PLoS Biology

Decoding the transcriptional program of epidermal cell morphogenesis

--Manuscript Draft--

Manuscript Number:	
Full Title:	Decoding the transcriptional program of epidermal cell morphogenesis
Short Title:	transcriptional control of epidermal cell
Article Type:	Research Article
Keywords:	Drosophila epidermal cell differentiation morphogenesis cell shape remodeling Ovo/Shavenbaby transcription factor regulatory gene network Cis Regulatory Modules enhancer tissue specific transcription ChIP transcriptome Shavenbaby downstream ge
Corresponding Author:	Serge Plaza Centre de Biologie du Developpement toulouse cedex 9, FRANCE
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Centre de Biologie du Developpement
Corresponding Author's Secondary Institution:	
First Author:	delphine Menoret
First Author Secondary Information:	
Order of Authors:	delphine Menoret
	Marc Santolini
	Isabelle Fernandes
	Rebecca Spokony
	Jennifer Zanet
	Ignacio Gonzalez
	Yvan Latapie
	Pierre Ferrer
	Hervé Rouault
	Kevin P White
	Philippe Besse
	Vincent Hakim
	Stein Aerts
	Francois Payre
	Serge Plaza

Order of Authors Secondary Information:	
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.
Suggested Reviewers:	
Opposed Reviewers:	Richart Mann, PhD PI, Columbia University rsm10@columbia.edu Specialist of transcriptional cell differentiation program triggered by Hox factors in Drosophila
	Michael Levine, PhD PI, University of California, Berkeley mlevine@berkeley.edu One of the best lab in the word addressing the questions of Transcriptional program during development
	Patrick Lemaire, PhD PI, CRBM, CNRS Patrick.lemaire@crbm.cnrs.fr His lab is deciphering the Transcriptional control of chordate morphogenesis
	Claude Desplan, PhD PI, Professor, NYU cd38@nyu.edu Strong experience of developmental mechanisms during development and adressing the questions of gene specific expression and functions.
	Scott Barolo, PhD PI, University of Michigan Medical School sbarolo@umich.edu working on CRM architecture and transcription factors in drosophile. Goog knowledge in the field.

Cover Letter

CBD
Centre de Biologie du Développement
www-cbd.ups-tlse.fr
CNRS/Université Paul Sabatier
118 route de Narbonne - Bât. 4R3 b3
31062 Toulouse cedex 9, France
T. 05.61.55.67.37
F. 05 61 55 65 07
UMR 5547

sous la co-tutelle de
cnrs
Université de Toulouse

Dear *PLoS Biology* Editor,

We are extremely pleased to send you the manuscript "Decoding the transcriptional program of epidermal cell morphogenesis" for consideration for publication as a Research article in *PLoS Biology***.**

While the general concepts of transcriptional regulation are well established, a current challenge resides in understanding the logic and physical elements implementing the interactions between Transcription Factors and their target genes within developing embryos (Rister and Desplan 2010). A central problem to be addressed is how enhancers, or Cis Regulatory Modules (CRM), are built and function to drive specific expression of a coherent set of genes, directly responsible for terminal differentiation.

Epidermal development in the fly embryo is an ideal model system for studying the genetics underlying morphological differentiation. We have shown previously that the Shavenbaby (Svb) transcription factor determines which (Frankel et al, Nature 2011) and when (Kondo et al, Science 2010) epidermal cells differentiate trichomes, through targeting the expression of terminal effectors of cell shape remodeling (Chanut et al., PLoS Biol 2006; Fernandes et al, Dev Cell 2010). In this paper, we report an interdisciplinary effort to identify CRM mediating Svb-dependent expression of effectors in trichome cells. This work is more than simply examine the expression pattern driven by these CRMs; we combine computational modeling to a fine dissection and mutagenesis of 14 CRMs driving trichome formation. This methodology allowed us to define, at an unprecedented level of detail, the functional architecture of CRMs mediating a transcription program of terminal differentiation *in vivo*.

