LETTER

doi:10.1038/nature13417

Enhancer loops appear stable during development and are associated with paused polymerase

Yad Ghavi-Helm[1], Felix A. Klein[1]*, Tibor Pakozdi[1]*, Lucia Ciglar[1], Daan Noordermeer[2], Wolfgang Huber[1] & Eileen E. M. Furlong[1]

Developmental enhancers initiate transcription and are fundamental to our understanding of developmental networks, evolution and disease. Despite their importance, the properties governing enhancer–promoter interactions and their dynamics during embryogenesis remain unclear. At the β-globin locus, enhancer–promoter interactions appear dynamic and cell-type specific[1,2], whereas at the *HoxD* locus they are stable and ubiquitous, being present in tissues where the target genes are not expressed[3,4]. The extent to which preformed enhancer–promoter conformations exist at other, more typical, loci and how transcription is eventually triggered is unclear. Here we generated a high-resolution map of enhancer three-dimensional contacts during *Drosophila* embryogenesis, covering two developmental stages and tissue contexts, at unprecedented resolution. Although local regulatory interactions are common, long-range interactions are highly prevalent within the compact *Drosophila* genome. Each enhancer contacts multiple enhancers, and promoters with similar expression, suggesting a role in their co-regulation. Notably, most interactions appear unchanged between tissue context and across development, arising before gene activation, and are frequently associated with paused RNA polymerase. Our results indicate that the general topology governing enhancer contacts is conserved from flies to humans and suggest that transcription initiates from preformed enhancer–promoter loops through release of paused polymerase.

Drosophila embryogenesis proceeds very rapidly, taking 18 h from egg lay to completion. Underlying this dynamic developmental program are marked changes in transcription, which are in turn regulated by characterized changes in enhancer activity. However, the role and extent of dynamic enhancer looping during this process remains unknown. To address this, we performed 4C-seq (chromosome conformation capture sequencing) experiments[5] anchored on 103 distal or promoter-proximal developmental enhancers (referred to as 'viewpoints'; Extended Data Fig. 1a), and constructed absolute and differential interaction maps for each, varying two important parameters: (1) developmental time, using embryos at two different stages, early in development when cells are multipotent (3–4 h after egg lay; stages 6–7), and mid-embryogenesis during cell-fate specification (6–8 h; stages 10–11); and (2) tissue context, comparing enhancer interactions in mesodermal cells versus whole embryo. To perform cell-type-specific 4C-seq in embryos, we established a modified version of BiTS-ChIP (batch isolation of tissue-specific chromatin for immunoprecipitation)[6]. Nuclei from covalently crosslinked transgenic embryos, expressing a nuclear-tagged protein only in mesodermal cells, were isolated by fluorescence-activated cell sorting (FACS; (>98% purity) and used for 4C-seq on 92 enhancers at 6–8 h and a subset of 14 enhancers at 3–4 h. The same 92 enhancers, and 11 additional regions, were also used as viewpoints in whole embryos at both time points (Extended Data Fig. 1b and Supplementary Table 1). The enhancers were selected based on dynamic changes in mesodermal transcription factor occupancy between these developmental stages[7,8] and the expression of the closest gene[9]. We thereby primed this study to detect dynamic three-dimensional (3D) interactions, focusing on developmental stages during which the embryo undergoes marked morphological and transcriptional changes.

All 4C-seq experiments had the expected signal distribution[5], with high concordance between replicates (median Spearman correlation 0.93). To assess data quality further, we examined ten known enhancer–promoter pairs (of the *ap*, *Abd-b*, *E2f*, *pdm2*, *Con*, *eya*, *stumps*, *Mef2*, *sli* and *slp1* genes), and in all cases recovered the expected interactions (Fig. 1 and Extended Data Fig. 1c–l). For example, using an enhancer of the *apterous* (*ap*) gene[10], we detect the expected interaction with the *ap* promoter, 17 kilobases (kb) away (Fig. 1), illustrating the high quality and resolution of the data.

In chromosome conformation capture assays, interaction frequencies decrease with genomic distance between regions[5]. To adjust for this, we modelled the 4C signal decay as a function of distance using a monotonously decreasing smooth function[11] (Extended Data Fig. 1b). Subtracting this trend, the residual interaction signal was converted to *z*-scores and interacting regions defined by merging neighbouring high-scoring fragments within 1 kb. Using this stringent approach, 4,247 high-confidence interactions were identified across all viewpoints and conditions, representing 1,036 unique interacting regions (Supplementary Table 2).

Each enhancer (viewpoint) interacted with, on average, ten distinct genomic regions (Extended Data Fig. 2a), less than half (41%) of which were annotated enhancers or promoters. Distal enhancers had a higher than expected interaction frequency with other enhancers (Extended Data Fig. 2b, $P = 2.4 \times 10^{-3}$). Similarly, promoter-proximal elements

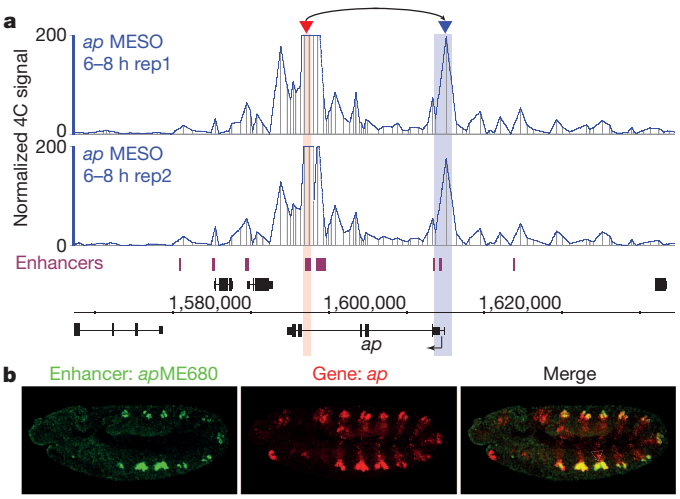


Figure 1 | A high-resolution view of enhancer interactions during *Drosophila* embryogenesis. **a**, 4C interaction map (viewpoint, red arrowhead) at the *ap* locus. The expected interaction with the promoter (blue arrowhead) of *ap* is observed. Known enhancers are indicated. **b**, Expression (*in situ* hybridization) of the *ap* gene (red) and expression driven by its interacting enhancer (*GFP*, green) at stage 11. MESO, mesoderm.

[1]European Molecular Biology Laboratory, Genome Biology Unit, D-69117 Heidelberg, Germany. [2]Swiss Federal Institute of Technology, School of Life Sciences, CH-1015 Lausanne, Switzerland.
*These authors contributed equally to this work.

had extensive interactions with distal active promoters, 98% of which are >10 kb away (Extended Data Fig. 2b, c, $P = 6 \times 10^{-4}$). Enhancer–promoter interactions, although not significantly enriched, involve active promoters, with high enrichment for H3K27ac and H3K4me3, and active enhancers, defined by H3K27ac, RNA Pol II and H3K79me3 (ref. 6) (Extended Data Fig. 2d, e). In contrast, contacts at inactive promoters are significantly depleted (Extended Data Fig. 2b). These results are similar to recent findings in human cells[12,13] and the mouse β-globin locus[1,2], indicating similarities in 3D regulatory principles from flies to humans.

