LETTER

doi:10.1038/nature13395

# Genome-scale functional characterization of *Drosophila* developmental enhancers *in vivo*

Evgeny Z. Kvon[1]†, Tomas Kazmar[1], Gerald Stampfel[1]*, J. Omar Yáñez-Cuna[1]*, Michaela Pagani[1], Katharina Schernhuber[1], Barry J. Dickson[1]† & Alexander Stark[1]

**Transcriptional enhancers are crucial regulators of gene expression and animal development[1] and the characterization of their genomic organization, spatiotemporal activities and sequence properties is a key goal in modern biology[2–8]. Here we characterize the *in vivo* activity of 7,705 *Drosophila melanogaster* enhancer candidates covering 13.5% of the non-coding non-repetitive genome throughout embryogenesis. 3,557 (46%) candidates are active, suggesting a high density with 50,000 to 100,000 developmental enhancers genome-wide. The vast majority of enhancers display specific spatial patterns that are highly dynamic during development. Most appear to regulate their neighbouring genes, suggesting that the *cis*-regulatory genome is organized locally into domains, which are supported by chromosomal domains, insulator binding and genome evolution. However, 12 to 21 per cent of enhancers appear to skip non-expressed neighbours and regulate a more distal gene. Finally, we computationally identify *cis*-regulatory motifs that are predictive and required for enhancer activity, as we validate experimentally. This work provides global insights into the organization of an animal regulatory genome and the make-up of enhancer sequences and confirms and generalizes principles from previous studies[1,9]. All enhancer patterns are annotated manually with a controlled vocabulary and all results are available through a web interface (http://enhancers.starklab.org), including the raw images of all microscopy slides for manual inspection at arbitrary zoom levels.**

Animal development depends on differential gene expression governed by genomic regulatory elements called enhancers[1,9], which are being studied extensively[2,3,8,10,11]. Many of the basic principles of developmental gene regulation have been elucidated in the fruitfly *Drosophila melanogaster*[1,12,13], and work over the past decades has characterized gene expression, transcription factor binding, chromatin features and enhancer activity in *Drosophila* at unprecedented levels[2,5–8,14–16]. This and the ability to obtain many embryos from all developmental stages[17] make *Drosophila* an ideal model in which to characterize spatiotemporal enhancer activities at a genomic scale and throughout embryogenesis.

To systematically characterize developmental enhancers in the *D. melanogaster* genome, we made use of transgenic fly lines (Vienna Tiles (VT) library), publicly available from the Vienna *Drosophila* RNAi Center (VDRC). Each line contains a transcriptional reporter construct with a ~2 kilobase (kb) genomic DNA fragment (enhancer candidate), minimal promoter and *GAL4* reporter gene integrated into an identical position in the fly genome[16], thus allowing the direct comparison of the candidates' activities (Fig. 1a, Extended Data Fig. 1a, b and Supplementary Table 1). Together, these fragments cover about 14 million base pairs or 13.5% of the non-coding, non-repetitive genome, with little or no bias regarding the distance to transcription start sites (TSSs; Extended Data Fig. 1c) or the embryonic expression of neighbouring genes (Extended Data Fig. 1d).

We developed a high-throughput pipeline to assess transcriptional enhancer activities in fly embryos by *in situ* hybridization against the *GAL4* reporter transcript. For each transgenic line, we acquired whole-slide images, each with about 400 embryos covering all stages of embryogenesis, and manually annotated the enhancer activity patterns using a controlled vocabulary[14] at six time intervals of embryogenesis (Extended Data Fig. 1e). The pipeline reported activities independent of fragment delineation and orientation and recovered 27 out of 28 known enhancers, whereas 13 out of 13 non-*Drosophila* controls were inactive (Extended Data Fig. 2a–c and Supplementary Information section 1). Results from re-testing 34 negative and 78 positive fragments using a different genomic site (on chromosome 2L instead of 3L) and reporter gene (*lexA*) were highly similar and the majority (82%) of enhancer activity patterns matched to the expression patterns of neighbouring genes, suggesting that we predominantly measured endogenous enhancer activities (Extended Data Fig. 2d–f and Supplementary Information section 1).

3,557 of all 7,705 tested candidate fragments (46%) were active in the embryo with diverse patterns that included gap and pair-rule patterns, all primary germ layers (Extended Data Figs 3a and 4a), and all major cell types and tissues (Fig. 1b and Extended Data Figs 3b and 4b). The fraction of active fragments increased about fivefold from ~7% in early embryos (stages 4–6) to ~35% for stages 15–16, consistent with the increase in organism complexity and the number of cell types (Fig. 1c). By contrast, the number of expressed genes remains roughly constant during embryogenesis (~1.3-fold increase[18]). Enhancer activities were much sparser than gene expression patterns both temporally and spatially: while 94% of all enhancers were only transiently active and only 0.8% were ubiquitous during the entire embryogenesis, this was true for 56.7% and 20.5% of the genes, respectively (Fig. 1d, e and Extended Data Fig. 5a–c). The temporal dynamics of enhancer activity was also apparent from changes of enhancer-associated chromatin features such as DNase I hypersensitivity (DHS), binding of co-activator CBP/p300, and presence of histone H3K27 acetylation mark assessed in entire embryos or in a tissue-specific manner[2,19,20] (Fig. 1f–h, Extended Data Fig. 6 and Supplementary Information section 2). Together, this confirms and quantifies the transient and dynamic nature of enhancer function and suggests that development progresses through increasingly complex gene regulation by enhancers with temporally and spatially restricted activities.

We next identified domains in the blastoderm embryo in which enhancers appeared co-regulated (that is, were coordinately active or inactive). Automated image segmentation and reverse clustering revealed distinct regions corresponding to the presumptive anterior and posterior endoderm, head and trunk mesoderm, procephalic neuroectoderm, and others, overall strongly resembling the established fate map of the blastoderm embryo[17] (Fig. 1i and Extended Data Fig. 4c). This suggests that cells within these domains have a common developmental fate, presumably due to shared *trans*-regulatory environments. Indeed, during late stages, early mesodermal enhancers were preferentially active in mesoderm derivatives (somatic, visceral and cardiac muscles), whereas early endodermal enhancers were active in endoderm derivatives (midgut and Malpighian tubules) (Extended Data Fig. 4d). These and equivalent trends for other presumptive tissues of the early embryo (Extended Data Fig. 4e–g)

[1]Research Institute of Molecular Pathology (IMP), Vienna Biocenter VBC, Dr Bohr-Gasse 7, 1030 Vienna, Austria. †Present addresses: Howard Hughes Medical Institute, Janelia Farm Research Campus, Ashburn, Virginia 20147, USA (B.J.D.); Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA (E.Z.K.).
*These authors contributed equally to this work.

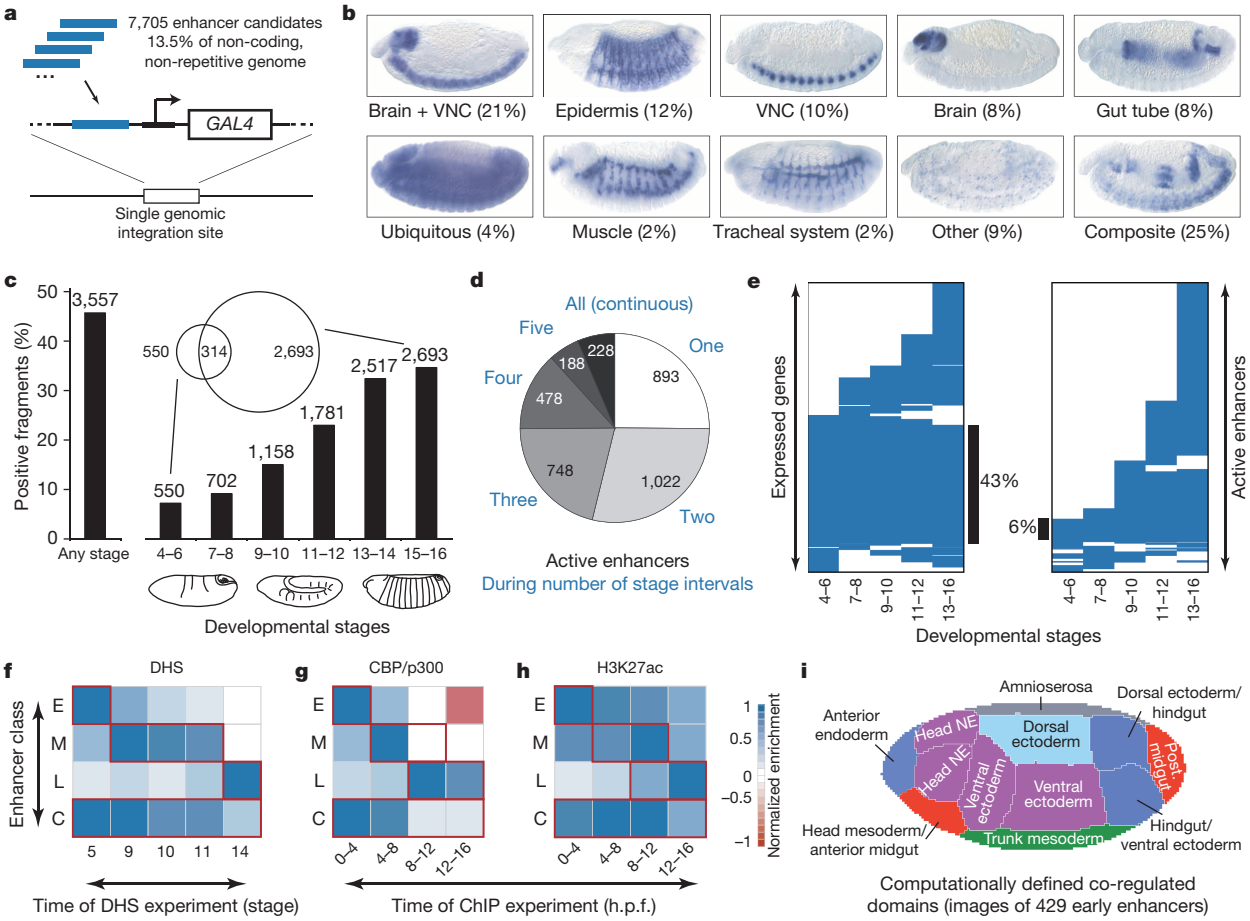



**Figure 1 | Enhancers display highly diverse and dynamic activity patterns across *Drosophila* development.** **a**, The VT library comprises transgenic flies with candidate fragments (blue) upstream of a transcriptional reporter (middle) in a constant genomic landing site (Extended Data Fig. 1). **b**, Proportion of enhancer activities in prominent tissues at stages 13–14 (representative embryos; Extended Data Figs 3 and 4). VNC, ventral nerve cord. **c**, The number of active enhancers increases during embryogenesis, with some overlap between early and late enhancers (Venn diagram in **c**). **d**, 3,329 (94%) embryonic enhancers are only transiently active. **e**, Temporal dynamics of gene expression (left, 5,134 genes[14]) and enhancer activity (right, 3,557 enhancers; black vertical lines indicate continuous expression or activity; Extended Data Fig. 5a). **f**–**h**, Heatmaps show the median enrichment of DNA accessibility[20] (**f**), CBP/p300 binding (**g**) and H3K27 acetylation (ac) marks[2] (**h**) on early (E), middle (M), late (L) and continuous (C) enhancers (rows) for experiments performed at different time points during *Drosophila* development (columns; red highlights coinciding time points; Extended Data Fig. 6). **i**, Co-regulated domains defined by reverse clustering of raw image data for 429 early enhancers resemble the embryo fate map[17] (Extended Data Fig. 4c). ChIP, chromatin immunoprecipitation; NE, neuroectoderm.

demonstrate that enhancer activities are consistent with the progression of development along defined cell lineages, highlighting the gene regulatory basis of development.

To analyse the locations of enhancers with respect to their putative target genes, we assigned enhancers to genes by manually matching enhancer activity and gene expression patterns (Figs 2a and 3a). For the 874 enhancers with the strongest activity patterns, we considered 3,681 genes within five genes up- and downstream of each enhancer (including host genes for intronic enhancers), that is, 9,293 enhancer–gene pairs. For 4,224 of these pairs (45%; 1,690 genes), expression patterns were available, resulting in 482 enhancer-to-gene assignments (of the enhancers for which all neighbouring genes were characterized, 82% could be assigned; Extended Data Fig. 2f and Supplementary Table 4). The assignments were supported by the location of chromosomal domain boundaries[21], binding sites of insulator proteins[22] and evolutionary chromosome breakpoints[23], all of which were depleted between enhancers and their assigned targets (Fig. 2b–d, Extended Data Fig. 7a–c and Supplementary Information section 3). Twenty-eight enhancers were assigned to and potentially regulate two genes, 23 of which were paralogues with very similar expression patterns (Supplementary Information section 4). During stages 4–6, 16 genes were assigned to enhancers with overlapping or identical activities reminiscent of shadow enhancers[24]. This is a considerable fraction (14%) among all 116 genes with multiple enhancers, in particular for developmental regulators (14 out of 16 genes are transcription factors; Supplementary Information section 5).

Along the linear genomic DNA sequence, the distances between the enhancers and the TSSs of their assigned target genes varied greatly: although many such pairs were close (21% were $<4$ kb), the median distance was 10 kb, and 28% of all inferred regulatory interactions were distal ($>20$ kb), up to more than 100 kb (Fig. 2e). However, considering the location of genes, the vast majority (88%) of all enhancers were located in the vicinity of their targets (Fig. 2f). Nevertheless, 12% of all enhancers were assigned across intervening genes and appeared to skip one (8%) or more (4%) genes to regulate a distal gene (Fig. 2f), as found for a *Sex combs reduced* (*Scr*) enhancer that lies beyond the *fushi tarazu* (*ftz*) gene[25]. Interestingly, enhancers were located almost as frequently upstream (30%) as downstream (22%) of their target genes (for example, the *SoxNeuro* (*SoxN*) locus; Extended Data Fig. 8), suggesting that no particularly preferred relative enhancer location might exist.