We observed that the function of effector CRM relies on very weak if any clustering of Svb binding sites, with only 1-3 sites within each CRM. Subtle constraints in adjacent nucleotides -not previously detected using *in vitro* approaches- help defining which, among thousands of similar motifs, are bound *in vivo* by Svb and mediate regulatory interactions. Furthermore, various combinations of additional cis-regulatory elements are required together with Svb binding sites to build a functional CRM. Our study is the first one to show how flexible architectures of different cis-regulatory elements can generate similar expression patterns coordinating a battery of target effector genes.

Our results are in sharp contrast with recent results from the analysis of early acting Gene regulatory networks. Collectively, the latter have shown the importance of homotypic clusters of TF binding sites (Segal et *al*., Nature 2008) and postulate the existence of a specific combination of cis-regulatory elements to generate a given transcription pattern (Zinzen et *al*., Nature 2009). Strikingly, our study of trichome cells and studies of segmentation or mesodermal networks revealed very different cis-regulatory codes. Our results thus suggest that the rules drawn from previous works may be restricted to enhancers acting at early stages to specify cell fates and instead those mediating terminal differentiation can display unexpected plasticity in their architecture. Obviously, more examples are required to evaluate whether this is a general feature of terminal differentiation programs (Rister and Desplan, 2010) and we believe that our approach and results will encourage future studies of this important problem in genome biology.

Thank you for your attention to this manuscript

With best wishes

Serge Plaza, PhD, CNRS & University of Toulouse

Decoding the transcriptional program of epidermal cell morphogenesis

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

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

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

[3] Laboratoire de Physique Statistique, CNRS, Université Pierre & Marie Curie, 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 cisTargetX 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 cisTargetX 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 (**C**n**GTT**a), with

a similar core consensus but extended to adjacent nucleotides (A**C**H**GTT**AK). 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 it 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 epifluoresence 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 transciption 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 $\Sigma_i (O_i - E)^2 / E$, 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.

<u>Figure 6:</u> 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, Erezyilmaz 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) vrille 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, Celniker 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