The extent of 3D connectivity is surprising given the relative simplicity of the *Drosophila* genome. On average, each promoter-proximal element interacted with four distal promoters and two annotated enhancers, whereas each distal enhancer interacted with two promoters and three other enhancers. These numbers are probably underestimates, as 60% of interactions involved intragenic or intergenic fragments containing no annotated *cis*-regulatory elements. Despite this, the level of connectivity is similar to that recently observed in humans, where active promoters contacted on average 4.75 enhancers and 25% of enhancers interacted with two or more promoters[13]. The multi-component contacts that we observe for *Drosophila* enhancers indicate topologically complex structures and suggest that, despite its non-coding genome being an order of magnitude smaller than humans, *Drosophila* may require a similar 3D spatial organization to ensure functionality.

Insulators, and associated proteins, are thought to have a major role in shaping nuclear architecture by anchoring enhancer–promoter interactions or by acting as boundary elements between topologically associated domains (TADs)[14–16]. Occupancy data from 0 to 12 h *Drosophila* embryos[17] revealed a 50% overlap of interacting regions with occupancy of one or more insulator protein. Insulator-bound interactions are enriched in enhancer elements, suggesting that insulators may have a role in promoting enhancer–enhancer interactions (Extended Data Fig. 3a–d). In contrast to mammalian cells[16], we observed no association between insulator occupancy and the genomic distance spanned by chromatin loops, although there was a modest increase in average interaction strength (Extended Data Fig. 3e, f). Conversely, 50% of interacting regions are not bound by any of the six *Drosophila* insulator proteins (Extended Data Fig. 3a, g), suggesting that these 3D contacts are formed in an insulator-independent manner, or are being facilitated by neighbouring interacting regions.

If enhancer 3D contacts are involved in transcriptional regulation, then genes linked by interactions with a common enhancer should share spatio-temporal expression, as recently reported[18–20]. For the four loci examined—*pdm2* (Extended Data Fig. 4a, b), *engrailed* (*en*; Extended Data Fig. 4c, d), *unc-5* (Extended Data Fig. 5c, d) and *charybde* (Fig. 2c, d, described below)—this is indeed the case. For example, the *pdm2* CE8012 enhancer interacts with both the *pdm2* and *nubbin* (*nub*, also known as *pdm1*) promoters, located 2.5 and 47 kb away, respectively. Both genes, producing structurally related proteins, are co-expressed in the ectoderm, overlapping the activity of the *pdm2* enhancer.

Although there are examples of long-range interactions in *Drosophila*, often involving Polycomb response elements (PREs)[15,21,22] and insulator elements[21], the vast majority of characterized enhancers are within 10 kb of their target gene, with few known to act over 50 kb (Fig. 2a and Supplementary Table 3). However, as investigators historically tested regions close to the gene of interest, characterized *Drosophila* enhancers are generally close to the gene they regulate. In contrast, although 4C cannot assess the full extent of short-range interactions (Extended Data Fig. 5a, b), it provides an unbiased systematic measurement of the distance of enhancer interactions, far beyond 10 kb.

The distance distribution of all significant interactions reveals extensive long-range interactions within the ~180 megabase (Mb) *Drosophila* genome; 73% span >50 kb, with the median interaction–viewpoint distance being 110 kb (Fig. 2a, b). Two striking examples of long-range interactions are the *unc-5* and *charybde* loci. The *unc-5* promoter interacts with multiple regions, including a weak but significant interaction

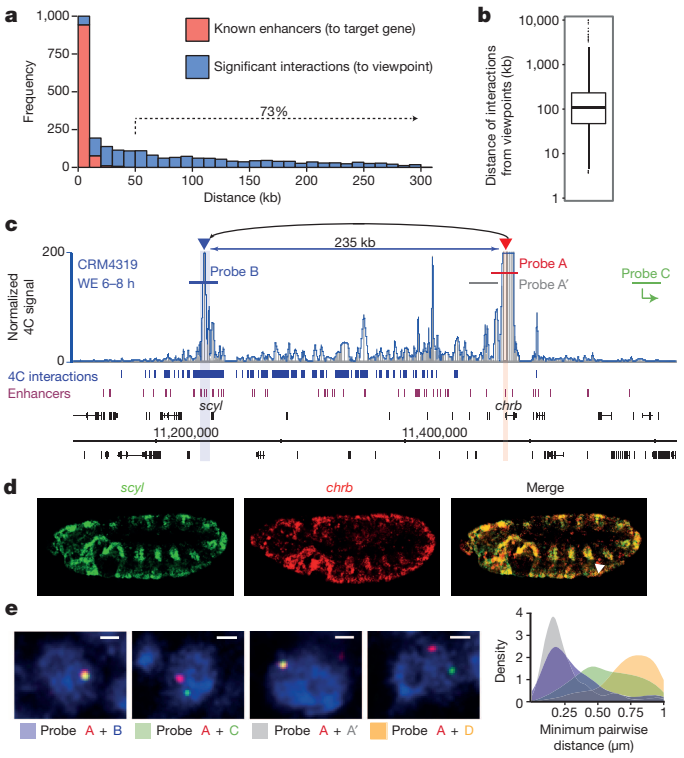


Figure 2 | Long-range interactions are widespread in the *Drosophila* genome. **a**, **b**, Distance distribution of all known *Drosophila* enhancers (red; literature based) and identified significant interactions (blue, $n = 1{,}983$) to their respective target gene or viewpoint, shown as a histogram (**a**) and box plot (**b**). **c**, 4C interaction map (viewpoint, red arrowhead) around the *scyl* and *chrb* loci. WE, whole embryo. Interaction with the *scyl* gene is highlighted (blue arrowhead). Significant 4C interactions, known enhancers and DNA FISH probes are indicated. **d**, Expression (*in situ* hybridization) of *chrb* (red) and *scyl* (green) at stage 11. **e**, DNA FISH images of representative nuclei, merging DAPI (blue), probe A (red) and probe B, C, A′ or D (green) channels. Scale bar, 1 μm. Density plot indicates measured distances between probe A and probe A′ (grey), D (beige) B (blue) or C (green).

with the promoter of *slit* (*sli*), at a distance of >500 kb (Extended Data Fig. 5c, d). These genes produce structurally unrelated proteins that are co-expressed in the heart, and are essential for heart formation.

A promoter-proximal element near the *charybde* (*chrb*) promoter has a strong interaction with the promoter of the *scylla* (*scyl*) gene, almost 250 kb away (Fig. 2c). Both genes are closely related in sequence and co-expressed throughout embryogenesis (Fig. 2d)[23]. These long-range interactions were confirmed by reciprocal 4C, using either the promoter of *chrb* or *scyl*, or an interacting putative enhancer as viewpoint (Extended Data Fig. 5e). We further verified this interaction using DNA fluorescence *in situ* hybridization (FISH) in embryos (Fig. 2e). As a control, we assessed the distance between the *chrb* promoter (probe A) and an overlapping probe A′ or a region on another chromosome (probe D), to determine the distances between regions very close or far away, respectively. Comparing the distance between the *chrb* and *scyl* promoters (probes A and B, Fig. 2c) showed a high, statistically significant co-localization (Fig. 2e; 37% co-localization; $P < 10^{-18}$; Extended Data Fig. 5f), in contrast to the distance between the *chrb* promoter and a non-interacting region with equal genomic distance (probes A and C; 5% co-localization).

