	cg6406	not epidermal	90,56	87,00	yes
Fip1	CG1078	ND	104,96	87,54	no
Ubp64E	CG5486	not epidermal	130,11	101,59	no
CG14636	CG14636	ep stripes	112,31	88,23	no
gp210	CG7897	ND	110,61	106,95	no
dbr	CG11371	ND	139,91	126,84	no
bru-2	CG43065	ND	118,64	112,13	no

CG3493	CG3493	ND	126,41	109,19	yes
betaggt-I	CG3469	ND	110,78	130,95	yes
gek	CG4012	ND	110,94	96,43	no
bip2	CG2009	not epidermal	146,43	130,23	yes

* : expression in mutant / WT (%)

Svb non-regulated genes
analysed par Hybridization in situ