Thirty-six per cent of the enhancers were intragenic and appeared to predominantly (79%) regulate their host genes, as exemplified by *Thrombospondin* (*Tsp*; Fig. 3a). However, 21% were assigned to a neighbouring gene instead (Fig. 3b), including an enhancer located inside the intron of *bric a brac 1* (*bab1*) that appears to activate *bab2* over a distance of

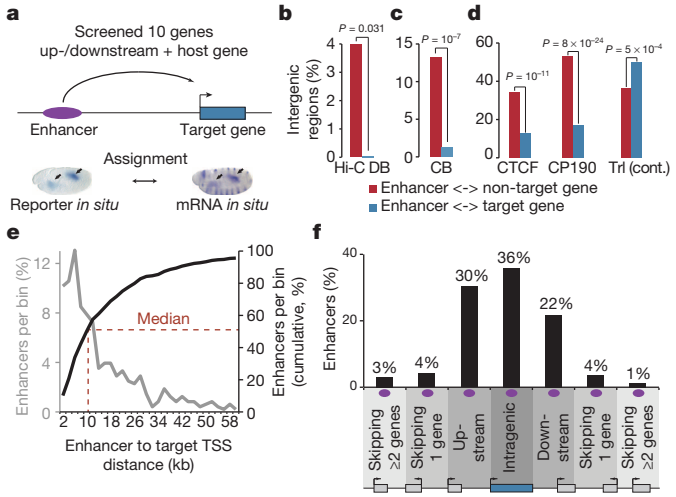



**Figure 2 | The organization of the *Drosophila cis*-regulatory genome.** **a**, Enhancer to target gene assignment based on enhancer activity and gene expression patterns. **b**–**d**, Chromosomal domain boundaries determined by Hi-C[21] (**b**), breakpoints during genome evolution[23] (**c**) and insulator binding sites[22] (**d**) show relative depletions between enhancers and their assigned target genes (blue) and enrichments between enhancers and non-targets (red), whereas the opposite is true for the activator Trl[22] (binomial *P* values are shown; see Extended Data Fig. 7a–c for additional insulator proteins and details). **e**, Genomic distances between enhancers and their assigned target gene TSSs in kb (grey, frequencies; black, cumulative). **f**, Frequency of enhancers (purple) at different genomic positions relative to their target genes (blue; schematic locus). Eighty-eight per cent of all enhancers are in the genes' genomic neighbourhoods within regulatory domains. CB, chromosomal breakpoints; DB, domain boundaries.

93 kb (Fig. 3c). *bab1* is not detectably expressed in the embryo during the corresponding developmental stage, which we found to be true more generally when intragenic enhancers regulated flanking genes rather than their host genes (Fig. 3d). Similarly, when intergenic enhancers were assigned to distal genes, the skipped genes were significantly less highly expressed than the target genes (Fig. 3d). Together these results support a predominantly local organization of the *Drosophila* genome into regulatory domains reminiscent of the chromosomal domains inferred from chromatin interactions[21].

The agreement of most enhancers' activities with the expression patterns of neighbouring genes (Figs 2f, 3a and Extended Data Figs 2f, 8) confirms that enhancer activity is predominantly context independent[9]. However, 18% of the enhancers could not be assigned to neighbouring genes (Extended Data Fig. 2f) and might be involved in more distal regulation (for example, ref. 26). For 19%, the activities were similar but broader and might thus be modulated in the endogenous sequence contexts in a more complex fashion (Supplementary Information section 1). Such context dependence is known for several loci in *Drosophila* (for example, the Hox locus[27]) and mouse (for example, *Fgf8* (ref. 28)), and enhancers in the bithorax complex indeed matched to gene expression patterns during early stages but appeared broader later (Extended Data Fig. 7d).

Many different enhancers showed similar or identical activity patterns in various embryonic tissues. For example, 263 were active throughout the central nervous system (CNS), 59 in midgut and 32 in macrophages (Extended Data Fig. 3), thus probably providing sufficient statistical power to discern predictive sequence signatures. Indeed, the motif content alone allowed the discrimination of enhancers from different functional classes using supervised machine learning in a cross-validated setting[29] (Extended Data Fig. 9a, b and Supplementary Table 5). The

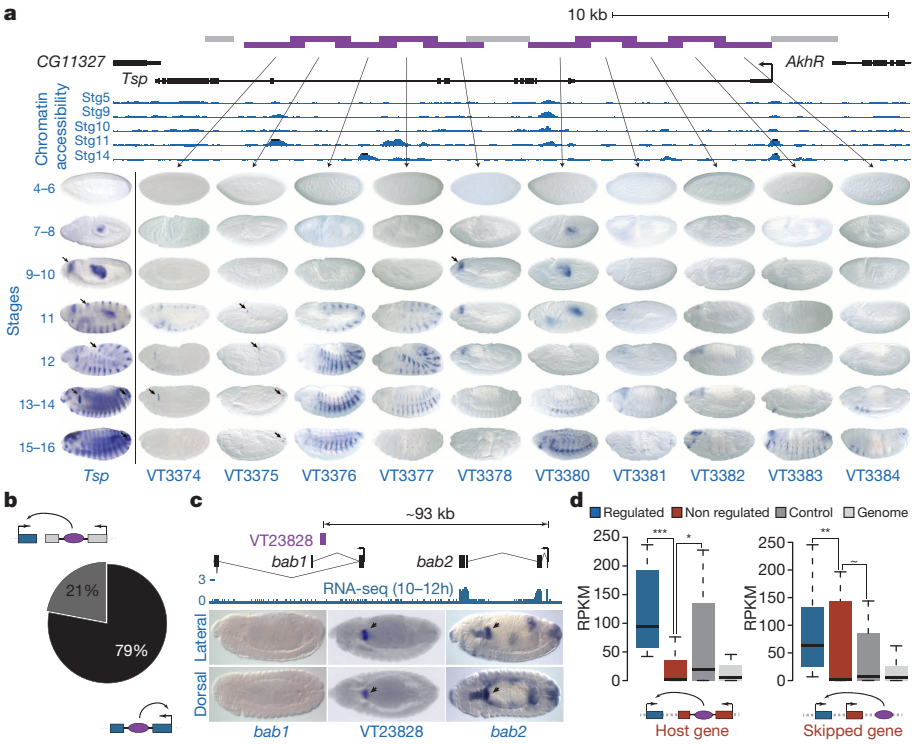



**Figure 3 | Intragenic enhancers in the *Drosophila* genome.** **a**, Enhancers in the *Tsp* locus. Top, UCSC Genome Browser screenshot including tested fragments (purple, positive; grey, negative) and DNA accessibility[20]. Bottom, embryos for all six time points of embryogenesis (left, *in situ* visualizing *Tsp* mRNA[14]; arrows highlight small expression/activity domains). **b**, Twenty-one per cent of intragenic enhancers are assigned to a neighbouring gene. **c**, A distal *bab2* enhancer (VT23828) in the intron of a neighbouring gene *bab1*. Top, UCSC Genome Browser screenshot including RNA-seq data for the corresponding stages[18]. Bottom, embryo images depicting the *bab1* and *bab2* expression during stages 13–14 (ref. 14) and VT23828's activity in the proventriculus (middle). **d**, Non-regulated host and skipped genes are often not expressed. Box plots show gene expression (reads per kilobase per million (RPKM)) values as measured by RNA-seq[18] for assigned target genes (blue) and non-regulated host genes (red, left) or skipped genes (red, right). Dark grey, unrelated neighbouring genes (control); light grey, all *D. melanogaster* genes. $***P = 10^{-8}$, $**P = 0.059$, $*P = 0.081$, $\sim P > 0.1$. Wilcoxon rank-sum test.

fraction of classes for which predictions were successful increased with the number of enhancers per class (Supplementary Table 5) but appeared to be independent of pattern complexity. This suggests that our understanding of regulatory sequences will benefit from the ongoing functional characterization of enhancers[4–7].

Different transcription factor motifs were strongly differentially distributed between the enhancer classes (Fig. 4a and Extended Data Fig. 9c). For example, early embryonic enhancers were enriched in motifs of the transcription factor Zelda, an important activator of embryonic gene expression[30]. Similarly, Twist (Twi) motifs were enriched in early mesodermal enhancers, Myocyte enhancing factor 2 (Mef2) motifs in late somatic muscle enhancers, and Pannier (Pnr) and Tinman (Tin) motifs in dorsal vessel enhancers, consistent with the established roles of the these transcription factors[8] (Fig. 4a and Extended Data Fig. 9c). To test whether predicted motifs are required for enhancer activity, we selected three midgut, four CNS and four anterior–posterior (A–P) enhancers (11 enhancers total), for which the successful predictions depended on GATA-like, Trithorax (Trl, also known as GAGA)-like, and Tramtrack (Ttk)-like motifs, respectively (Fig. 4a and Extended Data Fig. 9b, c). For each, we created reporter flies with an enhancer variant in which we disrupted the respective motifs by point mutations and compared the activity of the mutant and wild-type enhancers, both manually and by computational image analysis (Fig. 4b and Extended Data Fig. 10). In 10 out of 11 cases, the mutated enhancers were not active or had strongly reduced activity, validating the functional importance of the respective motifs.

Taken together this work complements efforts that study chromatin properties[2,19,20] or characterize enhancers at defined stages and in selected tissues[5–7,15,16]. Our results confirm and generalize principles and models from smaller scale studies (reviewed in refs 1, 9, 12) and suggest a high density of developmental enhancers in the *Drosophila* genome with an estimated total of ~41,000 enhancers or four enhancers per expressed protein-coding gene on average during embryogenesis alone. In addition, considering that enhancers that are exclusively active in larvae, pupae or the adult fly[5–7,15,16] (Supplementary Information section 6), we estimate between at least 50,000 to 100,000 developmental enhancers in the 170-megabase *D. melanogaster* genome. Even though the genome sequence properties (for example, repeat content and gene density) differ, this suggests that the 3-gigabase human genome could contain up to several million enhancers. In summary, the functional characterization of enhancers during the entire *Drosophila* embryogenesis adds a new level of functional annotation to the well-studied fly genome and elucidates global principles of *cis*-regulatory genome organization in animals, the importance of which for development, physiology, evolution and disease is becoming increasingly evident.

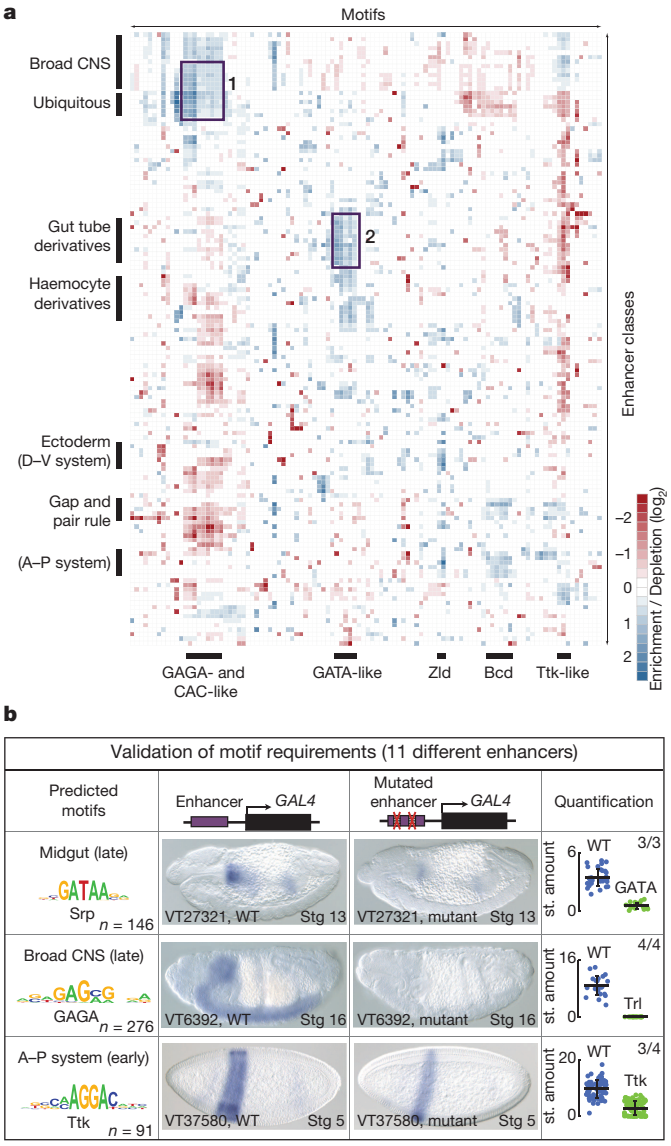



**Figure 4 | Prediction and validation of *cis*-regulatory motif requirements for tissue-specific enhancer activities.** **a**, Global *cis*-regulatory map of transcription factor motif enrichments in sequences of enhancers active in different tissues/cell types. Highlighted are Trl (GAGA) and CAC(N)$_N$CAC-like motifs enriched in CNS and ubiquitous enhancers (1) and GATA-like motifs enriched in midgut enhancers (2; see Extended Data Fig. 9c for the entire map). D–V, dorso–ventral. Zld, Zelda. **b**, Experimental validation of predicted *cis*-regulatory motif requirements. Shown are the most discriminative motifs (left), representative enhancers active in the midgut (stages 13–15), broad CNS (stages 15–16) and A–P system (stages 4–6) and their motif mutant variants (middle), and a quantification of the staining (st) intensities (right; all $P \leq 7 \times 10^{-10}$, Kolmogorov–Smirnov; see Extended Data Figs 9a, b and 10 for details and eight additional enhancers). WT, wild type.

## METHODS SUMMARY

We assessed enhancer activities of 7,705 genomic fragments of about 2 kb in embryos of transgenic *GAL4*-reporter (VT) fly strains obtained from the VDRC (http://stockcenter.vdrc.at/) by *in situ* hybridization. Embryos of each VT strain were manually annotated with a controlled vocabulary and positive strains were imaged. Motif analyses and support vector machine (SVM) predictions were performed as described in ref. 29. All fragment coordinates and annotations are in Supplementary Table 1 and at http://enhancers.starklab.org/.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 7 January; accepted 17 April 2014.**
**Published online 1 June 2014.**

1. Levine, M. Transcriptional enhancers in animal development and evolution. *Curr. Biol.* **20,** R754–R763 (2010).
2. The modENCODE Consortium. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **489,** 57–74 (2010).
3. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
4. Pennacchio, L. A. *et al. In vivo* enhancer analysis of human conserved non-coding sequences. *Nature* **444,** 499–502 (2006).
5. Jenett, A. *et al.* A GAL4-driver line resource for *Drosophila* neurobiology. *Cell Rep.* **2,** 991–1001 (2012).
6. Manning, L. *et al.* A resource for manipulating gene expression and analyzing *cis*-regulatory modules in the *Drosophila* CNS. *Cell Rep.* **2,** 1002–1013 (2012).
7. Jory, A. *et al.* A survey of 6,300 genomic fragments for *cis*-regulatory activity in the imaginal discs of *Drosophila melanogaster*. *Cell Rep.* **2,** 1014–1024 (2012).
8. Zinzen, R. P., Girardot, C., Gagneur, J., Braun, M. & Furlong, E. E. M. Combinatorial binding predicts spatio-temporal *cis*-regulatory activity. *Nature* **462,** 65–70 (2009).
9. Yáñez-Cuna, J. O., Kvon, E. Z. & Stark, A. Deciphering the transcriptional cis-regulatory code. *Trends Genet.* **29,** 11–22 (2013).
10. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457,** 854–858 (2009).