The reciprocal 4C revealed several intervening interactions that are consistently associated with loops to both the *scyl* and *chrb* promoter. We examined the activity of two of these in transgenic embryos. Both interacting regions can function as enhancers *in vivo*, recapitulating *chrb* expression in the visceral mesoderm (enhancer 1) and nervous system (enhancer 2) (Extended Data Fig. 5e, g).

When considering a 1-Mb scale around this region, the 4C interaction signal drops to almost zero just after the promoters of both genes (Extended Data Fig. 6a). This 'contained block' of interactions is reminiscent of TADs[14], although the boundaries don't exactly match TADs defined at late stages of embryogenesis[15], which may reflect differences in the developmental stages used. However, the boundaries do overlap a block of conserved microsynteny between drosophilids[24] spanning ~50 million years of evolution (Extended Data Fig. 6a), suggesting a functional explanation underlying the maintained synteny. Expanding this analysis across all viewpoints, ~60% of interactions are located within the same TAD and the same microsyntenic domain as the viewpoint (Extended Data Fig. 6b, c). In the case of the *chrb* and *scyl* genes, this constraint may act to maintain a regulatory association between a large array of enhancers, facilitating their interaction with both genes' promoters.

These examples, and the other 555 unique interactions >100 kb, provide strong evidence that long-range interactions are widely used within the *Drosophila* genome, potentially markedly increasing the regulatory repertoire of each gene.

As enhancer–promoter looping can trigger gene expression[25], it follows that enhancer contacts should reflect the dynamics of transcriptional changes during development and therefore be temporally associated with gene expression. To assess this, we directly compared looping interactions between the two different time points and tissue contexts. Given the non-discrete nature of chromatin contacts, we used the quantitative 4C-seq signal to identify differential interactions based on a Gamma-Poisson model and defined them as having >2-fold change and false discovery rate ≤10%.

Despite the marked differences in development and enhancer activity between these conditions, we found surprisingly few changes in chromatin interaction frequencies, with ~6% of interacting fragments showing significant changes between conditions (Extended Data Fig. 7; Fig. 3a and Extended Data Fig. 8a, red dots). Of these, 87 interactions were significantly reduced during mid-embryogenesis (6–8 h) compared to the early time point (3–4 h), and 90 interactions significantly increased. Similarly, 105 interactions had a higher frequency in mesodermal cells, compared to the whole embryo, and 34 interactions were lower.

For example, a promoter-proximal viewpoint in the vicinity of the *Antp* promoter identified many interactions, two of which are significantly decreased at 6–8 h, although the expression of the *Antp* gene itself increases (Extended Data Fig. 8b). For one region, the reduction in 4C interaction at 6–8 h corresponds to a loss in a H3K4me3 peak from 3–4 h to 6–8 h (asterisk), suggesting that this 3D contact is associated with the transient expression of an unannotated transcript. We examined the activity of the other interacting peak in transgenic embryos and showed that it acts as an enhancer, driving specific expression in the nervous system overlapping the *Antp* gene at 6–8 h (Extended Data Fig. 8c). Along with the two enhancers discovered at the *chrb* locus, this demonstrates the value of 3D interactions to identify new enhancer elements, even for well-characterized loci like *Antp*.

A viewpoint in the vicinity of the *Abd-B* promoter interacted with a number of regions spanning the bithorax locus, three of which correspond to previously characterized *Abd-B* enhancers; iab-5 (ref. 26), iab-7 and iab-8 (refs 26, 27) (Fig. 3b, c). The iab-7 and iab-8 enhancers are active in early embryogenesis, and have much reduced or no activity at the later time point[26,27]. Notably, although the loop to those two enhancers is strong at the early time point, it becomes significantly reduced later in development, when both enhancers' activities are reduced. Conversely, the iab-5 enhancer contacts the promoter at a much higher frequency later in development, at the stage when the enhancer is most active[26,27]. This locus therefore exhibits dynamic 3D promoter–enhancer contacts that reflect the transient activity of three developmental enhancers. It is interesting to note that in all loci examined, the dynamic contacts of specific elements are neighboured by stable contacts, as seen

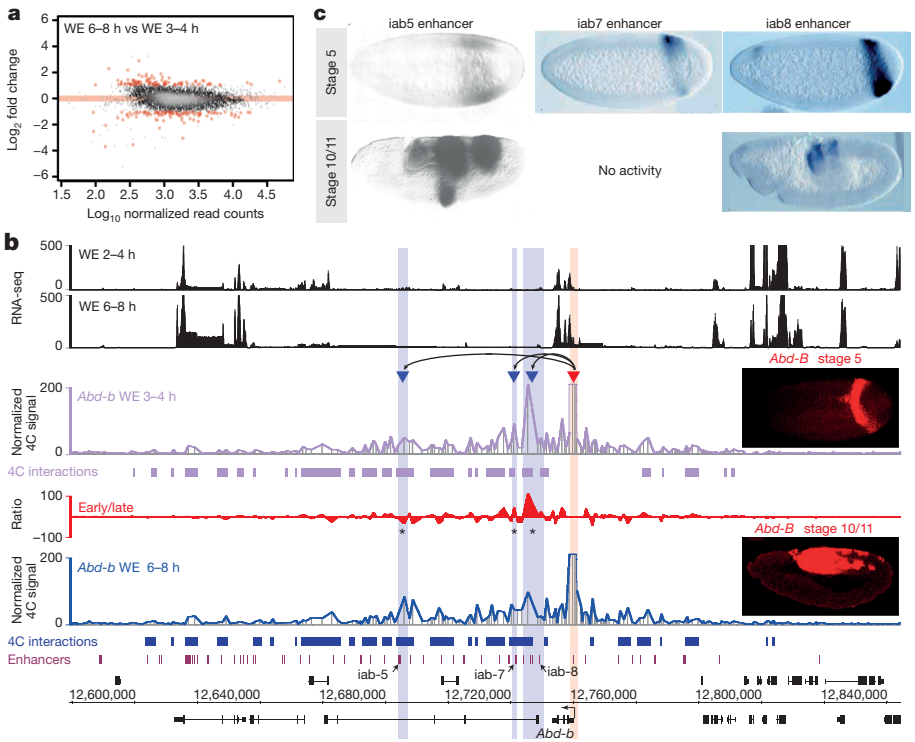


Figure 3 | Specific loci display localized differential interactions. **a**, MA plot of interaction signal between whole embryo 6–8 h and whole embryo 3–4 h (significant differential interactions, red dots). WE, whole embryo. **b**, 4C interaction map at the *Abd-b* locus. Top to bottom: RNA-seq signal (reads per kilobase per million mapped reads (RPKM), black) in whole embryo at 2–4 h and 6–8 h (ref. 9), 4C interaction map (viewpoint, red arrowhead) in whole embryo at 3–4 h (mauve) and 6–8 h (blue), and differential 4C signal (red) with significant differential 4C interactions (asterisk). Insets show the expression (*in situ* hybridization) of *Abd-B* (red) at stages 5 (2–4 h) and 11 (6–8 h). **c**, Expression (*in situ* hybridization) driven by iab5 (ref. 26), iab7 and iab8 (ref. 27) enhancers at stages 5 and 10–11. Embryo images in panel **c** reproduced with permission from: ref. 27, *Development*; ref. 26, Nature Publishing Group.

in the *Antp* and *Abd-B* loci. Dynamic changes, therefore, appear to operate in the context of larger, more-stable 3D landscapes.