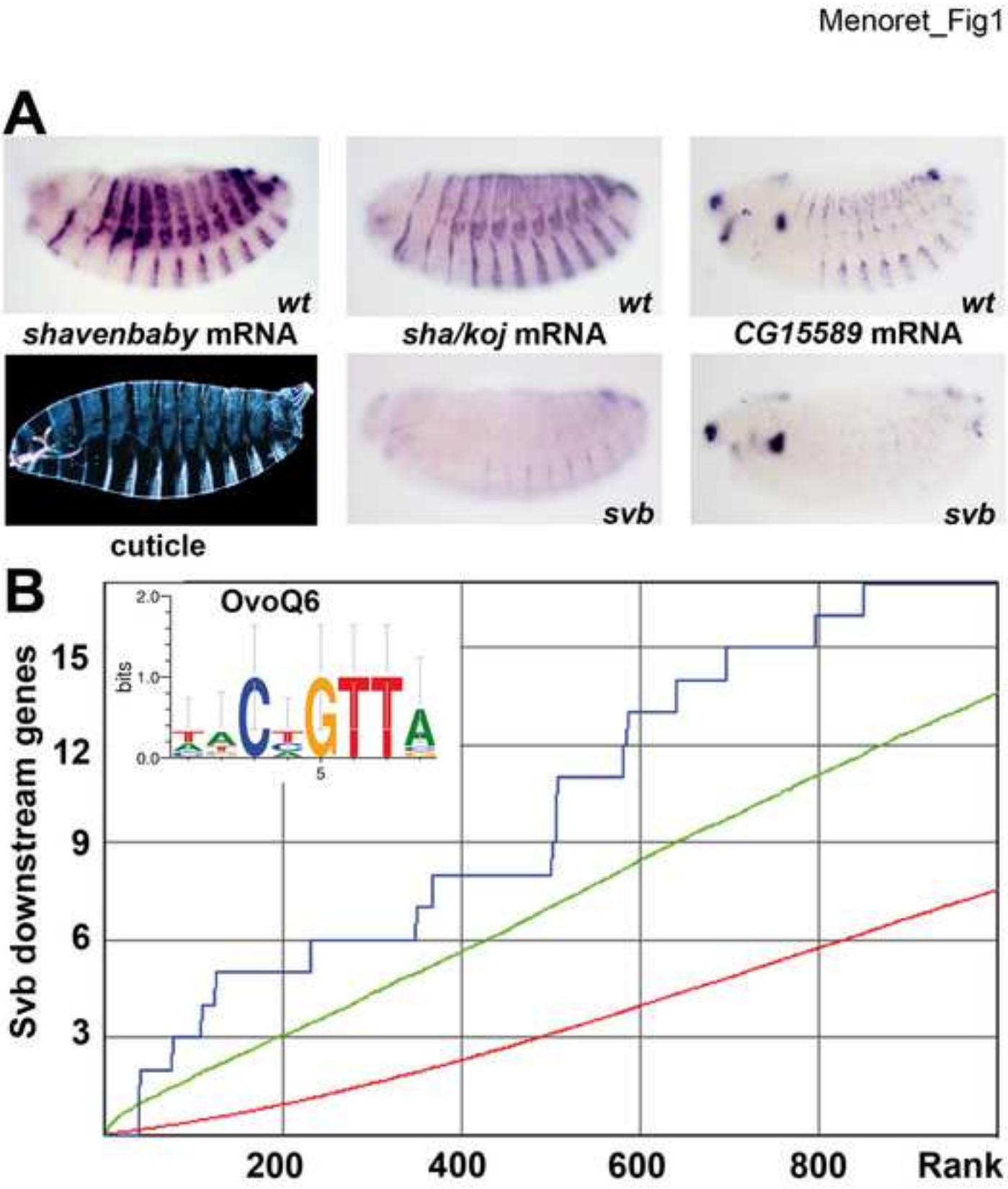


Figure2
Click here to download high resolution image

Menoret_Fig2

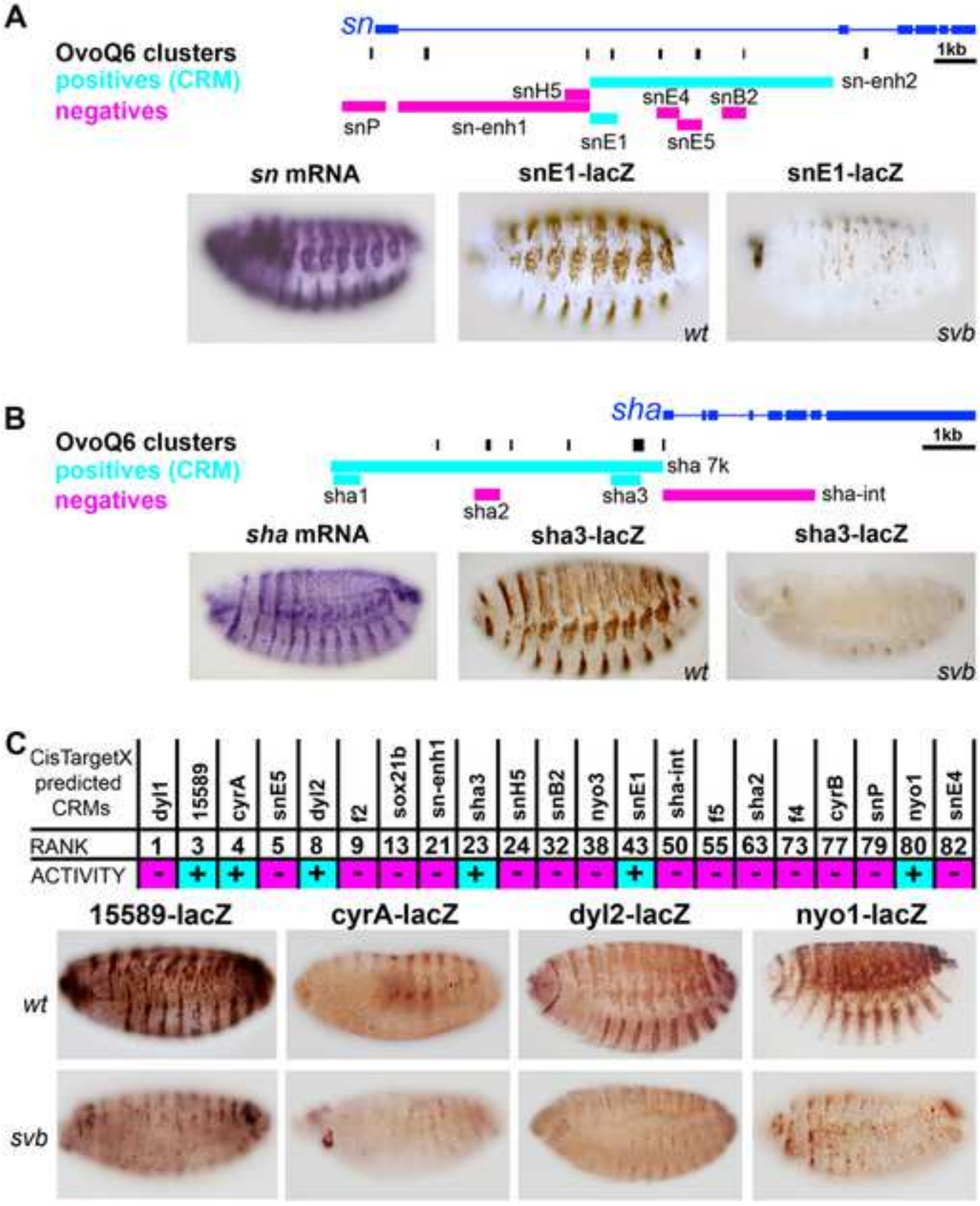