11. Shen, Y. *et al.* A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **448,** 116–120 (2012).
12. Zeitlinger, J. & Stark, A. Developmental gene regulation in the era of genomics. *Dev. Biol.* **339,** 230–239 (2010).
13. Rubin, G. M. & Lewis, E. B. A brief history of *Drosophila*'s contributions to genome research. *Science* **287,** 2216–2218 (2000).
14. Tomancak, P. *et al.* Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* **3,** RESEARCH0088 (2002).
15. Gallo, S. M. *et al.* REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in *Drosophila*. *Nucleic Acids Res.* **39,** D118–D123 (2011).
16. Pfeiffer, B. D. *et al.* Tools for neuroanatomy and neurogenetics in *Drosophila*. *Proc. Natl Acad. Sci. USA* **105,** 9715–9720 (2008).
17. Campos-Ortega, J. A. & Hartenstein, V. *The Embryonic Development of Drosophila melanogaster* (Springer, 1997).
18. Graveley, B. R. *et al.* The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471,** 473–479 (2011).
19. Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature Genet.* **44,** 148–156 (2012).
20. Thomas, S. *et al.* Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. *Genome Biol.* **12,** R43 (2011).
21. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148,** 458–472 (2012).
22. Nègre, N. *et al.* A comprehensive map of insulator elements for the *Drosophila* genome. *PLoS Genet.* **6,** e1000814 (2010).
23. Drosophila 12 Genomes Consortium. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450,** 203–218 (2007).
24. Hong, J.-W., Hendrix, D. A. & Levine, M. S. Shadow enhancers as a source of evolutionary novelty. *Science* **321,** 1314 (2008).
25. Calhoun, V. C., Stathopoulos, A. & Levine, M. Promoter-proximal tethering elements regulate enhancer-promoter specificity in the *Drosophila* Antennapedia complex. *Proc. Natl Acad. Sci. USA* **99,** 9243–9247 (2002).
26. Jack, J., Dorsett, D., Delotto, Y. & Liu, S. Expression of the cut locus in the *Drosophila* wing margin is required for cell type specification and is regulated by a distant enhancer. *Development* **113,** 735–747 (1991).
27. Maeda, R. K. & Karch, F. Gene expression in time and space: additive vs hierarchical organization of *cis*-regulatory regions. *Curr. Opin. Genet. Dev.* **21,** 187–193 (2011).
28. Marinić, M., Aktas, T., Ruf, S. & Spitz, F. An integrated holo-enhancer unit defines tissue and gene specificity of the *Fgf8* regulatory landscape. *Dev. Cell* **24,** 530–542 (2013).
29. Yáñez-Cuna, J. O., Dinh, H. Q., Kvon, E. Z., Shlyueva, D. & Stark, A. Uncovering *cis*-regulatory sequence requirements for context-specific transcription factor binding. *Genome Res.* **22,** 2018–2030 (2012).
30. Liang, H.-L. *et al.* The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. *Nature* **456,** 400–403 (2008).


**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank members of the Dickson laboratory VT project for cloning the candidate regions and generating transgenic flies and the VDRC (http://stockcenter.vdrc.at) for their maintenance and distribution. We are grateful to the IMP/IMBA Scientific Services, in particular BioOptics, Genomics, and IT for help and to C. H. Lampert (IST Austria) for advice. We thank M. Levine (UC Berkeley) and V. Hartenstein (UCLA) for their permission to reproduce figures. The Stark group is supported by a European Research Council (ERC) Starting Grant from the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 242922 awarded to A.S. and by the Austrian Science Fund (FWF, F4303-B09). Generation of transgenic lines was supported in part by an ERC Advanced Investigator Grant to B.J.D. Basic research at the IMP is supported by Boehringer Ingelheim GmbH.

**Author Contributions** E.Z.K. and A.S. conceived the study and wrote the paper. E.Z.K., G.S., M.P. and K.S. performed the screen, E.Z.K. annotated all patterns and performed validation experiments, E.Z.K. and T.K. performed the imaging, T.K. and G.S. analysed the image data, T.K. developed http://enhancers.starklab.org, J.O.Y.-C. performed all sequence analyses, B.J.D. provided transgenic flies prior to publication, E.Z.K., T.K., G.S., J.O.Y-C. and A.S. analysed data. A.S. supervised the project.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.S. (stark@starklab.org).

## METHODS

**Transgenic flies of the VT collection.** All transgenic flies used for the enhancer screen and for testing selected candidates in a second reporter setup were obtained from the VDRC (http://stockcenter.vdrc.at/). Each VT strain contains a transcriptional reporter construct consisting of a genomic sequence fragment (candidate enhancer) followed by the DSCP core promoter and *GAL4* reporter gene, which is integrated into the attP2 landing site on chromosome 3L[16]. Each VTL strain (Supplementary Table 2) contains a transcriptional reporter construct consisting of the same genomic fragment as corresponding VT strain followed by the DSCP core promoter and LexA::GAD reporter gene[31], which is integrated into the attP40 landing site on chromosome 2L.

**Genomic fragments tested in this study.** The 7,705 genomic fragments tested in this study are 2,022 bp on average and cover 13.5% of the entire *Drosophila* non-coding and non-repetitive genome (Extended Data Fig. 1a). Coding and repetitive regions were excluded during the design of the fragments (based on the FlyBase genome annotation r5.4), yet partly re-included if such regions were short and lead to extensive splitting of fragments (Supplementary Table 1). The average overlap between fragments was ~400 bp to avoid that enhancers are split. Information about all fragments and coordinates is available in Supplementary Table 1 and at http://enhancers.starklab.org/. See Extended Data Fig. 1a–d for information about the coverage and distribution of the 7,705 fragments.

**Embryo collection, fixation and automated 96-well whole-mount *in situ* hybridization.** Flies of each transgenic VT strain, carrying an insert of a defined ~2 kb enhancer candidate (Extended Data Fig. 1e), were maintained independently at 25 °C in plexiglass cages (10 × 6 × 6 cm). After 1 h of pre-laying, 0.5–15 h embryos (Bownes stages 2–16 (ref. 17)) were harvested from agar plates, fixated and stored at −20 °C using standard protocols[32].

For each VT strain ~100 µl of embryos (~400 embryos) were distributed into 96 1-ml deep-well plates (NUNC), together with a positive control (VT19021, active in early mesoderm), and processed by an automated *in situ* hybridization protocol using the Bravo Automated Liquid Handling Platform (Agilent Technologies). The protocol was adapted from refs 32 and 33. For consistency, we also re-stained and reanalysed the 149 lines tested before[34]. The antisense *GAL4* probe was generated as described earlier[16,34]. The antisense 1,473 bp *SoxN* probe was generated using primers 5′-CCGCGTTGCACGGACAT-3′ and 5′-**TGTAATACGACTCACTATAGGG**GATCGGGCGAGTCAGTCTGATAG-3′ (bold: T7 promoter sequence for *in vitro* transcription). The antisense 1,440 bp *lexA* probe was generated from a **pBPnlsLexA::GADflUw**[31] vector using primers 5′-ATGCCACCCAAGAAGAAGC-3′ and 5′-**TGTAATACGACTCACTATAGGG**GGACGGGAAGGAAATCG-3′ (bold: T7 promoter sequence). The conditions for *in situ* hybridization were optimized to provide the best signal-to-background ratio. Automated *in situ* hybridization and usage of the same antisense *GAL4* probe allowed robust and highly reproducible detection of enhancer activity across different enhancer candidates.

**Imaging and annotation of enhancer patterns.** After *in situ* hybridization, embryos of each VT strain (~400 embryos per slide) were mounted on barcoded microscope slides. We inspected each slide using a Zeiss Axiophot microscope and manually annotated activity patterns at six time intervals of embryo development (Bownes stages 4–6, 7–8, 9–10, 11–12, 13–14 and 15–16 corresponding to 1.5–3.3 h.p.f., 3.3–4 h.p.f., 4–5.3 h.p.f., 5.3–9.5 h.p.f., 9.5–11.3 h.p.f. and 11.3–15 h.p.f., respectively[14,17]) using a controlled vocabulary adapted from the BDGP[14]. During annotations, we did not consider background signals known to sporadically occur for the reporter system and landing site in the clypolabrum and cephalic furrow of stages 7–8 and in the anterior tip and the region of hypopharynx ventral of the stomodeum at stage 11 (refs 31, 35). We also manually annotated the strength of the signal for each annotation term semiquantitatively from 1 (very weak) to 5 (very strong; Extended Data Fig. 1e). A line was considered as positive if it contained at least one annotation term with strength >1 (excluding enhancers with only very weak (1) activities). This resulted in 3,557 positively annotated VT strains (4,480 (58.1%) if also considering very weak signals). Slides with positive VT strains were imaged by an automated whole-slide acquisition with Metafer Slide Scanning Platform (MetaSystems). We used a ×10 lens and a large aperture to get sufficient depth of field so that the enhancer activity from the complete volume of the embryo was well represented. The pictures were acquired in two channels: fluorescent and bright-field. The fluorescent channel was used for correct auto focusing on embryos and further image analysis. The bright-field channel was corrected for shading online on the microscope and served to detect the activity patterns. All bright-field images are available online at http://enhancers.starklab.org/.

**Experimental validations of pipeline and motif predictions.** All transgenic flies carrying reporters with *eve* stripe 2 enhancer-containing fragments (Extended Data Fig. 2a), fragments of non-*Drosophila* origin (negative controls; Extended Data Fig. 2c), and motif-mutated enhancers (Fig. 4b and Extended Data Fig. 10) were generated using the same strategy as flies carrying VT fragments. The activities of the corresponding reporter constructs were assayed identically to the screen.

**Embryo segmentation and extraction of enhancer patterns.** Individual embryos were identified and their respective binary segmentation extracted by thresholding of the fluorescence channel. First, an approximate global threshold was obtained from small windows at a coarse resolution sampled randomly from the whole slide—we searched for a maximum-a-posteriori estimate by fitting a Gaussian mixture model with expectation-maximization. Next, we refined the threshold at full resolution for each individual embryo. We selected the threshold so that the shape of the largest connected component was close to elliptical and the total length of the boundary between foreground and background was minimal. We also split closely positioned and/or touching embryos. From the resulting segmented objects, we kept only those classified as embryos by a support vector machine (SVM) on the basis of general shape features (area, mean-square error of an ellipse fit, minor axis, eccentricity, solidity). This discarded all the fluorescent clutter and broken/incorrectly segmented embryos. Enhancer patterns were extracted as bright-field intensities of pixels inside the fitted ellipse after positioning the embryos horizontally and resampling the images down to 128 × 64 pixels.

**Reverse clustering of enhancer patterns.** Reverse clustering was performed on 636 representative blastoderm stage embryo images in lateral orientation from 429 strains active in stages 4–6. Hence a 636-dimensional vector of enhancer activities represented each of the 6,415 pixels inside the ellipse fitted to the embryos. We used k-means clustering with k-means++ initialization and ten restarts. The approach is analogous to reverse clustering of spatial gene expression patterns[36].

**Defining enhancer classes for fatemap heatmaps.** To identify patterns that significantly frequently co-occurred with each other (in early and late embryos; Extended Data Fig. 4d–g), we obtained confident enrichment values by correcting the ratios/enrichment values according to the respective counts (Wilson correction as in ref. 37 using a *z*-score of 1.67) between the enhancers that were annotated as driving expression in both patterns (pattern 1 and pattern 2) and the total of enhancers of pattern 1 active in the same stage as pattern 2 over the ratio between the enhancers with pattern 2 and the total number of enhancers.

**Calculating occupancy of ChIP and DHS signals on enhancer DNA.** To assess the occurrence of enhancer associated chromatin features (histone marks, CBP/P300 binding[2] and DNA accessibility[20]) on enhancers, we divided the tested fragments in four classes on the basis of their temporal activity profile (Extended Data Fig. 6a). The early (E) class included enhancers active exclusively at stages 4–8 (1.5–4 h.p.f.; embryo cellularization and gastrulation). Late (L) enhancers drive broad expression only at stages 13–16 (9.5–15 h.p.f.; core events of organogenesis). Middle (M) enhancers drive broad expression only at stages 9–12 (4–9.5 h.p.f.; germband elongation and shortening). Continuous (C) enhancers were active at all stages (1.5–15 h.p.f.). For each fragment from all the different classes (E, L, M, C and negative at all stages fragments), we calculated the median ChIP and DHS signal across the fragment. We then quantile-normalized the median signals across the different experimental data sets to allow direct comparison between the data sets. For every enhancer class, we then calculated the median of the ChIP and DHS signal at each time point and normalized it by subtracting the corresponding median of negative class. Finally, we normalized the resulting values to emphasize the temporal dynamics of the signals such that the time point (Fig. 1f–h and Extended Data Fig. 6a) or the enhancer class (Extended Data Fig. 4d, e) with the highest signal was 1. BiTS-ChIP data[19] was processed equivalently.

**Refining of minimal enhancers using chromatin data.** To refine enhancer fragments identified in this study to the putative minimal elements, we used the published DHS[20], CBP/P300 and H3K4me1 data[2]. We first refined the 4,480 active or very weakly active VT fragments to the boundaries of overlapping DHS regions, which resulted in the refinement of 2,580 (57.6%) such fragments to 3,914 refined putative minimal elements/enhancers. The 1,900 remaining VT fragments that did not overlap with DHS regions, we refined by the union of CBP/P300-bound and H3K4me1-marked regions, resulting in the refinement of 439 (9.8%) additional fragments to 600 minimal elements. The list of all refined minimal elements is provided in Supplementary Table 3.

**Gene to enhancer assignment analysis.** For the assignment of 874 strong enhancers to their target genes we considered all first to fifth degree neighbouring genes. We defined first-degree neighbours on the basis of FlyBase v5.53 (2013_05)[38] as the genes immediately upstream and immediately downstream of the fragment's genomic position. In case the fragment was fully contained within the gene, we included the host gene as a first-degree neighbour. We defined *N*th-degree neighbours analogously to the first-degree neighbours by reaching $N-1$ genes further out both upstream and downstream, that is, reaching to the *N*th gene in both directions. We considered an enhancer to fully match a gene if its activity recapitulated a subset of the gene expression pattern for all stages during which the enhancer was active. A partial match requires matching patterns in at least one stage but not throughout all stages. To avoid potential false positive assignments, we only considered genes with good-quality *in situ* stainings (for example, *Dichaete*: http://insitu.fruitfly.org/cgi-bin/ex/report.pl?ftype=1&ftext=CG5893/) and excluded poor-quality *in situ*

stainings from the analysis (for example, *homothorax*: http://insitu.fruitfly.org/cgi-bin/ex/report.pl?ftype=1&ftext=CG17117/). This analysis resulted in 482 assignments for strong enhancers that comprise 220 different gene loci (the most frequent locus occurs 12 times and the median over all contributing gene loci is 1).

As the BDGP assessed the embryonic expression patterns for 7,686 (46%; FlyBase v5.53 (ref. 38)) of the *D. melanogaster* genes, data for one or more flanking genes were lacking for most enhancers. This made it difficult to interpret situations in which enhancers do not match to any of the flanking genes with characterized expression. We therefore performed an additional analysis that we restricted to enhancers (both strong and weak) for which expression data on all flanking genes were available (Extended Data Fig. 2f). Considering only first-degree neighbours, this was true for 354 of all 3,557 enhancers. For 94 enhancers, all first- and second-degree neighbours were characterized (Extended Data Fig. 2f). All enhancer–gene pairs identified in this study are provided in Supplementary Table 4.