Ninety-four per cent of enhancer interactions showed no evidence of dynamic changes across time and tissue context, which is remarkable given the marked developmental transitions during these stages (Fig. 3a, Extended Data Fig. 8a and Supplementary Table 4). To investigate this further, we examined enhancer–promoter interactions of genes switching their expression state between time points or tissue contexts. The *ap* gene, for example, is not expressed at 2–4 h but is highly expressed during mid-embryogenesis (6–8 h) (Fig. 4a). Despite the absence of expression, the interaction between the *ap*ME680 enhancer and the *ap* promoter is already present at 3–4 h, several hours before the gene's activation (Fig. 4a). To examine this more globally, we selected differentially expressed genes, going either from on-to-off or off-to-on (Extended Data Fig. 9). Even for these dynamically expressed genes, there was no correlation with changes in their promoter–enhancer contacts (Fig. 4b). We observe similar 'stable' interactions between tissue contexts. Genes predominantly expressed in the neuroectoderm at 6–8 h, for example, have interactions at the same locations in whole embryos and purified mesodermal nuclei at 6–8 h, despite the fact that they are not expressed in the mesoderm at this stage (Extended Data Fig. 8d–g).

Pre-existing loops were recently observed in human and mouse cells, and suggested to prime a locus for transcriptional activation[4,13]. However, why they are formed and how transcription is eventually triggered remains unclear. To investigate this, we focused on the subset of genes that have both off-to-on expression and no evidence for differential interactions (20 genes; differentially expressed with stable loops (DS) genes; Supplementary Table 5 and Extended Data Fig. 9). Despite changes in their overall expression, DS genes have similar levels of RNA polymerase II (Pol II) promoter occupancy at both time points (Fig. 4c). The presence of promoter-bound Pol II in the absence of full-length transcription is indicative of Pol II pausing. Using global run-on sequencing (GRO-seq) data[28] to define a stringent set of paused genes, we observed that most (75%) DS genes are paused (15 of 20 DS genes; Fig. 4d and Extended Data Fig. 9b, d), and have a significantly higher pausing index (Fig. 4e). This percentage is significantly higher than expected by chance when sampling over all off-to-on genes (Fig. 4d), and is robust to using a strict (Fig. 4d) or more relaxed (Extended Data Fig. 9e) definition of Pol II pausing[28]. This association is very evident when examining specific loci (Fig. 4a and Extended Data Fig. 10), showing Pol II occupancy, short abortive transcripts, and loop formation before the gene's expression. Taken together, these results indicate that 'stable' chromatin loops are associated with the presence of paused Pol II at the promoter.

To understand how transcription is ultimately activated, we examined changes in DNase I hypersensitivity[29] at the promoter of DS genes. DNase I hypersensitivity is significantly increased at interacting promoters at the stages when the gene is expressed (Fig. 4f), suggesting that the recruitment of additional transcription factor(s) later in development might act as the trigger for transcriptional activation.

In summary, our data reveals extensive long-range interactions in an organism with a relatively compact genome, including pairs of co-regulated genes contacting common enhancers often at distances greater than 200 kb. Comparing enhancer contacts in different contexts revealed that chromatin interactions are very similar across developmental time points and tissue contexts. Enhancers therefore do not appear to undergo long-range looping *de novo* at the time of gene expression, but are rather already in close proximity to the promoter they will regulate. Within this 3D topology, highly dynamic and transient contacts would not be visible when averaging over millions of nuclei. As transcription factor

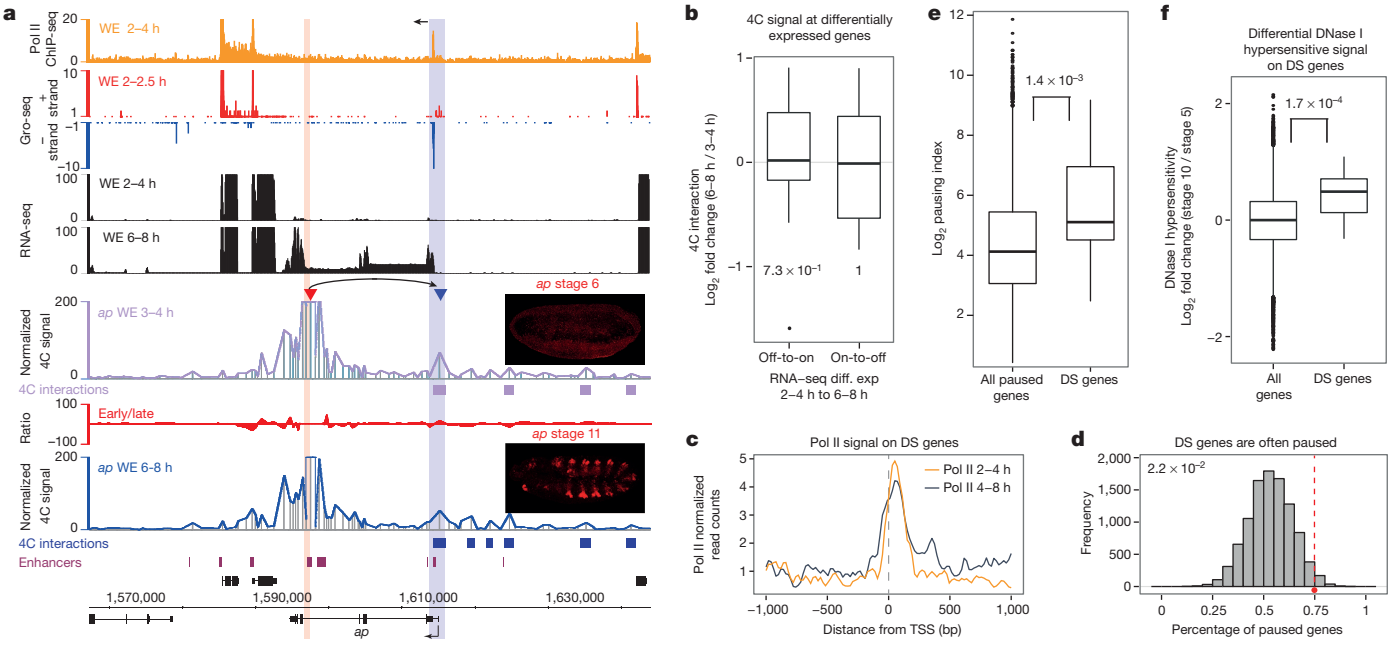


Figure 4 | Interactions are stable across developmental time and associated with paused polymerase. **a**, 4C interaction map at the *ap* locus. Top to bottom: Pol II signal (reads per genomic content (RPGC)) in whole embryo at 2–4 h (orange), GRO-seq signal in whole embryo at 2–2.5 h (plus strand, red; minus strand, blue)[28], RNA-seq signal (RPKM) in whole embryo at 2–4 h and 6–8 h (black)[9], 4C interaction map (viewpoint, red arrowhead) in whole embryo at 3–4 h (mauve) and 6–8 h (blue), and differential 4C signal (red). WE, whole embryo. Significant 4C interactions and known enhancers are indicated. Insets show the expression (*in situ* hybridization) of *ap* at stages 6 (3–4 h) and 11 (6–8 h). **b**, Differentially expressed genes, going from off-to-on ($n = 21$) or on-to-off ($n = 8$), have no significant differences in the frequency of 4C interactions at their promoter (two-sided Wilcoxon test). **c**, Pol II signal (RPGC) is enriched at the promoter of differential genes with stable interactions (DS genes) at 2–4 h, even though the genes are not expressed. **d**, Expected distribution of paused genes (using top 50% of paused genes[28]) by sampling 10,000 times differential off-to-on genes (red dotted line, observed percentage of paused DS genes). *P* value is indicated. **e**, $\log_2$ pausing index in whole embryo at 2–2.5 h (ref. 28) of DS genes ($n = 19$) is significantly different from all paused genes[28] ($n = 7{,}734$; two-sided Wilcoxon test). **f**, Box plot showing differential DNase I hypersensitivity ($\log_2$ fold change stage 10/stage 5) at the promoter of DS genes ($n = 18$) and all *Drosophila* mRNA genes ($n = 10{,}409$, see Methods). *P* value from a two-sided Wilcoxon test. Boxes depict the interquartile range (IQR) with the median as a horizontal thick line. Upper and lower whiskers extend to 1.5 times the IQR, and points represent outliers.