Figure3
Click here to download high resolution image

Menoret_Fig3

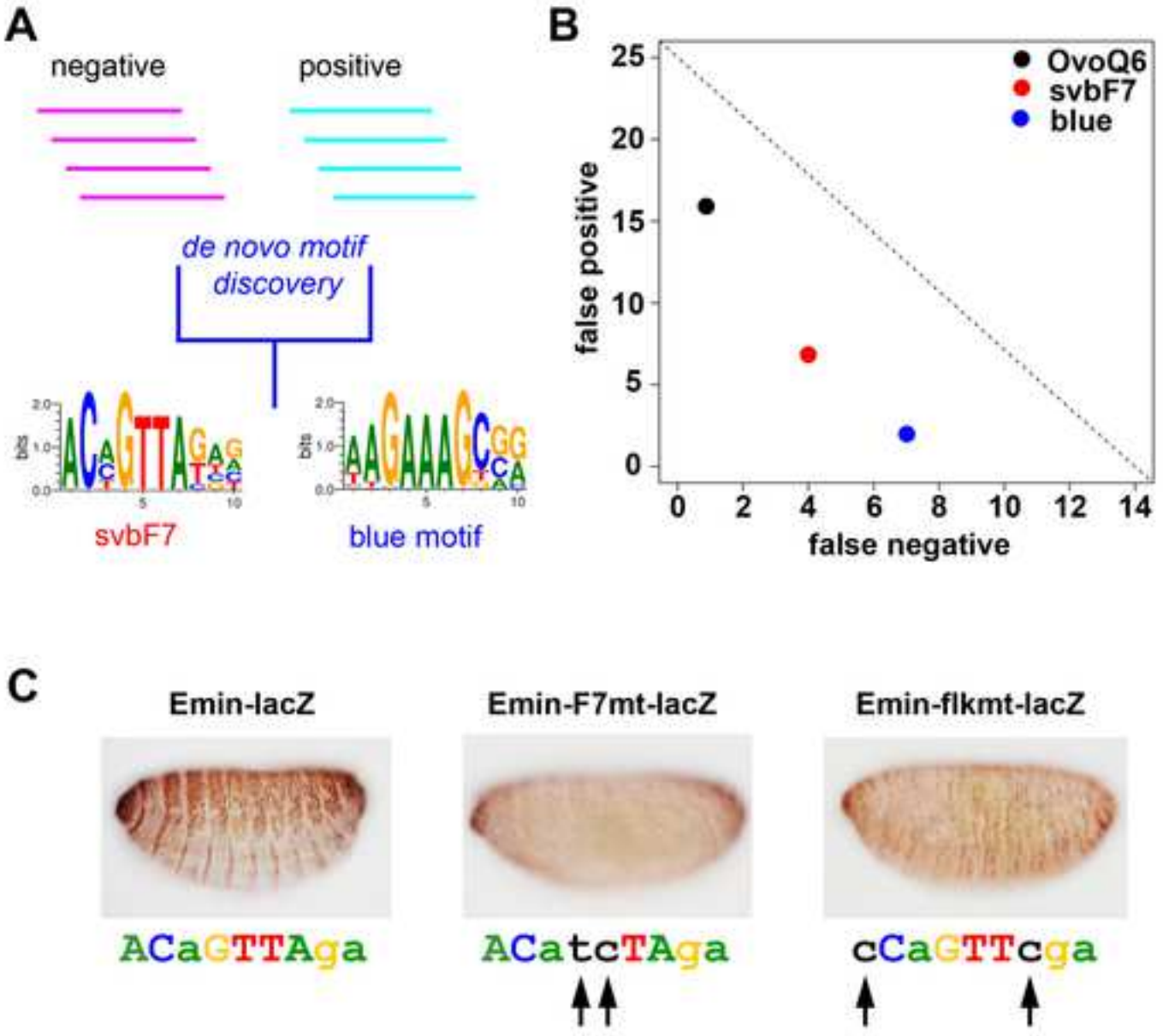