**Gene expression analysis.** To compare the overall gene expression levels in the embryo for target and non-target genes (Fig. 3d), we used published RNA-seq data performed at different embryonic stages[18]. For each gene we calculated the RPKM as an average of the expression in all the developmental stages where the corresponding enhancer was active.

**Calculating the frequency of chromosomal breakpoints.** To obtain the coordinates of chromosomal breakpoints (Fig. 2c and Extended Data Fig. 7b) we used pairwise (*D. melanogaster* centric) genome alignments for *D. melanogaster* with each of the 11 other sequenced Drosophilid species[23,37]. For each of the 11 pairwise alignments, the alignment blocks were sorted on the basis of their chain score and considered the top 1%, 0.5%, 0.1%, 0.05% and 0.01% of the regions. For each of these cutoffs, we selected the starts and ends of the respective regions from each of the 11 pairwise alignments and considered them as chromosomal breakpoints. We then pooled the breakpoints from all species.

**Motif analysis.** All motif analyses were performed as described earlier[29]. In brief, to calculate motif enrichment, we counted motif occurrences of known and predicted motifs from ref. 37 within the enhancers on the basis of a position weight matrixes (PWM)-matching-cutoff $P = 1/4{,}096$. We assessed the statistical significance of the enrichment for each motif by a hypergeometric $P$ value.

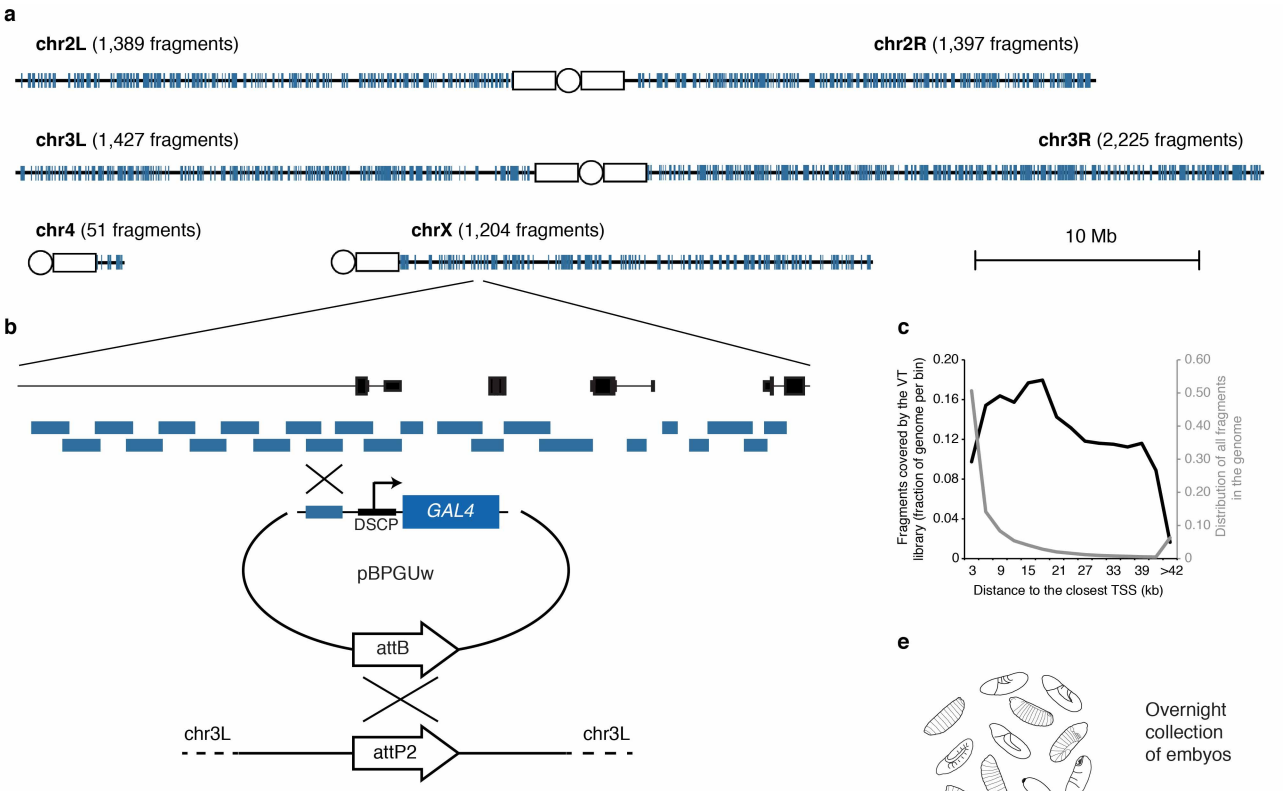
**SVM predictions.** SVM predictions were performed to discriminate VT fragments driving a particular expression pattern against either a set of negative fragments or a set of fragments that were active at the same stage but in a non-overlapping pattern. We used a manual implementation of leave-one-out cross-validation as described earlier[29]. The total number of motif instances for each of the fragments was used as features for SVM. To identify the most discriminative motifs, we first clustered similar motifs on the basis of their position weight matrixes (PWM) to reduce redundancy. For each of the resulting 101 clusters, we chose as representative the motif that best discriminated between the fragments of interest (positive set) and the respective control set (see above). We then performed feature selection by backward elimination to remove motifs with low predictive value[29].

**Embryo segmentation and extraction of enhancer patterns.** We detected individual embryos and extracted their respective binary segmentation by thresholding of the fluorescence channel. First, we identified an approximate global threshold based random sampling of small windows throughout the whole slide at a coarse resolution—we searched for a maximum-a-posteriori estimate by fitting a Gaussian mixture model with expectation-maximization algorithm. Next, we refined the threshold at full resolution for each individual embryo separately. We selected the threshold so that the shape of the largest connected component was close to elliptical and the total length of the boundary between foreground and background was minimal. We also split closely positioned and/or touching embryos. From the resulting segmented objects, we kept only those whose shape resembled an embryo. To this end we used an SVM with general shape features (area, mean-square error of an ellipse fit, minor axis, eccentricity, solidity). This discarded all the fluorescent clutter and broken/incorrectly segmented embryos. Finally, we extracted enhancer patterns from the bright-field channel as greyscale intensities of pixels inside the fitted ellipse while positioning the embryos horizontally and resampling the images down to $128 \times 64$ pixels.

**Quantification of mutation phenotypes.** For each mutation phenotype, we took lateral images of mutated embryos and the corresponding wild-type embryos. We set a threshold on intensities as mean of thresholds estimated using Otsu's method from the images of wild-type embryos only. For all images, we extracted the staining intensity as inverted greyscale intensities of all the pixels below the obtained threshold (staining = threshold − intensity if intensity < threshold). We estimated the amount of staining in each individual embryo as the sum of all these staining intensities. To account for the remaining variance in embryo size and pose, we normalize this estimate by the total area of the foreground. Finally, we plot all the enhancer strength estimates for both wild-type and mutated embryos. To test for a change in enhancer strength, we compared the histograms of staining amount of all wild-type embryos versus all mutated embryos using two-sample Kolmogorov–Smirnov test.

**Statistical analyses and computations.** All statistical computations to obtain binomial, hypergeometric or Wilcoxon $P$ values, to calculate the multiple testing corrected $P$ values and to generate the heatmap plots were done in R (version 2.14). Clustering was done using the clustering tool Cluto for k-means. To intersect coordinates and identify overlapping regions, we used the tool intersectBed from BEDTools[39].


31. Pfeiffer, B. D. *et al.* Refinement of tools for targeted gene expression in *Drosophila*. *Genetics* **186,** 735–755 (2010).
32. Weiszmann, R., Hammonds, A. S. & Celniker, S. E. Determination of gene expression patterns using high-throughput RNA in situ hybridization to whole-mount *Drosophila* embryos. *Nature Protocols* **4,** 605–618 (2009).
33. Lécuyer, E., Parthasarathy, N. & Krause, H. M. Fluorescent *in situ* hybridization protocols in *Drosophila* embryos and tissues. *Methods Mol. Biol.* **420,** 289–302 (2008).
34. Kvon, E. Z., Stampfel, G., Yáñez-Cuna, J. O., Dickson, B. J. & Stark, A. HOT regions function as patterned developmental enhancers and have a distinct *cis*-regulatory signature. *Genes Dev.* **26,** 908–913 (2012).
35. Fisher, W. W. *et al.* DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. *Proc. Natl Acad. Sci. USA* **109,** 21330–21335 (2012).
36. Frise, E., Hammonds, A. S. & Celniker, S. E. Systematic image-driven analysis of the spatial *Drosophila* embryonic expression landscape. *Mol. Syst. Biol.* **6,** 345 (2010).
37. Stark, A. *et al.* Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* **450,** 219–232 (2007).
38. St Pierre, S. E., Ponting, L., Stefancsik, R. & McQuilton, P. The FlyBase Consortium. FlyBase 102–advanced approaches to interrogating FlyBase. *Nucleic Acids Res.* **42,** D780–D788 (2014).
39. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
40. Small, S., Blair, A. & Levine, M. Regulation of even-skipped stripe 2 in the *Drosophila* embryo. *EMBO J.* **11,** 4047–4057 (1992).
41. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser–a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35,** D88–D92 (2007).
42. Hartenstein, V. *Atlas of Drosophila development* (1993).
43. Erives, A. & Levine, M. Coordinate enhancers share common organizational features in the *Drosophila* genome. *Proc. Natl Acad. Sci. USA* **101,** 3851–3856 (2004).
44. Perry, M. W., Boettiger, A. N. & Levine, M. Multiple enhancers ensure precision of gap gene-expression patterns in the *Drosophila* embryo. *Proc. Natl Acad. Sci. USA* **108,** 13570–13575 (2011).
45. Gajewski, K. *et al.* Pannier is a transcriptional target and partner of Tinman during *Drosophila* cardiogenesis. *Dev. Biol.* **233,** 425–436 (2001).
46. Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R. & Eisen, M. B. Zelda binding in the early *Drosophila melanogaster* embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet.* **7,** e1002266 (2011).
47. Nien, C.-Y. *et al.* Temporal coordination of gene networks by Zelda in the early *Drosophila* embryo. *PLoS Genet.* **7,** e1002339 (2011).
48. Nambu, J. R., Lewis, J. O., Wharton, K. A. & Crews, S. T. The *Drosophila* single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell* **67,** 1157–1167 (1991).
49. Ochoa-Espinosa, A. *et al.* The role of binding site cluster strength in Bicoid-dependent patterning in *Drosophila*. *Proc. Natl Acad. Sci. USA* **102,** 4960–4965 (2005).
50. Baylies, M. K. & Bate, M. *twist*: a myogenic switch in *Drosophila*. *Science* **272,** 1481–1484 (1996).
51. Nguyen, H. T., Bodmer, R., Abmayr, S. M., McDermott, J. C. & Spoerel, N. A. D-mef2: a *Drosophila* mesoderm-specific MADS box-containing gene with a biphasic expression profile during embryogenesis. *Proc. Natl Acad. Sci. USA* **91,** 7520–7524 (1994).



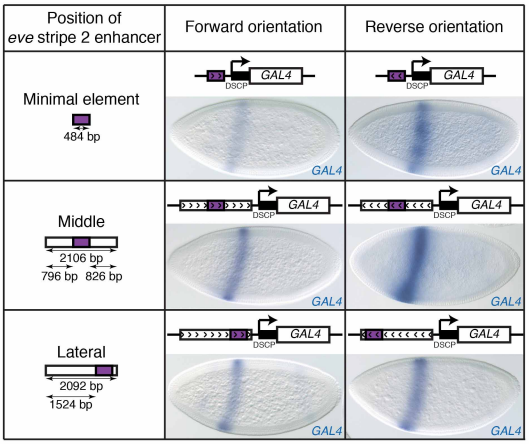

Staining Intensity: (1) very weak; (2) weak; (3) average; (4) strong; (5) very strong

d

| Category of genes | Stage of development | Number of genes | Source | Genome-wide | | Enhancer library | | Deviation |
|---|---|---|---|---|---|---|---|---|
| | | | | Total number of fragments | Number of fragments uniquely assigned to genes | Total number of fragments | Number of fragments uniquely assigned to genes | |
| Any expressed genes | 2–24 h.p.f. | 10,928 | RNA-seq (RPKM > 1)* | 65,911 | 48,595 | 7,705 | 6,290 | **1.11** |
| Highly expressed genes | 2–24 h.p.f. | 2,700 | RNA-seq (Top 20% by RPKM)* | 65,911 | 9,397 | 7,705 | 913 | **0.83** |
| Ubiquitous | Stages 4–16 (1.5–15 h.p.f.) | 877 | mRNA *in situ* hyb.** | 65,911 | 3,281 | 7,705 | 452 | **1.18** |
| Salivary gland | Stages 13–16 (9.3–15 h.p.f.) | 129 | mRNA *in situ* hyb.** | 65,911 | 974 | 7,705 | 41 | **0.36** |
| Tracheal system | Stages 13–16 (9.3–15 h.p.f.) | 289 | mRNA *in situ* hyb.** | 65,911 | 2,123 | 7,705 | 203 | **0.82** |
| Hypopharynx | Stages 13–16 (9.3–15 h.p.f.) | 335 | mRNA *in situ* hyb.** | 65,911 | 2,457 | 7,705 | 235 | **0.82** |
| Fat body | Stages 13–16 (9.3–15 h.p.f.) | 276 | mRNA *in situ* hyb.** | 65,911 | 1,136 | 7,705 | 114 | **0.86** |
| Yolk | Stages 13–16 (9.3–15 h.p.f.) | 218 | mRNA *in situ* hyb.** | 65,911 | 940 | 7,705 | 99 | **0.90** |
| Foregut | Stages 13–16 (9.3–15 h.p.f.) | 407 | mRNA *in situ* hyb.** | 65,911 | 2,743 | 7,705 | 294 | **0.92** |
| Muscle system | Stages 13–16 (9.3–15 h.p.f.) | 670 | mRNA *in situ* hyb.** | 65,911 | 2,485 | 7,705 | 278 | **0.96** |
| Epipharynx | Stages 13–16 (9.3–15 h.p.f.) | 316 | mRNA *in situ* hyb.** | 65,911 | 2,192 | 7,705 | 253 | **0.99** |
| Posterior spiracle | Stages 13–16 (9.3–15 h.p.f.) | 216 | mRNA *in situ* hyb.** | 65,911 | 1,539 | 7,705 | 180 | **1.00** |
| Esophagus | Stages 13–16 (9.3–15 h.p.f.) | 191 | mRNA *in situ* hyb.** | 65,911 | 1,438 | 7,705 | 169 | **1.01** |
| Ventral epidermis | Stages 13–16 (9.3–15 h.p.f.) | 627 | mRNA *in situ* hyb.** | 65,911 | 4,527 | 7,705 | 548 | **1.04** |
| Visceral muscle | Stages 13–16 (9.3–15 h.p.f.) | 370 | mRNA *in situ* hyb.** | 65,911 | 1,971 | 7,705 | 244 | **1.06** |
| Hindgut | Stages 13–16 (9.3–15 h.p.f.) | 805 | mRNA *in situ* hyb.** | 65,911 | 4,103 | 7,705 | 516 | **1.08** |
| Garland cell | Stages 13–16 (9.3–15 h.p.f.) | 163 | mRNA *in situ* hyb.** | 65,911 | 802 | 7,705 | 101 | **1.08** |
| Head epidermis | Stages 13–16 (9.3–15 h.p.f.) | 478 | mRNA *in situ* hyb.** | 65,911 | 3,669 | 7,705 | 463 | **1.08** |
| Proventriculus | Stages 13–16 (9.3–15 h.p.f.) | 337 | mRNA *in situ* hyb.** | 65,911 | 2,093 | 7,705 | 265 | **1.08** |
| Midgut chamber | Stages 13–16 (9.3–15 h.p.f.) | 196 | mRNA *in situ* hyb.** | 65,911 | 689 | 7,705 | 88 | **1.09** |
| Anal pad | Stages 13–16 (9.3–15 h.p.f.) | 319 | mRNA *in situ* hyb.** | 65,911 | 1,716 | 7,705 | 222 | **1.11** |
| Dorsal epidermis | Stages 13-16 (9.3-15 h.p.f.) | 701 | mRNA *in situ* hyb.** | 65,911 | 5,210 | 7,705 | 679 | **1.11** |
| Crystal cell | Stages 13–16 (9.3–15 h.p.f.) | 102 | mRNA *in situ* hyb.** | 65,911 | 481 | 7,705 | 65 | **1.16** |
| Somatic muscle | Stages 13–16 (9.3–15 h.p.f.) | 361 | mRNA *in situ* hyb.** | 65,911 | 1,482 | 7,705 | 209 | **1.21** |
| Central brain glia | Stages 13–16 (9.3–15 h.p.f.) | 257 | mRNA *in situ* hyb.** | 65,911 | 2,892 | 7,705 | 409 | **1.21** |
| Lymph gland | Stages 13–16 (9.3–15 h.p.f.) | 115 | mRNA *in situ* hyb.** | 65,911 | 657 | 7,705 | 93 | **1.21** |
| Ventral nerve cord | Stages 13–16 (9.3–15 h.p.f.) | 1,127 | mRNA *in situ* hyb.** | 65,911 | 7,431 | 7,705 | 1,607 | **1.85** |
| Brain | Stages 13–16 (9.3–15 h.p.f.) | 1,137 | mRNA *in situ* hyb.** | 65,911 | 7,305 | 7,705 | 1,625 | **1.90** |
| Ventral midline | Stages 13–16 (9.3–15 h.p.f.) | 206 | mRNA *in situ* hyb.** | 65,911 | 1,734 | 7,705 | 405 | **2.00** |