binding is sufficient to force loop formation[25], our results suggest a model where through transcription factor–enhancer occupancy, an enhancer loops towards the promoter and polymerase is recruited, but paused in the majority of cases. The subsequent recruitment of transcription factor(s) or additional enhancers at preformed 3D hubs most likely triggers activation by releasing Pol II pausing. Such preformed topologies could thereby promote rapid activation of transcription[3,4,13]. At the same time, as paused promoters can exert enhancer-blocking activity[30], the presence of paused polymerase within these 3D landscapes could safeguard against premature transcriptional activation, but yet keep the system poised for activation.

METHODS SUMMARY

Staged *Drosophila* embryos were collected at 3–4 h or 6–8 h after egg lay and fixed in 1.8% formaldehyde for 15 min at room temperature. Thirty million nuclei were used for each 4C template preparation, enough for on average ten viewpoints. Libraries were amplified from 320 ng of 4C template (primer sequence in Supplementary Table 6), and 100 multiplexed libraries were sequenced over on average five HiSeq2000 lanes, using 100-base-pair (bp) single-end reads. Two independent biological replicates were analysed for each condition.

Online Content Methods, along with any additional 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 18 December 2013; accepted 29 April 2014.
Published online 2 July 2014.

1. Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active β-globin locus. *Mol. Cell* **10,** 1453–1465 (2002).
2. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). *Nature Genet.* **38,** 1348–1354 (2006).
3. Montavon, T. *et al.* A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147,** 1132–1145 (2011).
4. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502,** 499–506 (2013).
5. Van de Werken, H. J. G. *et al.* 4C technology: protocols and data analysis. *Methods Enzymol.* **513,** 89–112 (2012).
6. 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).
7. Junion, G. *et al.* A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* **148,** 473–486 (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. Graveley, B. R. *et al.* The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471,** 473–479 (2011).
10. Capovilla, M., Kambris, Z. & Botas, J. Direct regulation of the muscle-identity gene apterous by a Hox protein in the somatic mesoderm. *Development* **128,** 1221–1230 (2001).
11. Tolhuis, B. *et al.* Interactions among Polycomb domains are guided by chromosome architecture. *PLoS Genet.* **7,** e1001343 (2011).
12. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489,** 109–113 (2012).
13. Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503,** 290–294 (2013).
14. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
15. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148,** 458–472 (2012).
16. Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153,** 1281–1295 (2013).
17. Nègre, N. *et al.* A comprehensive map of insulator elements for the *Drosophila* genome. *PLoS Genet.* **6,** e1000814 (2010).
18. Schoenfelder, S. *et al.* Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature Genet.* **42,** 53–61 (2010).
19. Fanucchi, S., Shibayama, Y., Burd, S., Weinberg, M. S. & Mhlanga, M. M. Chromosomal contact permits transcription between coregulated genes. *Cell* **155,** 606–620 (2013).
20. Zhang, Y. *et al.* Chromatin connectivity maps reveal dynamic promoter–enhancer long-range associations. *Nature* **504,** 306–310 (2013).
21. Cléard, F., Moshkin, Y., Karch, F. & Maeda, R. K. Probing long-distance regulatory interactions in the *Drosophila melanogaster* bithorax complex using Dam identification. *Nature Genet.* **38,** 931–935 (2006).
22. Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F. & Orlando, V. Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nature Cell Biol.* **9,** 1167–1174 (2007).
23. Scuderi, A., Simin, K., Kazuko, S. G., Metherall, J. E. & Letsou, A. *scylla* and *charybde*, homologues of the human apoptotic gene *RTP801*, are required for head involution in *Drosophila*. *Dev. Biol.* **291,** 110–122 (2006).
24. Engström, P. G., Sui, S. J. H., Drivenes, Ø., Becker, T. S. & Lenhard, B. Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Res.* **17,** 1898–1908 (2007).
25. Deng, W. *et al.* Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149,** 1233–1244 (2012).
26. Busturia, A. & Bienz, M. Silencers in abdominal-B, a homeotic *Drosophila* gene. *EMBO J.* **12,** 1415–1425 (1993).
27. Zhou, J., Ashe, H., Burks, C. & Levine, M. Characterization of the transvection mediating region of the abdominal-B locus in *Drosophila*. *Development* **126,** 3057–3065 (1999).
28. Saunders, A., Core, L. J., Sutcliffe, C., Lis, J. T. & Ashe, H. L. Extensive polymerase pausing during *Drosophila* axis patterning enables high-level and pliable transcription. *Genes Dev.* **27,** 1146–1158 (2013).
29. Thomas, S. *et al.* Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. *Genome Biol.* **12,** R43 (2011).
30. Chopra, V. S., Cande, J., Hong, J.-W. & Levine, M. Stalled Hox promoters as chromosomal boundaries. *Genes Dev.* **23,** 1505–1509 (2009).

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

Acknowledgements This work was technically supported by the EMBL Genomics Core and FACS core facilities. We thank all members of the Furlong laboratory for discussions and comments, in particular D. Garfield, J. Reddington and I. Schor for important suggestions. Embryo images in Fig. 3c were used with permission from *Development*[27] and Nature Publishing Group[26] as these fly strains no longer exist. This work was supported by a DFG (FU 750) grant to E.E.M.F., an EMBO post-doctoral fellowship to Y.G.-H., and the EC FP7 project 'Radiant' grant to F.A.K. and W.H.

Author Contributions Y.G.-H. and E.E.M.F. designed the study, analysed the results and wrote the manuscript. Y.G.-H. performed 4C experiments, DNA *in situ* hybridization and imaging and performed data analysis. F.A.K., T.P. and W.H. developed and performed 4C-seq bioinformatics analysis. L.C. generated all transgenic strains and performed 4C-PCR reactions, RNA *in situ* hybridizations and imaging. D.N. was involved in 4C primer design. T.P. and F.A.K. contributed equally to the study. All authors discussed the results and commented on the manuscript.

Author Information All raw data, which consists of 2,587 demultiplexed files, have been submitted to the EBI European Nucleotide Archive and ArrayExpress databases, accession numbers ERP004524 and E-MTAB-2180, respectively. To enable the community to browse through the data, 4C-seq interaction data is available in a customized web browser at http://furlonglab.embl.de/4CBrowser. 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 E.E.M.F. (furlong@embl.de).