Figure4
Click here to download high resolution image

Menoret _Fig4

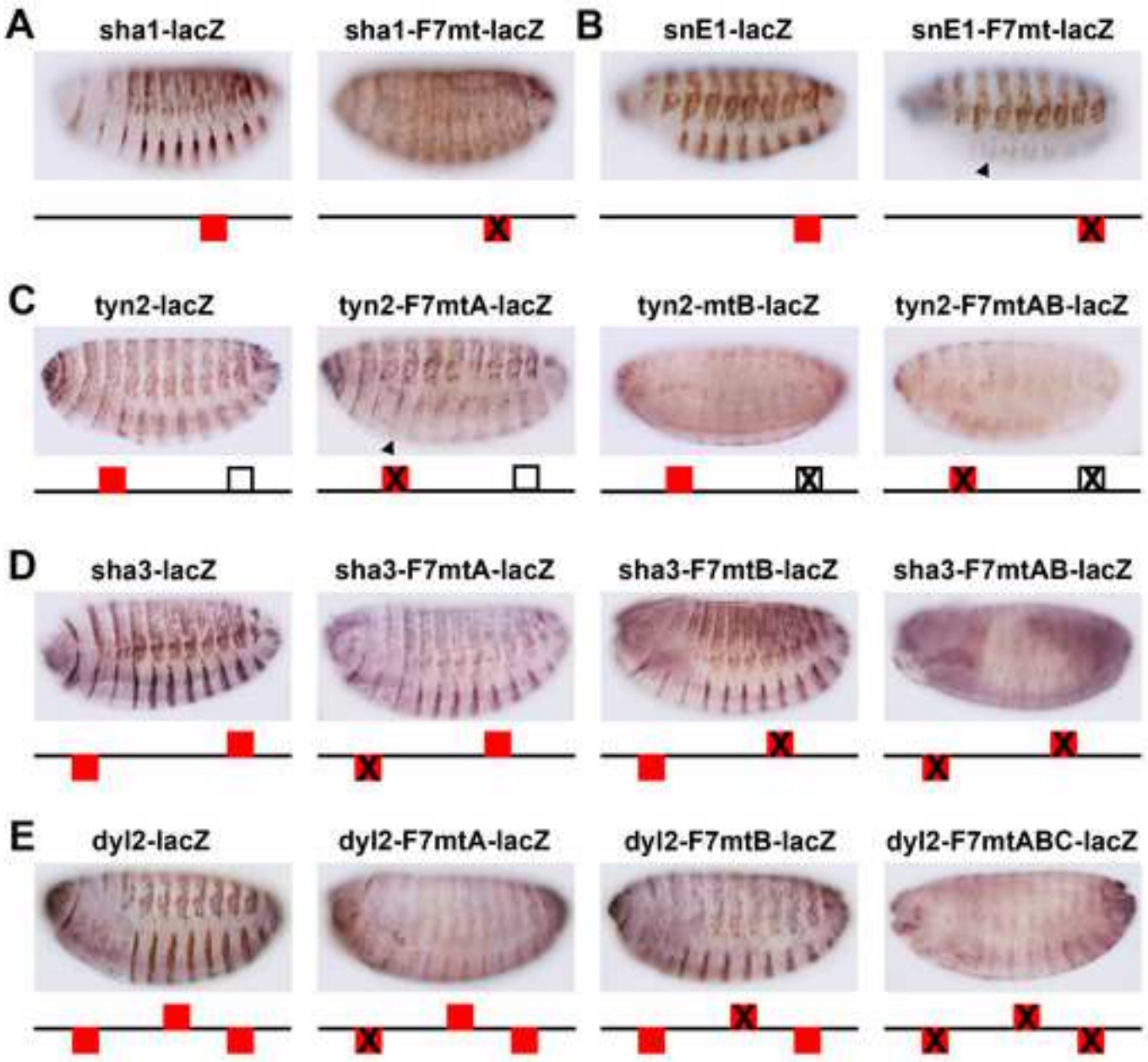


Figure5