**Extended Data Figure 1 | A high-throughput *in vivo* enhancer screen in *Drosophila* embryos.** **a**, The distribution of tested candidate fragments (blue) across the *D. melanogaster* genome (fragments in heterochromatic regions of the chromosomes are not shown). **b**, Schematic representation of the candidate fragments in the VT library used. The cartoon displays a genomic locus (top) with genes (black) and enhancer candidates (blue). The library contains each candidate fragment in a constant transcriptional *GAL4* reporter (middle) integrated in a constant genomic landing site. **c**, Coverage of the genome by the VT library. Shown is the distribution of the tested DNA fragments in the *Drosophila* genome with respect to the distance to the closest gene TSS. **d**, Distribution of the tested DNA fragments in the genome with respect to the expression of the closest gene at relevant developmental stages (2–24 hours post fertilization (h.p.f.)). Each region was uniquely assigned to the closest gene (sixth and eighth columns), and the fraction of genes which are expressed (RPKM values higher than 1), highly expressed (top 20% of genes according to their RPKM values), or active in various tissues of the stages 13–16 embryo is indicated (*embryonic developmental time course RNA-seq data[18]; **RNA *in situ* hybridization data from BDGP[14]; 65,911 fragments cover the whole non-coding, non-repetitive genome and serve as baseline). The deviation from the genome average is typically within the range of 0.8- to 1.2-fold. The few exceptions are genes active in CNS (1.8–2.0-fold enriched) and salivary glands (0.36-fold). **e**, Pipeline for the assessment of enhancer activities in *Drosophila* embryos. We collected about 400 embryos representing all stages of embryogenesis per transgenic VT strain and performed whole-mount 96-well *in situ* hybridization with an antisense *GAL4* probe. Positively stained embryos were imaged and their enhancer activity pattern was annotated manually with a controlled vocabulary. A representative enhancer (VT4454) active in a subset of the dorsal epidermis is shown. All images and annotations are available online (http://enhancers.starklab.org/).

a



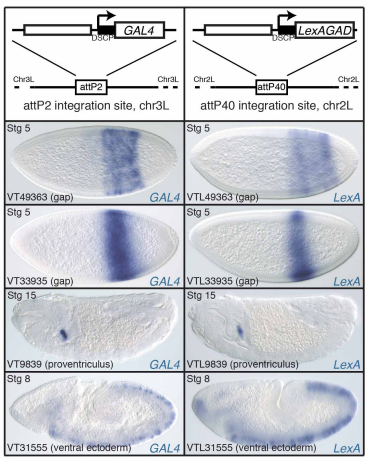

b

| REDfly enhancer name | Target gene | VT fragment | Chr | Start | End | Positive | Match |
|---|---|---|---|---|---|---|---|
| slp1_5-4 | *slp1* | VT1967 | chr2L | 3820069 | 3822166 | Yes | Yes |
| Mef2_I-D[s] | *Mef2* | VT14335 | chr2R | 5817792 | 5819902 | Yes | Yes |
| Mef2_II-A | *Mef2* | VT14336 | chr2R | 5819511 | 5821746 | Yes | Yes |
| Mef2_II-B[L] | *Mef2* | VT14337 | chr2R | 5821284 | 5823462 | Yes | Yes |
| eve_260-hsp70lacZ | *eve* | VT14361 | chr2R | 5861246 | 5863407 | Yes | Partialy |
| eve_stripe_3+7 | *eve* | VT14362 | chr2R | 5862739 | 5864868 | Yes | Yes |
| gsb_fragIV | *gsb* | VT22199 | chr2R | 20944201 | 20946383 | Yes | Yes |
| gsb_GLE | *gsb* | VT22200 | chr2R | 20945508 | 20947613 | Yes | Yes |
| h_betah34 | *h* | VT27677 | chr3L | 8656203 | 8658330 | Yes | Yes |
| h_h7_element | *h* | VT27678 | chr3L | 8657952 | 8660041 | Yes | Yes |
| h_stripe_6 | *h* | VT27679 | chr3L | 8659642 | 8661749 | Yes | Yes |
| h_302 | *h* | VT27680 | chr3L | 8661294 | 8663516 | Yes | Yes |
| Ilp4_Mes3_early_embryonic_enhancer | *Ilp4* | VT28266 | chr3L | 9796953 | 9797904 | Yes | Yes |
| kni_KH | *kni* | VT33934 | chr3L | 20688420 | 20690975 | Yes | Yes |
| lab_0.5 | *lab* | VT37468 | chr3R | 2492009 | 2494238 | Yes | Partially |
| Dfd_NAE | *Dfd* | VT37535 | chr3R | 2624667 | 2626918 | Yes | Yes |
| ftz_5'B | *ftz* | VT37567 | chr3R | 2684413 | 2686617 | Yes | Yes |
| Scr_0.8XH | *Scr* | VT37568 | chr3R | 2686006 | 2688195 | Yes | Yes |
| ato_RE | *ato* | VT38330 | chr3R | 4098467 | 4100635 | Yes | Yes |
| hb_HZ526 | *hb* | VT38550 | chr3R | 4526703 | 4528807 | Yes | Yes |
| ems_elementIV | *ems* | VT41282 | chr3R | 9720242 | 9722389 | No | No |
| srp_A7.1EB | *srp* | VT42352 | chr3R | 11816411 | 11818615 | Yes | Partially |
| tin_tinD | *tin* | VT45194 | chr3R | 17209116 | 17211326 | Yes | Yes |
| bap_bap3 | *bap* | VT45197 | chr3R | 17216118 | 17218205 | Yes | Yes |
| fkh_salivary_gland_enhancer | *fkh* | VT48946 | chr3R | 24418578 | 24420833 | Yes | Yes |
| vnd_early_embryonic_enhancer | *vnd* | VT54910 | chrX | 485819 | 487844 | Yes | Yes |
| oc_otd-186 | *oc* | VT58873 | chrX | 8546539 | 8548888 | Yes | Yes |
| oc_otd-186 | *oc* | VT58874 | chrX | 8548254 | 8550343 | Yes | Yes |

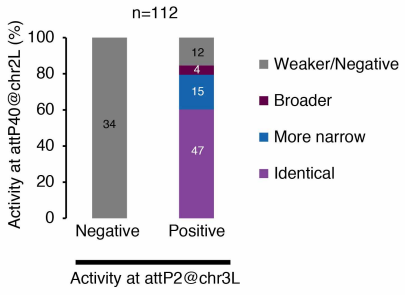
c

| Name of the element | Type | Size | Expression in Mouse embryos | Expression in *Drosophila* embryos |
|---|---|---|---|---|
| pBPGUw | Empty reporter vector | 1,711 | - | * |
| Element 1862 | Human enhancer | 2,606 | Heart | Very weak in ectoderm at stage 12 |
| Element 1670 | Human enhancer | 1,628 | Heart | No activity |
| Element 463 | Human enhancer | 1,159 | Heart | No activity |
| Element 268 | Human enhancer | 1,112 | CNS | No activity |
| Element 1022 | Human enhancer | 1,379 | CNS | No activity |
| Element 488 | Human enhancer | 1,755 | CNS | No activity |
| Element 1258 | Human enhancer | 1,547 | CNS | Very weak in ectodern at stage 16 |
| Element 1082 | Human enhancer | 1,970 | CNS | No activity |
| Element 1453 | Human enhancer | 2,653 | CNS | No activity |
| Element 1344 | Human enhancer | 1,662 | Brain | No activity |
| Element 138 | Human enhancer | 2,549 | CNS | No activity |
| Element 141 | Human enhancer | 1,971 | Eye | No activity |