METHODS

4C-seq on *Drosophila* embryos. 4C templates were generated as previously described with minor modifications[31]. *Drosophila* embryos (*twi*^{PEMK}::*SBP-His2B Drosophila* strain[32]) collected (after 3 pre-lays) at 3–4 h (stage 6–7) or 6–8 h (stage 10–11) after egg lay[33] were covalently crosslinked in 1.8% formaldehyde for 15 min at room temperature. Nuclear extraction and sorting was carried as described previously[34]. 30 million nuclei were used for each 4C template preparation (on average enough for 10 viewpoints) using DpnII and NlaIII as the first and second restriction enzyme, respectively. Libraries were amplified from 320 ng of 4C template (primer sequence in Supplementary Table 6) using touchdown PCR (94 °C 2 min; 8 cycles of 94 °C 15 s, 55 °C 1 min, 68 °C 5 s; 18 cycles of 94 °C 15 s, 63 °C 1 min, 63 °C 5 s + 10 s per cycle; 68 °C 10 s; hold at 22 °C). 100 multiplexed libraries were sequenced over on average five HiSeq2000 lanes using 100-bp single-end reads. Two independent biological replicates were performed for each condition. To ensure enough coverage, each sample has on average >5 million reads[35] (Supplementary Table 1).

Two-colour 3D-DNA FISH (fluorescent *in situ* hybridization). Two-colour DNA FISH was performed as previously described[36], on dechorionated overnight embryos fixed for 20 min with 4% formaldehyde. DNA FISH probes were constructed from BAC clones 67I11 (A), 99B18 (A′), 15O19 (B), 66C01 (C) and 141G06 (D) (P[acman][37]). FISH probes were labelled using the FISH Tag DNA Multicolor kit (Alexa Fluor 488 dye and Alexa Fluor 594 dye) (Life Technologies). The slides were mounted using ProLong Gold antifade reagent with DAPI (Life Technologies) and imaged on a Zeiss LSM 780 confocal microscope. 3D stacks were collected, and relative distances between FISH signals were analysed in approximately 200 to 800 nuclei using Imaris software (Bitplane). Two probes were considered co-localized when the distance between the centres of FISH signal was below 0.2 μm.

Transgenic reporter assays. To assay 4C interacting regions for enhancer activity, appropriate genomic regions were cloned in front of a minimal *Hsp70* promoter driving a *lacZ* reporter gene in the reporter vector pH-Pelican[38]. The coordinates of the cloned regions are as follows: *chrb* enhancer 1: chr3L, 11,421,058–11,423,149; *chrb* enhancer 2: chr3L, 11,505,174–11,506,253; new *Antp* enhancer: chr3R, 2,743,008–2,748,681. All constructs were injected according to standard methods[39] into the *w1118* line. Stably integrated transgenic lines were balanced, homozygosed and used for embryo collections, and subjected to *in situ* hybridization to examine *lacZ* expression.

Double fluorescent RNA *in situ* hybridization. Fluorescent RNA *in situ* hybridization was performed as described previously[40]. The following ESTs were used to generate antisense digoxigenin-, fluorescein-, or biotin-labelled probes: LD16125 (*en*), LD19406 (*unc-5*), GH13089 (*E2f*), LD32080 (*chrb*), LD22812 (*scy*), LD33666 (*Antp*), SD05618 (*ap*), RE34565 (*pdm2*), RE34782 (*nub*), RE47096 (*Abd-B*), AT24588 (*inv*), IP16087 (*sli*), LP21457 (*stumps*). Enhancer activity was detected using a probe against a reporter gene (*lacZ*, *GFP* or *GAL4*) cloned downstream of the enhancer and a minimal promoter. The probes were detected using peroxidase-conjugated antibodies (Roche, Invitrogen) and developed using the TSA fluorescence system (Perkin Elmer). Embryos were imaged with a Zeiss LSM 510 META microscope.

4C-seq data analysis from raw reads to significant peaks. FASTQ files representing the sequencing libraries were demultiplexed using the first 12 bases of the viewpoint primers. The Hamming distance between primers was ≥3. During demultiplexing, the primer sequence was trimmed, while the restriction enzyme cutting site was kept. Reads were aligned to the *Drosophila* dm3 genome using Novoalign with the default parameters. For short fragments, some reads contained the religation site and additional sequence of the ligated restriction fragment. Therefore unaligned reads were scanned for the restriction enzyme cutting sequence of the second restriction enzyme. When such a site was found the read was trimmed after this site. The trimmed reads were realigned, and the combined set of alignments was analysed further.

We generated a '4C reference genome' based on the recognition sequences of the DpnII (GATC) restriction enzyme within the *Drosophila* genome. Fragments of the first cutter that did not contain a cutting site for the second restriction enzyme were removed. We also removed fragments if both fragment ends were shorter than 20 nt. The aligned reads were mapped to the 4C reference genome and only reads starting exactly at a restriction fragment end with the correct orientation were assigned to a fragment end. The counts for both fragment ends were summed into one count value per fragment. The percentage of undigested (or self-ligated) 4C products was estimated by counting the number of reads that fell within the restriction fragment adjacent to the viewpoint in the direction of the sequencing primer. The obtained counts were divided by the total number of reads for each library. The median percentage of undigested (or self-ligated) reads was 7.2% over all the libraries.

For each viewpoint, count values >40 on the viewpoint's chromosome were transformed using a variance stabilizing transformation implemented in the R package *DESeq2*. On the transformed count values, the decrease in 4C signal with genomic distance from the viewpoint was fitted using a smooth monotonous local regression with the R package *fda*. The fragments directly adjacent to the viewpoint fragment, which typically had very strong signals, were excluded from the fit up to a distance where a local minimum in the 4C signal was observed. The resolution of our 4C experiments around the viewpoint is illustrated in Extended Data Fig. 5a, b. The median distance to the viewpoint of the first fragment included in the fit was 4.3 kb. The 'first identified interaction' category (Extended Data Fig. 5b) represents the distance of the first valid DpnII fragment to each viewpoint, and had a median distance of 13.8 kb. The residuals from the fit were used to calculate *z*-scores for each fragment. Nominal *P* values were calculated from the *z*-scores via the standard Normal distribution and adjusted for multiple testing using the method of Benjamini and Hochberg[41]. We also calculated the interaction fold change (referred to as IFC) between the fit and the observed (transformed) counts as an alternative quantitative measurement of the interaction strength. 99.67% of the high-confidence interactions had an IFC greater than 2, and 95.60% had an IFC greater than 3.

Fragments were selected that had a *z*-score >3 (corresponding to a nominal *P* value of 0.001) in both biological replicates, and FDR-adjusted *P* value <0.1 in at least one replicate. Selected fragments within 1 kb of each other were merged into regions, which were considered significantly interacting. The total number of reads, filtered reads and interactions for each viewpoint in each condition is summarized in Supplementary Table 1.

To identify significant differences in interaction frequencies between different developmental stages and conditions, we used a test based on the Gamma-Poisson model, following the Differential Binding Analysis approach established for comparative ChIP-seq analysis[42]. Library normalization factors were first calculated using the fitted values. For comparison between different conditions for a single viewpoint (used for both visualizations and differential analysis), a reference normalization fit curve was constructed by taking the fragment-based median of all condition-specific fit curves per viewpoint. The individual fragment counts were scaled by the ratio of condition-specific regression fits to the reference curve. The R package *DESeq2* was used to find differentially interacting fragments, defined as fragments with an absolute $\log_2$ fold change greater than 1 and an FDR ≤ 10% (Fig. 3a and Extended Data Fig. 8a).