Click here to download high resolution image

Menoret_Fig5

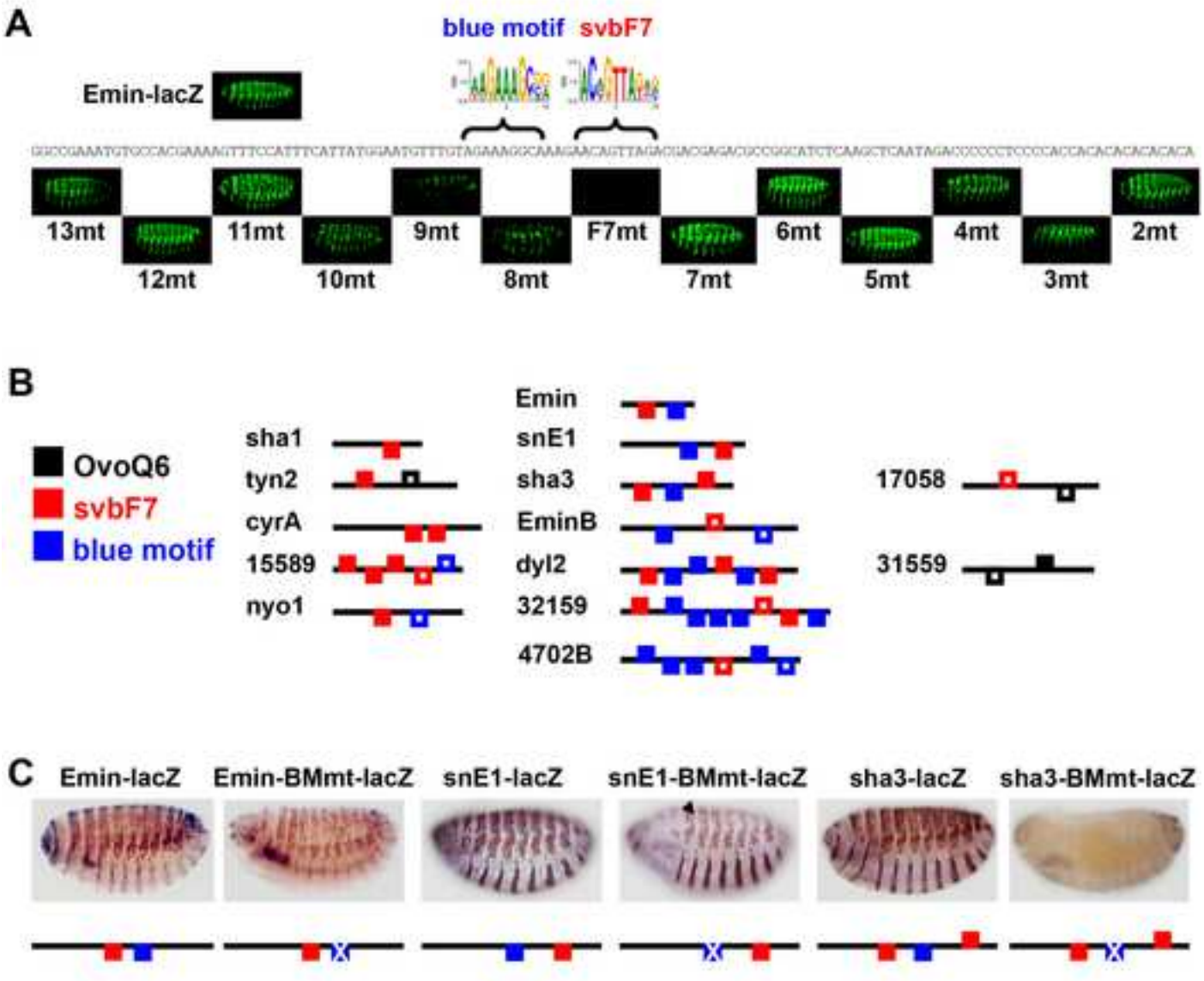


Figure6
Click here to download high resolution image