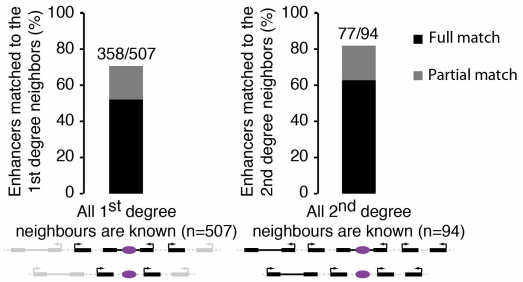
d



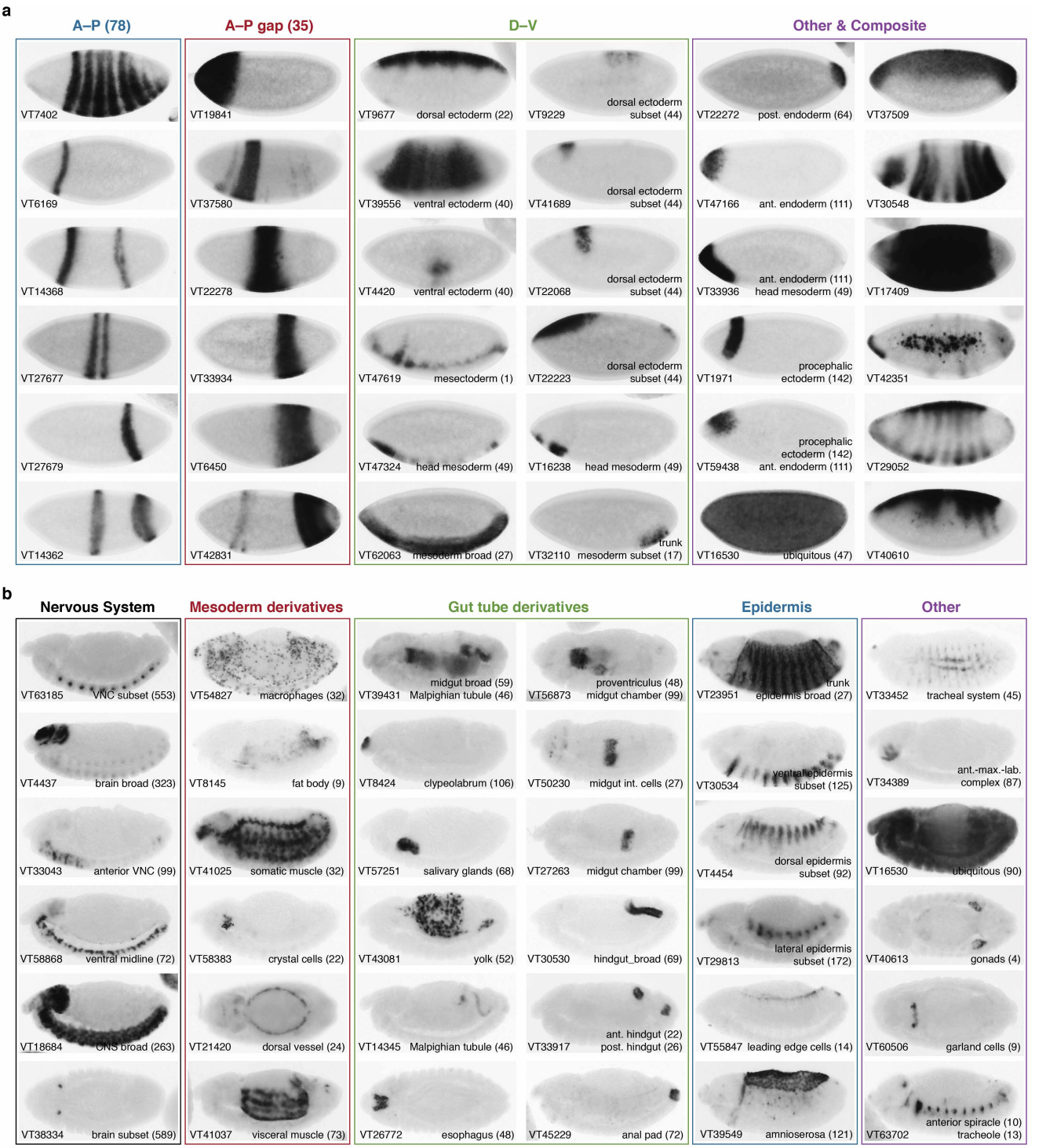

e



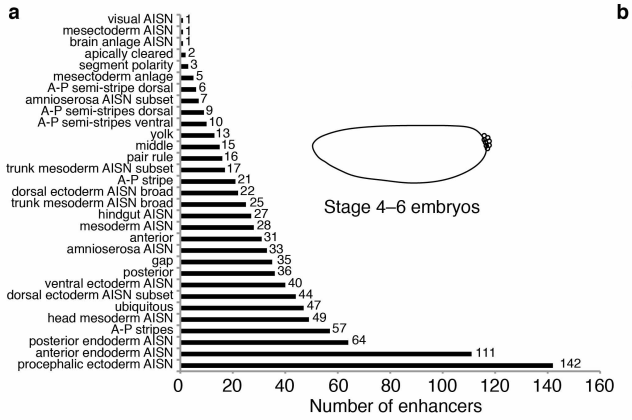

f

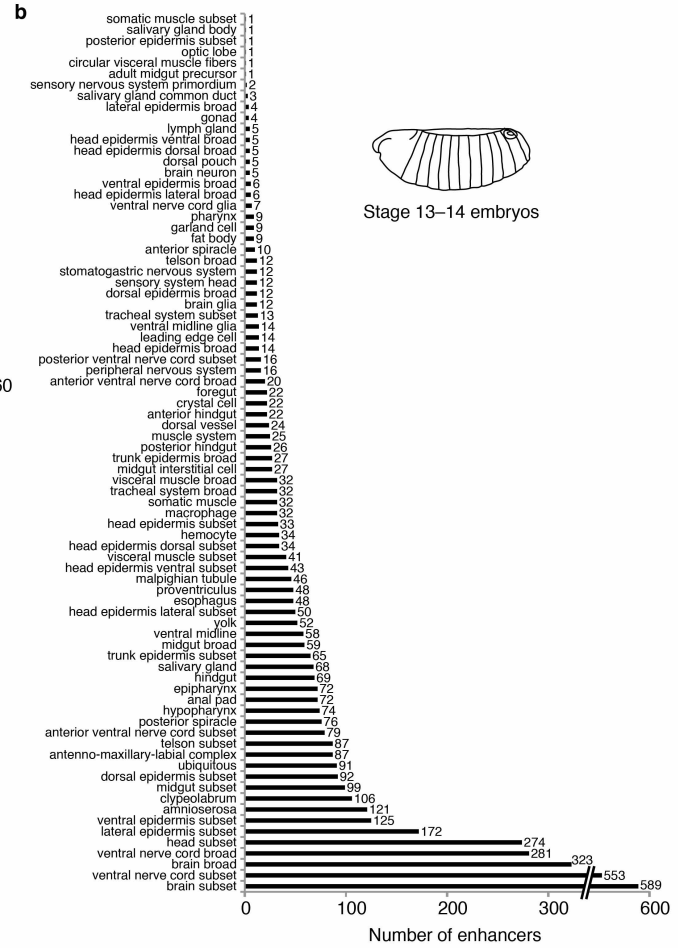


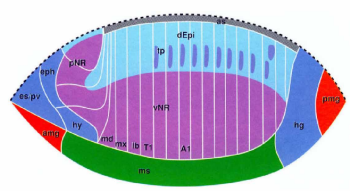

**Extended Data Figure 2 | Validation of the screening pipeline and context-independence of enhancer activities.** **a**, Reporter activity of the *eve* stripe 2 minimal enhancer[40] in the context of longer genomic fragments (the average size of the fragments tested in this study is 2,022 base pairs (bp)). The minimal 484 bp *eve* stripe 2 enhancer[40] was cloned in different positions within longer fragments corresponding to the enhancer's extended genomic context: minimal element alone (first row), in the middle (second row) or towards one side of a ~2 kb fragment (third row), each in two different orientations (forward, second column; reverse, third column). In all cases the endogenous *eve* stripe 2 activity[40] was reproduced. **b**, Recovery of known *Drosophila* enhancers. Twenty-eight previously published non-redundant enhancers (first column, enhancers with available images from the REDfly database[15]) fully overlapped with tested VT fragments (third column). For 24 of them, the VT fragment fully or partially recapitulated the published pattern. Of the four REDfly enhancers, for which the activity was not fully reproduced, three partially matched to the pattern of the VT fragment and only one was inactive in our screen. **c**, Activity of the human DNA sequences in the *Drosophila* embryos. Only two of 12 human ultra-conserved enhancer sequences[41] that we tested in the screen's reporter setup were very weakly active, below the threshold we used to defined active enhancers. The inactivity of all 12 non-*Drosophila* fragments serves as a measure of specificity and demonstrates that less than 1 in 12 (~8%) of random fragments are expected to be active. **d**, The majority of enhancers show identical activity patterns in the context of two different transcriptional reporter systems (different reporter vector and genomic position/landing site). Shown are representative enhancer activities of the fragments using the *GAL4* reporter integrated at the attP2 landing site (VT strains, left column) or using a LexAGAD reporter integrated at the attP40 landing site (VTL strains, right column; Supplementary Information section 1). We re-tested a total of 112 fragments in VTL strains, of which 34 were negative and 78 positive in the original screen in VT strains. **e**, Fraction (*y* axis) and total number (numbers inside bars) of fragments that were negative (right bar) or positive (left bar) in VT embryos that in VTL embryos displayed identical (purple), more narrow (dark blue), broader (dark red) and weaker or no activity (grey). The activities measured in both independent reporter systems agreed very well (Supplementary Table 2) and according to our experience, most differences likely stem from differences in *in situ* sensitivities rather than enhancer function. **f**, Systematic comparison of enhancer activities to the expression patterns of the neighbouring genes for enhancers for which the expression patterns of all respective neighbours are available[14]. Shown is the fraction of enhancers (bar heights), which fully (black) or partially (grey) matched to the first (left) and first or second (right) degree neighbouring genes.



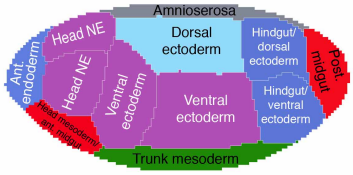

**Extended Data Figure 3 | Enhancers display diverse tissue and cell-type-specific activity patterns.** Shown are representative A–P, D–V, composite and other activity patterns observed in early embryos (**a**, stages 4–6, 1.5–3.3 h.p.f.) and in various tissues/cell types of late embryos (**b**, stages 13–14, 9.5–11.3 h.p.f.). The number of VT fragments per pattern class is indicated in parenthesis (see Extended Data Fig. 4a, b for detailed numbers on all patterns).



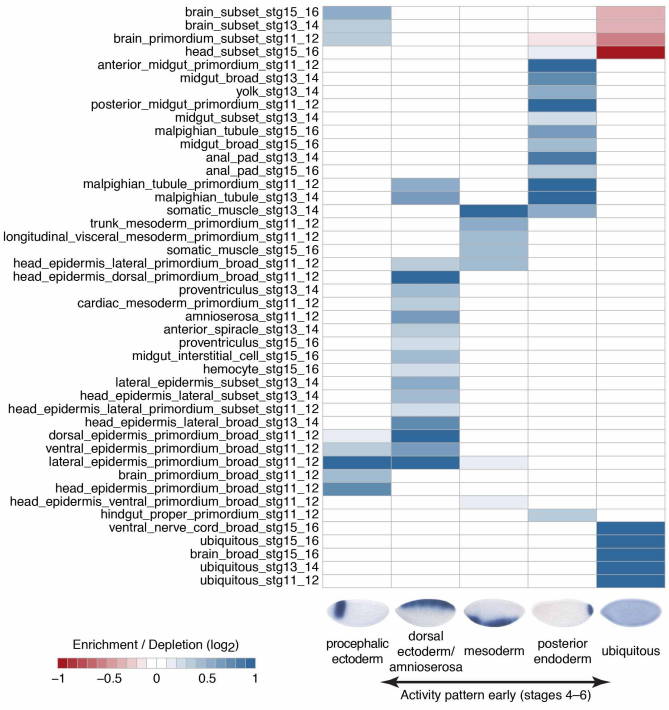



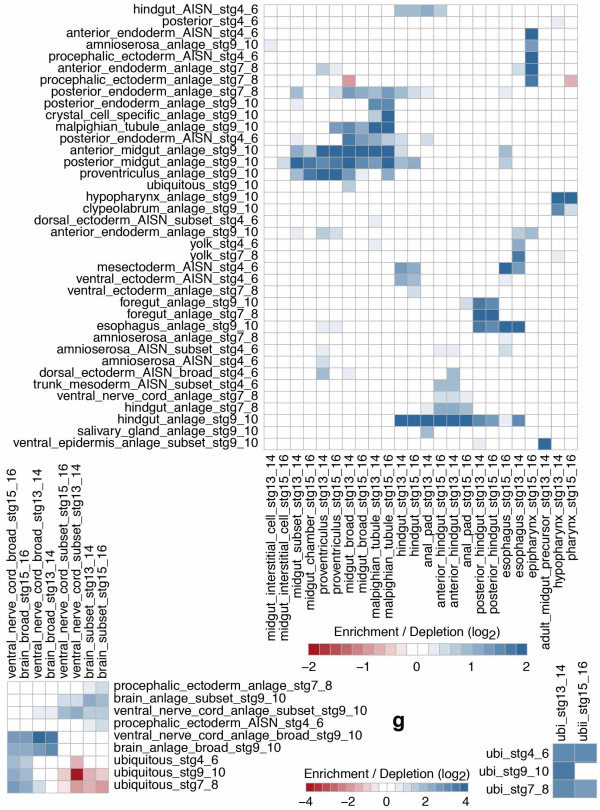




Fate map of the *Drosophila* blastoderm

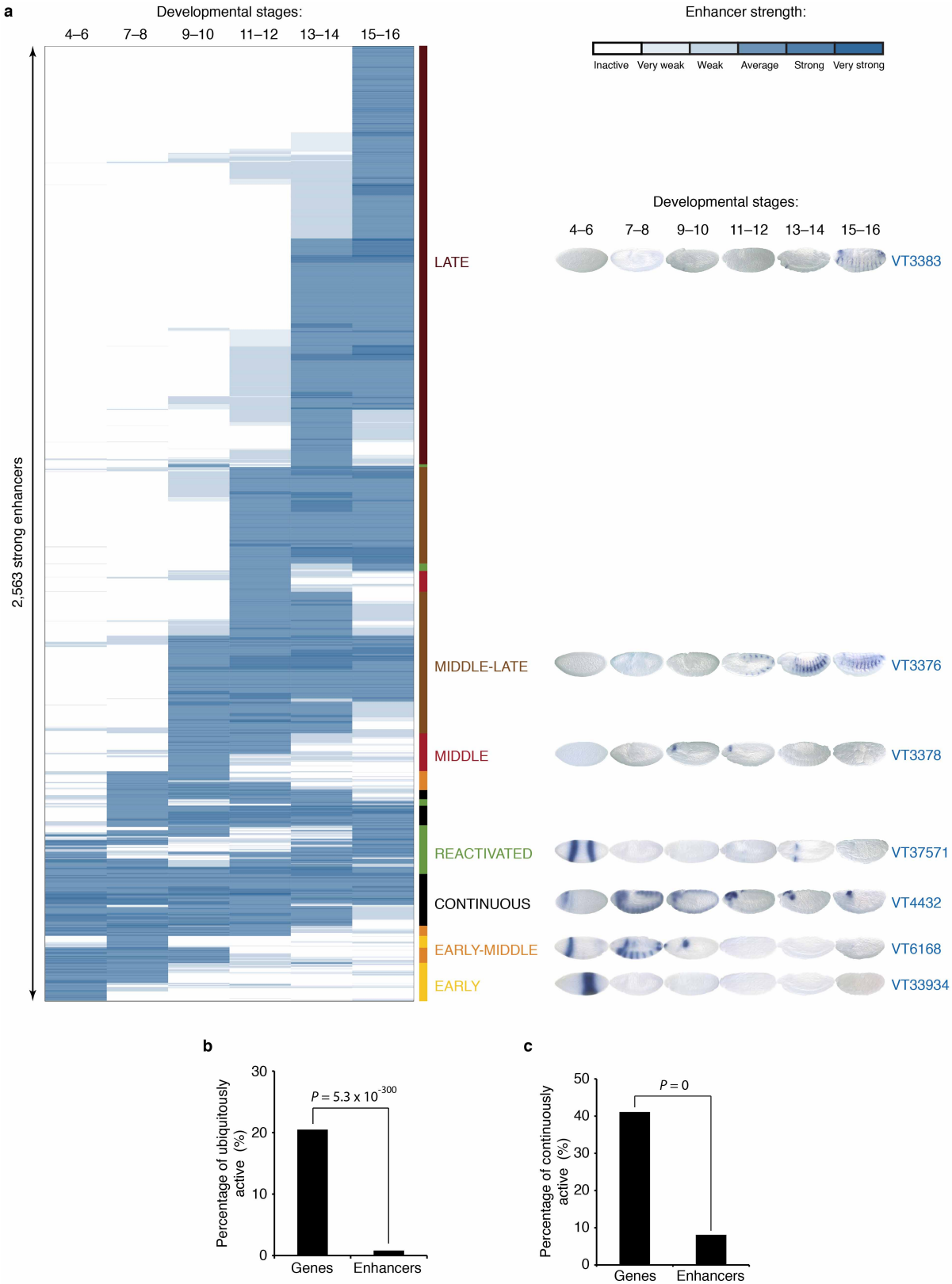



Computationally defined co-regulated domains
(Images of 429 early enhancers)