Correlations to active promoters and enhancers (associated with Extended Data Fig. 2). The viewpoints used in this study were defined as 'distal enhancers' when they were: (1) located >1 kb from the closest transcriptional start site (TSS) and (2) devoid of H3K4me3 signal to avoid unannotated TSS, and 'promoter-proximal' viewpoints otherwise (Extended Data Fig 1a). This classification ensures that interactions with 'distal-enhancer' viewpoints were truly involving an enhancer, while interactions with 'promoter-proximal' viewpoints could involve either the promoter-proximal element and/or the nearby promoter (within 1 kb), given the resolution of 4C experiments. Note, we therefore only examined interactions between promoter-proximal viewpoints and distal (that is, >1 kb away) promoter and enhancer elements, which are generally >10 kb away (Extended Data Fig. 5b).

The feature distribution of interacting regions in Extended Data Fig. 2b was based on unique 4C interactions (to observe the feature distribution in a condition-unspecific way, we generated a non-redundant set of interactions for each viewpoint across all conditions) and examined by dividing the genome into five categories: (1) enhancers, formed by enhancers tested in transgenic embryos (CAD2 (ref. 6)) and ChIP-defined *cis*-regulatory modules (CRMs) based on transcription factor occupancy (TF8008 (ref. 8) and Cardiac[7]), which have been filtered for TSS proximity (1 kb), redundancy (with the hierarchy CAD2> TF8008> Cardiac), and potential unannotated TSS using H3K4me3 from mesodermal cells at 6-8h[6] (based on H3-subtracted signal using a cut-off estimated using linear discriminant analysis trained on a set of active and inactive genes, as described previously[6]); (2) active or (3) non-active promoters (±1 kb around each annotated TSS obtained from FlyBase v5.47 (ref. 43)); (4) intragenic regions; and (5) intergenic regions that do not contain an annotated enhancer as defined above. Promoters were annotated as active or non-active based on the RPKM values from RNA-seq at either 2–4 h or 6–8 h of embryo development[9], with a threshold selected to define a gene as active based on the local minima in the log-RPKM distribution[44], as illustrated in Extended Data Fig. 9a. The observed distribution of feature overlaps shown in Extended Data Fig. 2b were compared to a background set of interactions (or expectation regions). This background was based on random sampling of valid DpnII fragments throughout the genome, which were matched for mappability, G+C content and width to the observed set of interactions using the MatchIt algorithm (http://imai.princeton.edu/research/files/matchit.pdf) with Mahalanobis distance. In this way we constructed an unbiased set of regions that share similar properties to the 4C interacting regions, to allow for a fair comparison without prior knowledge of epigenetic patterning or expression state. This set of background regions was then associated

with particular features of the genome (active promoter, non-active promoter, enhancer) using the same rules as the observed interactions.

For Extended Data Fig. 2c–e, profiles of histone modifications (H3-subtracted) and Pol II binding (input-subtracted) from 6–8 h ChIP-seq performed in the mesoderm[6] were summarized in a 2-kb window around each TSS or enhancer midpoint on the observed interactions for whole embryo 6–8 h and the same set of background interactions as described above. The thick lines in Extended Data Fig. 2c–e represents a 10% trimmed mean, and confidence intervals have been interpolated from bootstrap estimation.

Association of 4C interacting region with insulator binding (associated with Extended Data Fig. 3). Each significant interaction from the whole embryo 6–8 h condition was examined for overlap with insulator binding, using the seven available insulators data sets from modENCODE ChIP-chip experiments (E0–12 h)[17] (six different insulator proteins, with two ChIP experiments for CTCF using N- or C-terminal-specific antibodies), based on a 1% FDR cut-off for ChIP peak calling[45] and at least a 1-bp overlap between the defined regions. Insulator groups were constructed based on the observed co-overlap on 4C interactions, and only insulator groups associated with 20 or more interactions were considered for further analyses. In enrichment tests, each insulator group was compared to a constant set of non-insulator bound interactions (which represents 50.3% of all 4C unique interactions). *P* values were derived from Fisher's Exact Test, with a minimal statistical significance set to <10% FDR.

The distribution of distances over which enhancers interact (associated with Fig. 2). For each interacting region, the genomic distance was calculated from the midpoint of the significant region to the midpoint of the viewpoint region (in blue). The distance from all known *Drosophila* enhancers, using a comprehensive database based on literature information of enhancer activity from transgenic embryos (RedFly[46] and CAD2 (ref. 6)), to their experimentally linked target gene is shown in red.

Most 4C interacting regions fall within the same genomic block as the corresponding viewpoint (associated with Extended Data Fig. 6b, c). We used previously defined topological domains[15], and blocks of conserved microsynteny between *Drosophila melanogaster* and 4 other drosophilids[24], to determine if interacting regions were present within the same domain or block or if they were 'broken' over many domains. For each domain or block that a viewpoint was contained in, its interacting regions (averaged over all conditions) were assigned as being either in the same or neighbouring domains/blocks (either upstream or downstream), based on the midpoint of the region (with a 50% overlap rule). Error bars represent standard error of the median. To test the significance of observed interaction containment within the same topological domains and syntenic blocks, we first divided viewpoints into two categories—'middle' and 'edge' (periphery)—depending on whether the viewpoint region is located around the middle or on the outer quarter of domains/block, respectively. Then, per viewpoint we constructed the same number of interacting regions as observed, sampling their distance from the viewpoint from log-normal distribution, where log mean and standard deviation were estimated per viewpoint and condition. The expected regions within the sampled distance were then randomly placed either upstream or downstream of the tested viewpoint, and assigned as being either in the same or neighbouring domains/blocks. We tested if the 'containment' of interactions within the same block was significantly different between the observed and expected interactions using a two-sided Mann–Whitney *U*-test.

Stable 3D interactions correlate with Pol II pausing and increased DNase I hypersensitivity (associated with Fig. 4 and Extended Data Figs 9, 10). To assess the correlation between dynamic gene expression and changes in 4C interaction contacts we used whole-embryo RNA-seq from 2–4 h and 6–8 h of embryonic development[9] and called differentially expressed genes using *DESeq2* (ref. 47) (10% FDR, absolute $\log_2$ fold change ≥ 1). Genes were categorized into those going from on-to-off and off-to-on between 2–4 h and 6–8 h of development, based on the additional requirement that the gene's expression changes from non-active to active (or vice versa) using the local minima of RNA-seq log-RPKM distribution[44] (shown in Extended Data Fig. 9a). This resulted in two stringent sets of differentially expressed genes: 459 high-confidence off-to-on genes and 277 high-confidence on-to-off genes. The highest 4C count fold change was used to represent the interactions in the vicinity of off-to-on or on-to-off genes at 6–8 h of development. To test if there was a significant deviation from zero in the differential 4C signal for genes whose expression went from off-to-on or on-to-off (Fig. 4b), we used one-sample two-sided Wilcoxon test.

Static 4C interactions were defined as promoter interactions that were not called as differential between whole embryo 3–4 h and whole embryo 6–8 h (628 interactions), having an FDR or $\log_2$ fold change below the thresholds described above. Intersection of the promoters with static interactions with off-to-on gene expression resulted in a group of 20 genes that have strong differences in gene expression from 2–4 h to 6–8 h of embryonic development but non-differential promoter 3D contacts. These genes are referred to as DS genes for differentially expressed genes with stable loops.