Menoret_Fig6

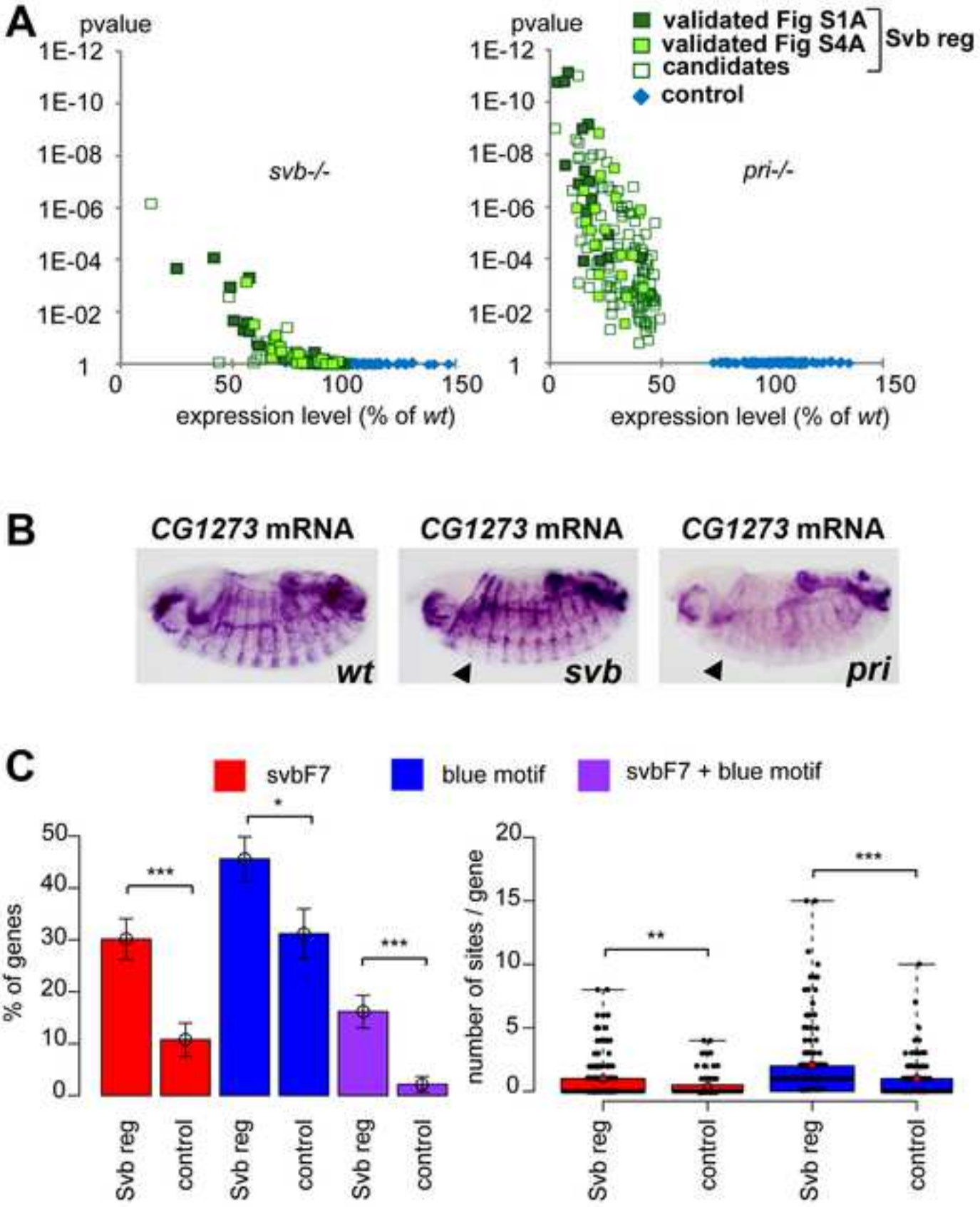
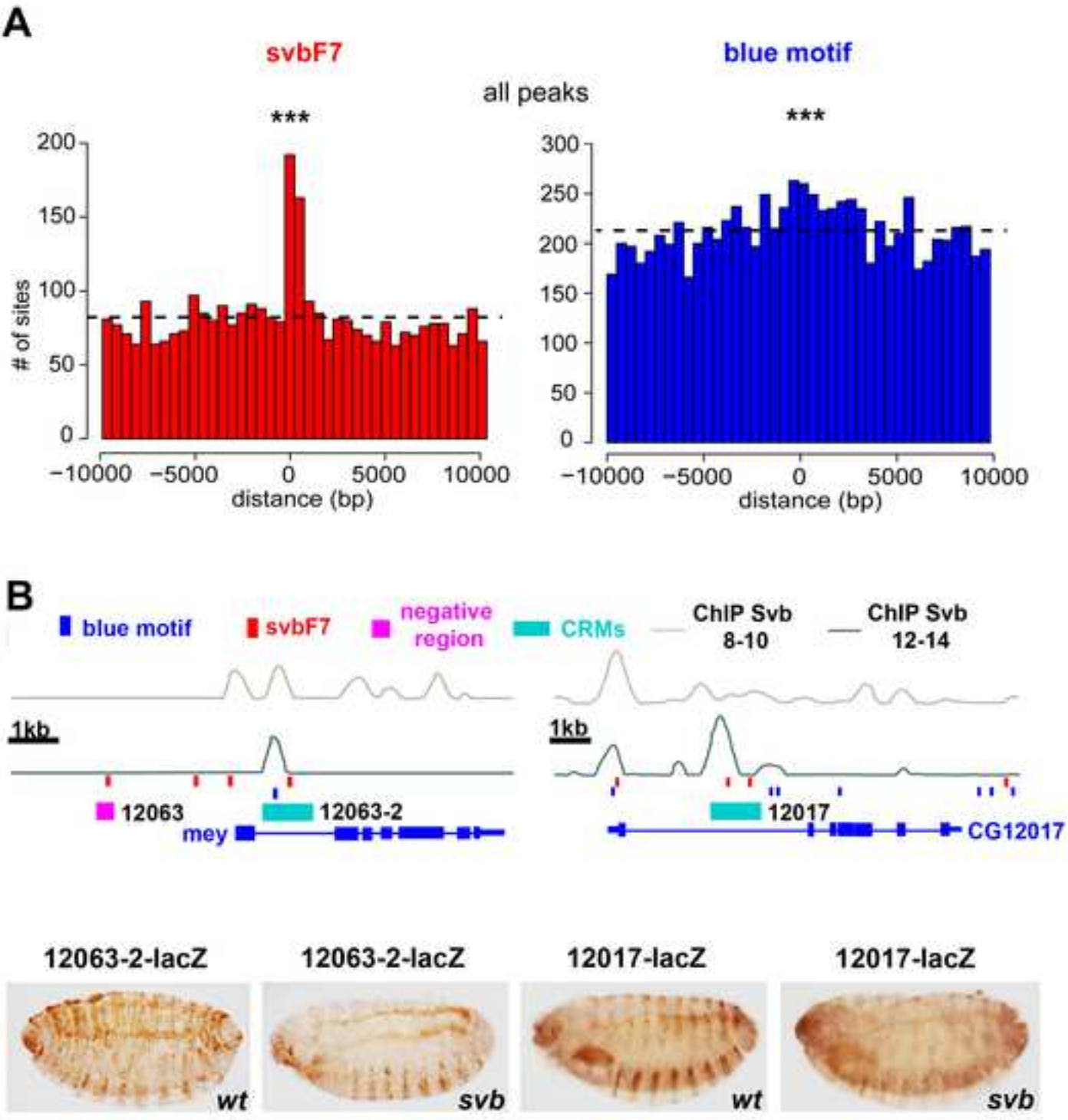


Figure7
Click here to download high resolution image



Supporting Information

Click here to download Supporting Information: Menoret et al_SupMaterial.pdf