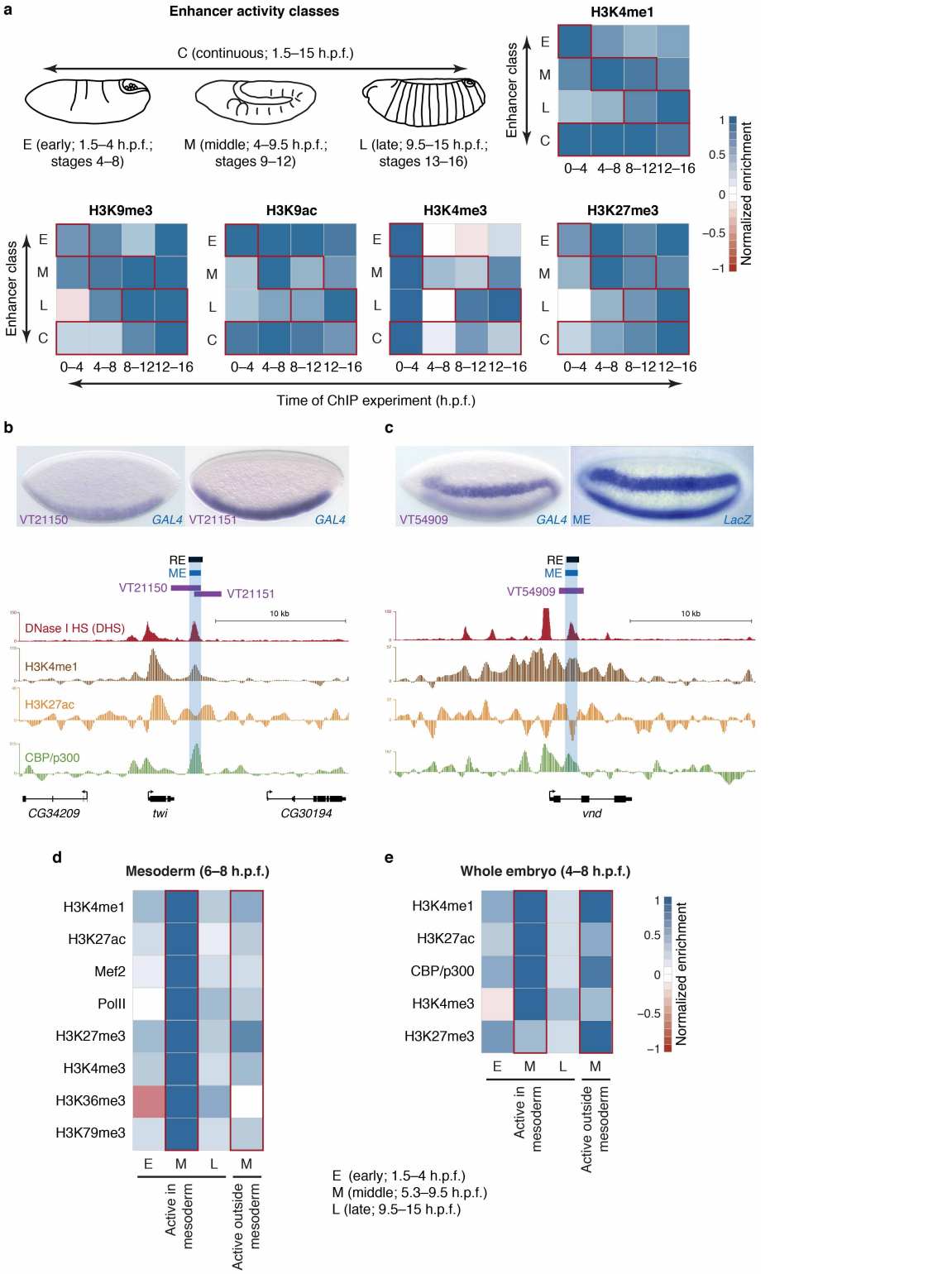


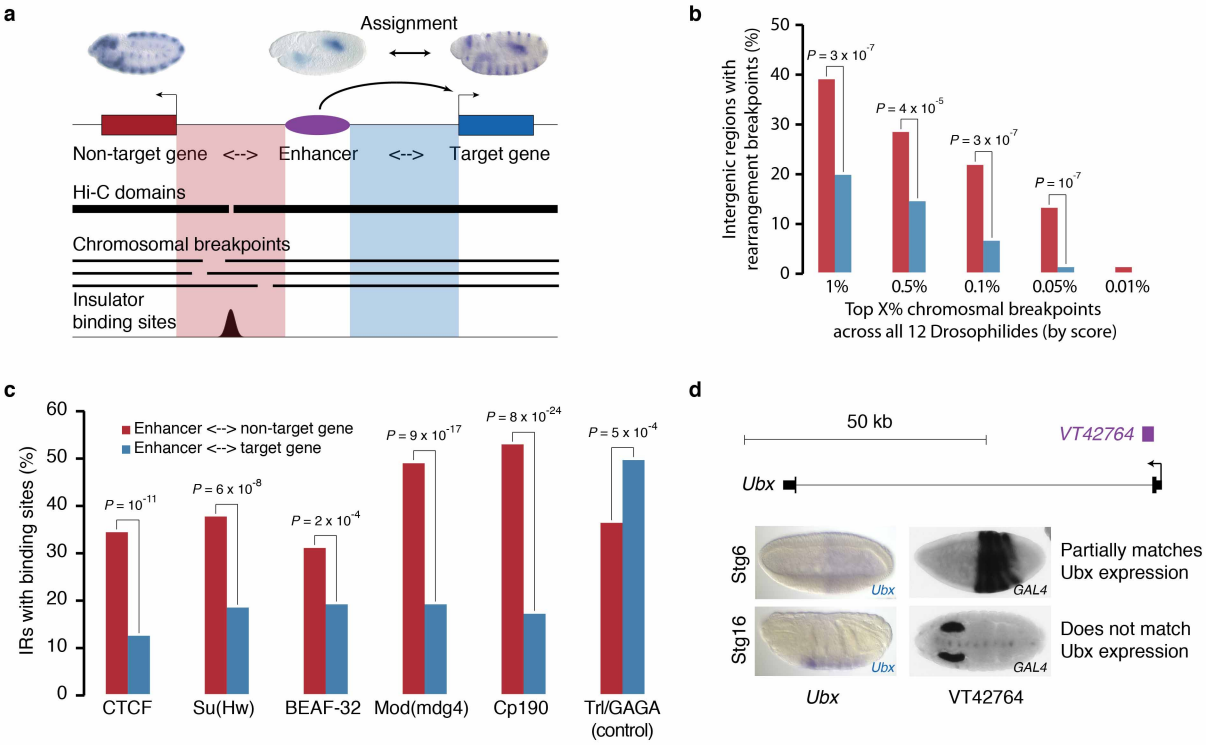
**Extended Data Figure 4 | Spatial activity patterns trace tissue fatemap and development.** **a**, **b**, Total number of VT fragments active in a given anatomical structure (**a**, stages 4–6; or **b**, stages 13–14). **c**, Co-regulated domains resemble the embryo fate map. The schematic embryo on top shows the fate map of the *Drosophila* blastoderm (reproduced from ref. 42 with permission) and the one below shows co-regulated domains determined by unbiased reverse clustering of the raw microscope images for 429 early enhancers. These domains correspond to the following presumptive germ layers (colour coded and manually annotated, from left to right): anterior endoderm (corresponds to es/pv. (oesophagus/proventriculus), eph (epipharynx) and hy (hypopharynx) of the fate map on the top; dark blue), anterior midgut (corresponds to amg; red), head mesoderm (corresponds to ms (mesoderm)), head neuroectoderm (head NE; corresponds to pNR (procephalic neurogenic region); purple), ventral ectoderm (corresponds to vNR (ventral neurogenic region); purple and dark blue), trunk mesoderm (corresponds to ms; green), dorsal ectoderm (corresponds to dEpi (dorsal epidermis); light blue and dark blue), amnioserosa (corresponds to as; grey) hindgut (corresponds to hg; dark blue) and posterior midgut (corresponds to pmg; red). **d**, Enhancer activities during embryogenesis follow the presumptive tissue fate map. Columns represent enhancers active in major presumptive cell types or ubiquitously in all cell types of the early *Drosophila* embryo. Rows show the corresponding most strongly enriched annotation terms at later stages (rows for which all corrected enrichment values were below $2^{0.25}$ ($\sim$1.19) were excluded; rows and columns are clustered). **e**–**g**, The same as in **d** but now going from late to early stages. For enhancers active in the late embryo (stages 13–16, columns) specifically in the gut (**e**), CNS (**f**) or ubiquitously (**g**), the most enriched terms for earlier stages are shown (stages 4–10; rows). Only enhancers that are active in both early and late stages were considered and rows and columns were clustered.




**Extended Data Figure 5 | Enhancer activities are highly dynamic during *Drosophila* embryogenesis.** **a**, The majority of all enhancers are active only during specific time points of embryo development. Heatmap representation of 2,563 strong enhancers, clustered on the basis of their temporal activity profiles during six-stage intervals of *Drosophila* embryogenesis. Seven major temporal groups of enhancers are indicated on the right, together with representative transgenic embryos for each of the groups. **b**, **c**, Enhancers display activity patterns that are temporally and spatially sparser than gene expression patterns (statistical significance was estimated by binomial *P* values shown above the bars). **b**, The fraction of genes that are ubiquitously expressed at all stages (20.5%) was >25 fold higher than the fraction of ubiquitously active enhancers (0.8%). **c**, Similarly, the percentage of continuously expressed genes (41.1%) was >5 fold higher than the percentage of continuously active enhancers (8.1%). We considered stages 4 through 16 (1.5–14 h.p.f.) from the Berkeley Drosophila Genome Project (BDGP)[14].

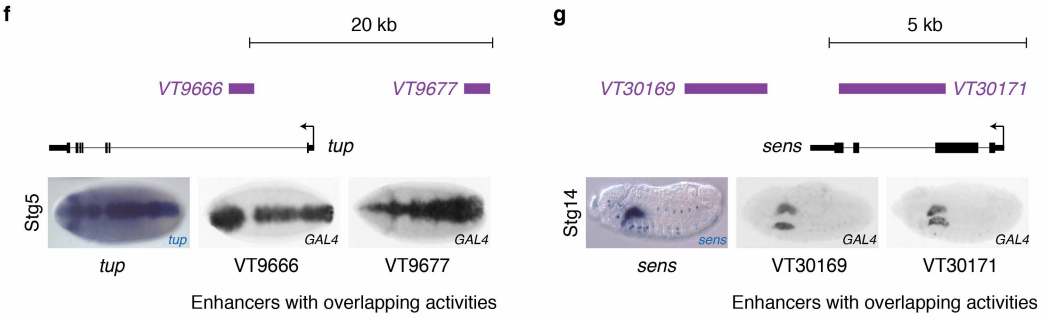



**Extended Data Figure 6 | The dynamics of enhancer activities are reflected in the dynamics of the chromatin landscape.** **a**, H3K4me1, H3K9me3, H3K9ac, H3K4me3 and K3K27me3 histone marks[2] of early (E), middle (M), late (L) and continuous (C) enhancers (data from whole embryos). **b**, **c**, ChIP and DHS signals help to refine minimal enhancer elements. UCSC Genome Browser screen shots showing the *twist* (*twi*; **b**) and *ventral nervous system defective* (*vnd*; **c**) genomic loci including stage 5 DNA accessibility[20] and 0–4 h.p.f. ChIP data for H3K4me1, H3K27ac and CBP/p300 (ref. 2). Displayed are genomic fragments tested in this study (purple), which could be refined to smaller elements (RE, refined elements; grey) that coincide with known minimal enhancers (ME; blue[15,19]). The corresponding transgenic embryos show the respective enhancer activities during stage 5 in mesoderm (**b**) or neuroectoderm (**c**; right image is reproduced from ref. 43 with permission, copyright (2004) National Academy of Sciences, USA; ME and LacZ added to image for clarity). See Supplementary Table 3 for the full list of refined elements. **d**, Tissue-specific 6–8 h.p.f. ChIP signals for histone marks, the Mef2 transcription factor and PolII binding (data from Batch isolation of tissue-specific chromatin for immunoprecipitation (BiTS-ChIP)[19]) on early (E), middle (M) and late (L) mesodermal enhancers and on enhancers active outside the mesoderm (analysis as in ref. 19 but evaluated for enhancers from this study). **e**, The corresponding ChIP signals from whole embryos[2] for comparison.

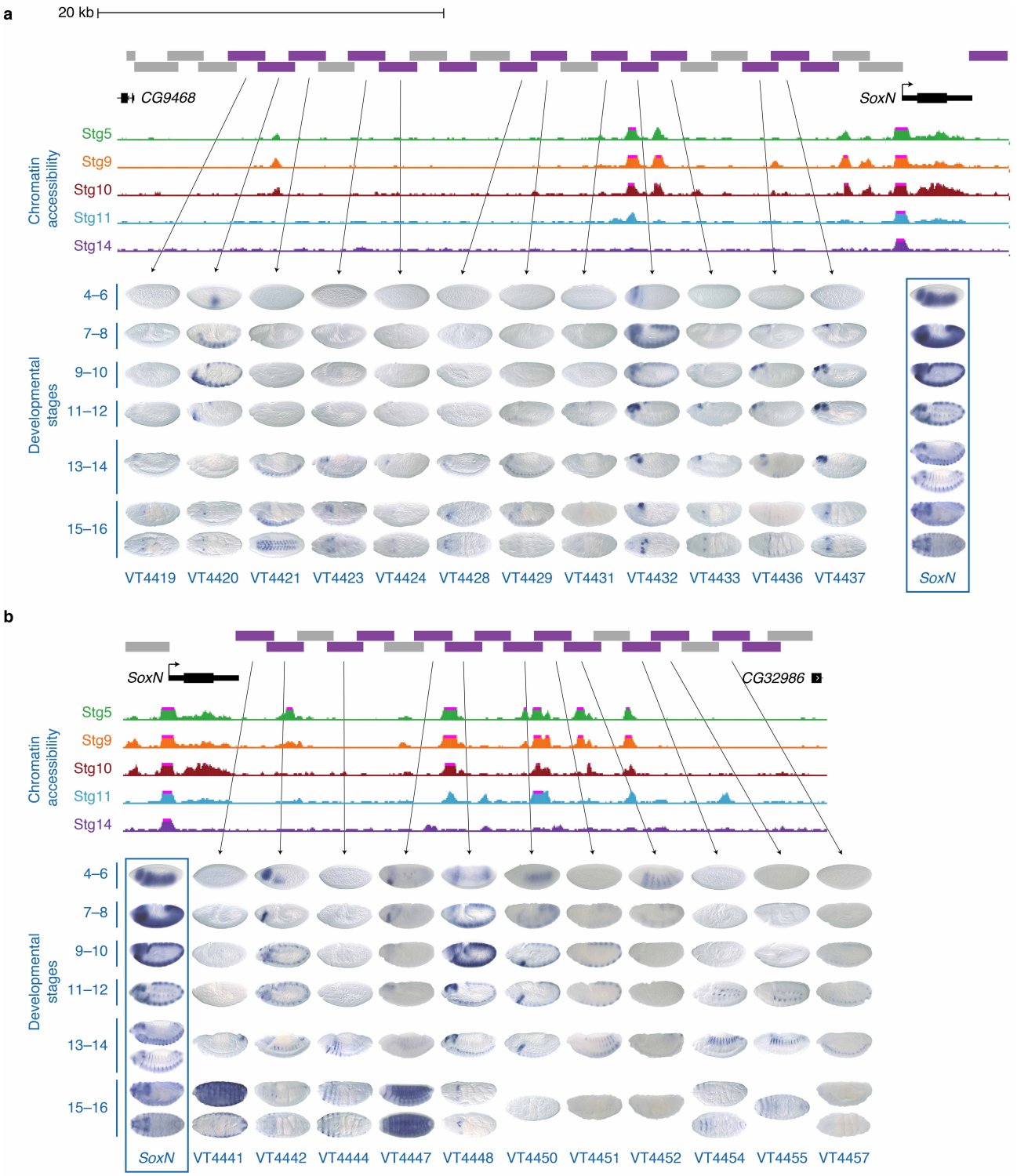



e

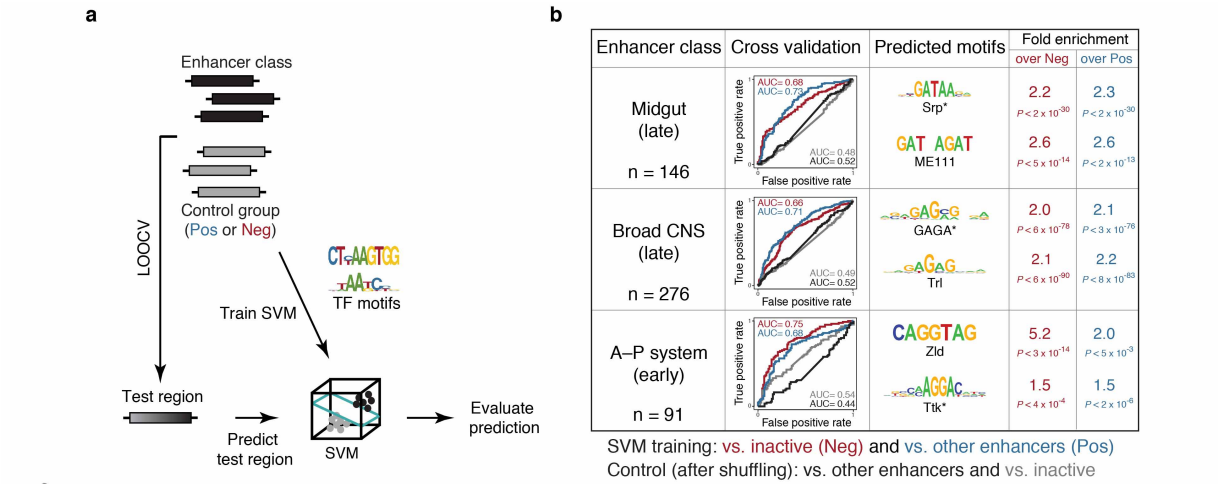
| Gene Name | FBgn | Activity Pattern at stages 4–6 | Number of enhancers | Enhancer IDs | Overlap | Reference |
|---|---|---|---|---|---|---|
| *Abd-B* | FBgn0000015 | posterior | 3 | VT42831,VT42837,VT42848 | Almost full | |
| *bab2* | FBgn0025525 | posterior_endoderm_AISN | 3 | VT23827,VT23832,VT23870 | Partial | |
| *CG10479* | FBgn0035656 | head_mesoderm_AISN | 2 | VT26323,VT26329 | Partial | |
| *D* | FBgn0000411 | gap | 2 | VT30548,VT30559 | Partial | |
| *dan* | FBgn0039286 | ventral_ectoderm_AISN | 2 | VT47167,VT47173 | Almost full | |
| *Dl* | FBgn0000463 | procephalic_ectoderm_AISN | 2 | VT44107,VT44122 | Partial | |
| *Dll* | FBgn0000157 | dorsal_ectoderm_AISN | 2 | VT22068,VT22094 | Partial | |
| *esg* | FBgn0001981 | ventral_ectoderm_AISN | 2 | VT7841,VT7847 | Partial | |
| *gt* | FBgn0001150 | anterior_endoderm_AISN | 3 | VT55791,VT55793,VT55797 | Partial | |
| *gt* | FBgn0001150 | gap | 2 | VT55790,VT55795 | Partial | |
| *hb* | FBgn0001180 | procephalic_ectoderm_AISN | 2 | VT38545,VT38550 | Partial | |
| *oc* | FBgn0004102 | procephalic_ectoderm_AISN | 2 | VT58867,VT58873 | Almost full | Perry et al., 2011 |
| *Optix* | FBgn0025360 | anterior_endoderm_AISN | 2 | VT13348,VT13350 | Partial | |
| *prd* | FBgn0003145 | A-P_stripe | 2 | VT6165,VT6169 | Partial | |
| *SoxN* | FBgn0029123 | procephalic_ectoderm_AISN | 3 | VT4432,VT4442,VT4447 | Partial | |
| *SoxN* | FBgn0029123 | ventral_ectoderm_AISN | 4 | VT4420,VT4442,VT4448,VT4450 | Partial | |
| *tup* | FBgn0003896 | dorsal_ectoderm_AISN | 2 | VT9666,VT9677 | Almost full | |
| *vnd* | FBgn0261930 | ventral_ectoderm_AISN | 2 | VT54907,VT54910 | Almost full | Hong et al., 2008 |
| *vnd* | FBgn0261930 | procephalic_ectoderm_AISN | 2 | VT54905,VT54910 | Almost full | |



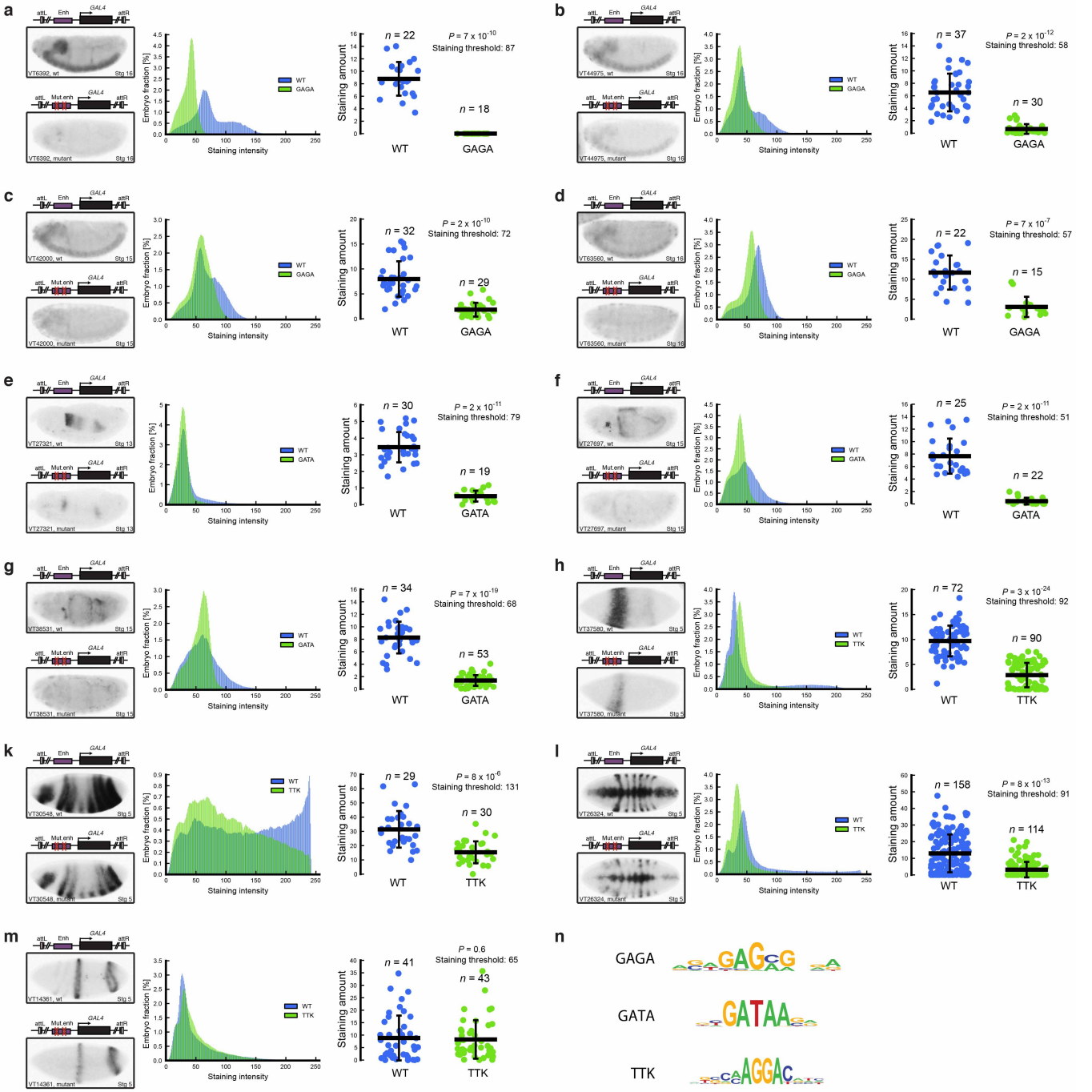
**Extended Data Figure 7 | The *cis*-regulatory organization of the *Drosophila* genome.** **a**, We inspected the intergenic regions (IRs) between intergenic enhancers (purple) and their assigned target genes (blue) and the neighbouring gene that was not assigned (red) for the location of chromosomal domain boundaries as determined by Hi-C[21] (Fig. 2b), breakpoints during evolutionary chromosomal rearrangements (between 12 *Drosophila* species[23]), and insulator protein binding sites[22]. We restricted this analysis to the 151 intergenic enhancers for which both immediately flanking genes were characterized and that were uniquely assigned to one of these first-degree neighbours (see also Fig. 2a). **b**, Breakpoints during evolutionary genome rearrangements (from 12 *Drosophila* species[23]) are significantly reduced between enhancers and their target genes (blue) compared to enhancers and genes they do not regulate (red). Shown are different score cutoffs to define breakpoints with increasing stringency (left to right), which all show a consistent trend and an increasing difference with increasing stringency. **c**, The location of insulator protein binding sites correlate with enhancer–target gene assignments. Bar plots show the fraction of IRs between enhancers and their assigned target genes (blue) or non-target genes (red) that contain at least one binding site for one of the insulator proteins CTCF, Suppressor of Hairy wing (Su(Hw)), Boundary element-associated factor of 32kD (BEAF-32), Modifier of mdg4 (Mod(mdg4)), Centrosomal protein 190kD (Cp190), or the transcriptional activator Trl as a control (1% false discovery rate regions from ref. 22). Statistical significance in **b**, **c** was estimated by binomial *P* values (above the bars). **d**, An enhancer located in the *Ultrabithorax* (*Ubx*) intron does not fully match the *Ubx* expression pattern (overview of the *Ubx* locus on top). The VT42746 fragment drives a characteristic gap gene expression pattern in the early stage 6 embryo (upper right embryo), which recapitulates the known *Ubx* pattern at that stage. During stage 15, however, it displays strong activity in salivary glands (lower right embryo) which does not match *Ubx* expression[14]. **e**, Genes that appear to be regulated by more than one enhancer with identical or overlapping activity patterns (shadow enhancers[24]) at stages 4–6, including known shadow enhancers[24,44]. **f**, **g**, Examples of shadow enhancers in the *tailup* (*tup*, **f**) and *senseless* (*sens*, **g**) gene loci. Displayed are genomic fragments (purple) that act as enhancers with overlapping activity patterns. The corresponding transgenic embryos highlighting enhancer activity during stage 5 in dorsal ectoderm and amnioserosa (**f**) or during stage 14 in salivary glands (**g**) together with wild-type embryos stained against *tup* (**f**) or *sens* (**g**) mRNA[14] are shown below.




**Extended Data Figure 8 | The *cis*-regulatory organization of the *SoxN* locus.** **a**, **b**, The regulatory landscape of the *SoxN* locus (**a**, upstream; **b**, downstream). UCSC Genome Browser screenshot including DNA accessibility data as determined by DHS-seq[20] and genomic fragments tested in this study (top: positive fragments are purple, negative grey). For strong enhancers, the corresponding *GAL4*-stained transgenic embryos are shown below at six time points of embryo development. The boxed embryos show the *in situ* hybridization against *SoxN* mRNA for each developmental stage.



**Extended Data Figure 9 | Predicting *cis*-regulatory motifs requirements across *Drosophila* tissues and cell types.** **a**, Schematic overview of our approach. We trained a support vector machine (SVM) to distinguish between functionally similar enhancers (black solid blocks) and control fragments (grey blocks: Neg, negative regions; Pos, enhancers with different activity patterns) solely on the basis of their motif content. We excluded each fragment in turn for testing, trained the SVM on the remaining fragments, and predicted the test fragment as described before[29]. **b**, Representative enhancers active in the midgut (stages 13–15), broad CNS (stages 15–16) and A–P system (stages 4–6) and the corresponding receiver-operating-characteristic (ROC) curves together with area under the curve (AUC) values for predictions of each of the groups (red curve, versus inactive control regions; blue curve, versus other enhancers) and for controls, for which we randomized the enhancers' class assignments (black curve, versus inactive control regions; grey curve, versus other enhancers). The predictions were not successful when we shuffled the enhancers' assignments between classes, demonstrating that the predictive signals reside within the enhancer sequences rather than stemming merely from the computational procedure per se[29] (see Supplementary Table 5 for more enhancer groups and technical details). The most discriminative transcription factor motifs together with their enrichments over control regions are shown in columns 3–5. **c**, Motif-to-tissue associations revealed by motif-enrichment analyses. The heatmap columns represent a subset of all known *Drosophila* transcription factor motifs[29,37] which were discriminative during supervised machine learning and the rows show different enhancer classes. Each matrix cell shows the enrichment of the corresponding motif in enhancers of the corresponding enhancer class ($\log_2$); only rows and columns for which at least one matrix cell has enrichment values $\geq 2^{0.7}$ are shown and the matrix rows and columns are sorted by bi-clustering. Further highlighted are broad CNS and ubiquitous enhancers enriched in Trl (GAGA) and CAC(N)NCAC-like motifs (region 1); midgut tube enhancers enriched in GATA-like motifs (region 2); embryonic heart enhancers enriched in motifs for Tinman and Pannier[45] (region 3); early A–P and D–V enhancers enriched in Zelda motifs[30,46,47] (regions 4, 6a and 7a); ventral midline enhancers enriched in motifs for Single-minded[48] (Sim) (region 5); anterior endoderm, procephalic ectoderm and A–P system enhancers enriched in motifs for Bicoid (Bcd)[49] and Ttk (regions 6b and 7b); early mesoderm enhancers enriched in Twi motifs[50] (region 8) and late somatic muscle enhancers enriched in Mef2 motifs[51] (region 9).




**Extended Data Figure 10 | Validation of *cis*-regulatory motif requirements for enhancer activity.** **a**–**m**, Wild-type (WT; top cartoon) and motif-mutated (mutant; bottom cartoon) enhancers were tested by the transcriptional reporter gene assay used here. Shown are the representative embryos for WT (top) and mutant (bottom) enhancers. The activity of 10 out of 11 enhancers (**a**–**l**) was abolished or strongly reduced as we quantified directly from the strengths of the *in situ* signal in the original microscopy images. Histogram (middle) shows the distribution of *in situ* staining intensities in different parts of the embryo for WT (blue) and corresponding mutant (green) enhancer averaged over all independent embryos of the same genotype. Right, quantification of inferred enhancer strengths for WT (blue) and mutant (green) enhancers in transgenic reporter tests. Shown are the *in situ* staining levels for individual reporter-bearing embryos (dots), mean (horizontal lines) and standard deviations (*P* values according to Kolmogorov–Smirnov). The activity of four different broad CNS enhancers was dependent on presence of Trl (GAGA) motifs (**a**–**d**), the activity of three different midgut enhancers on the presence of GATA motifs (**e**–**g**), and the activity of three out of four tested early A–P enhancers was dependent on the presence of AGGAC (Ttk) motifs (**h**–**l**). **n**, position weight matrixes of the transcription factor motifs tested in **a**–**m**.