Paused genes were previously defined based on GRO-seq data overlapping the early time point[28]. Here, we used two definitions: a stringent set which includes the top 50% of genes Saunders *et al.*[28] defined as paused based on GRO-seq, or using 'all' genes they defined as paused[28]. We then tested the likelihood of DS genes being paused by comparing the observed number of paused genes to 10,000 random samples of equal size (20 genes) from the total set of off-to-on genes (459 genes). A *P* value was calculated as the number of events where random selection contained more or equal amount of paused genes than the observed case (Fig. 4d and Extended Data Fig. 9e). The test was robust to using either the list of 'all' paused genes, as defined by Saunders *et al.*[28] ($P = 0.0071$; Extended Data Fig. 9e) or using a more stringent set of paused genes, taking the top 50% of paused genes defined by Saunders *et al.*[28] ($P = 0.02$, shown in Fig. 4d), and indicates a strong association between the presence of preformed static interactions and paused polymerase.

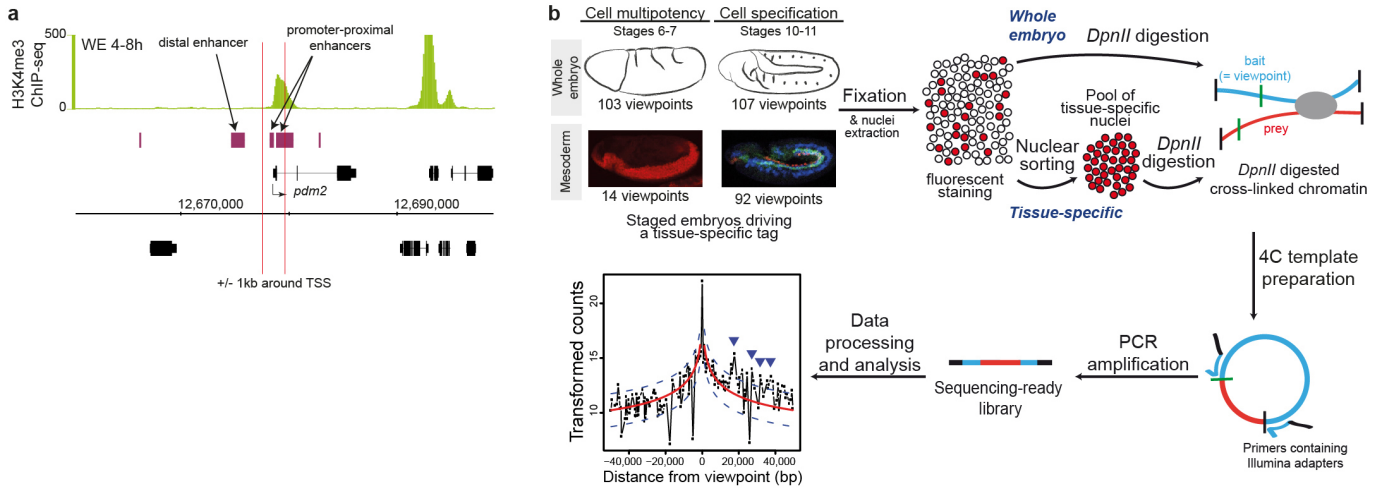
The expression of DS genes, using whole embryo RNA-seq data[9], was significantly different between the two developmental stages, as expected, given that they were selected to be both differentially expressed and go from non-active to active from 2–4 h to 6–8 h (Extended Data Fig. 9b). However, these genes are also expressed at significantly lower levels at 2–4 h than the entire set of strongly paused genes (Extended Data Fig. 9b). Similarly, the levels of nascent gene transcription, based on GRO-seq RPKM signal[28], of DS genes at the promoter were significantly higher than on the gene body, as observed for the complete set of paused genes (Extended Data Fig. 9d).

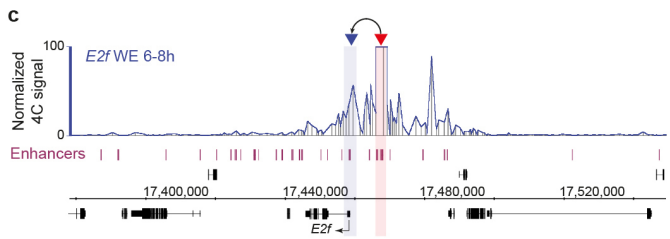
To visualize the levels of Pol II on paused genes, we first performed read shifting on samples from 2–4 h (ref. 48) and 4–8 h (ref. 49) to correct for protein positioning (57 bp for 2–4 h and 82 bp for 4–8 h sample as estimated using ArchTeX[50]), followed by normalization by genome coverage (RPGC). The graph line represents a mean value over each position (Fig. 4c and Extended Data Fig. 10).

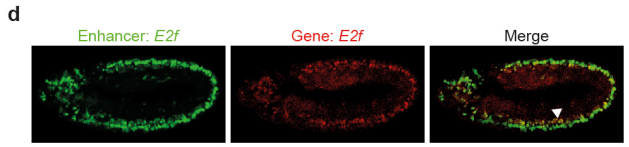
DNase I hypersensitive read counts[29] were summarized in 1-kb windows around the gene's TSS, using all mRNA-producing genes (as annotated in Flybase v5.47) with a transcript length >1 kb and ignoring chrU/Uextra and chrM (in total 11,042 genes). After library size correction, *DESeq2* was used for differential DNase I hypersensitivity analysis between stage 5 and stage 10, using default parameters. Note, 633 genes had only background DNase I signal and were therefore excluded after the differential analysis, leaving 10,409 genes. Stage 5 represents 2–4 h, and stage 10 6–8 h of development. The magnitude of DNase I differential change between both stages was examined for two categories: (1) all genes' promoters (10,409 expressed genes with transcript length >1 kb); and (2) DS genes' promoters (18 genes). To test for differences between DNase I hypersensitivity at the DS genes' promoters and the 10,409 background genes' promoters, a two-sample two-sided Wilcoxon test was used on the $\log_2$ DNase I fold change (Fig. 4f).

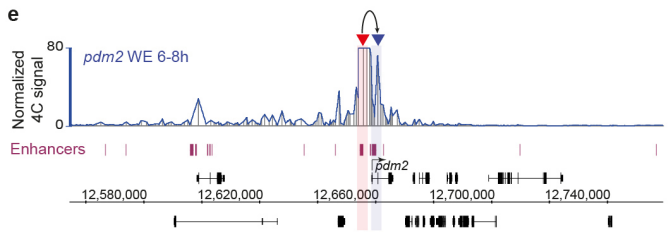
31. Noordermeer, D. *et al.* The dynamic architecture of Hox gene clusters. *Science* **334,** 222–225 (2011).
32. Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature Genet.* **44,** 148–156 (2012).
33. Campos-Ortéga, J. A. & Hartenstein, V. *The Embryonic Development of* Drosophila melanogaster (Springer, 1997).
34. Bonn, S. *et al.* Cell type-specific chromatin immunoprecipitation from multicellular complex samples using BiTS-ChIP. *Nature Protocols* **7,** 978–994 (2012).
35. van de Werken, H. J. G. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nature Methods* **9,** 969–972 (2012).
36. Bantignies, F. *et al.* Polycomb-dependent regulatory contacts between distant Hox loci in *Drosophila*. *Cell* **144,** 214–226 (2011).
37. Venken, K. J. T., He, Y., Hoskins, R. A. & Bellen, H. J. P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science* **314,** 1747–1751 (2006).
38. Barolo, S., Carver, L. A. & Posakony, J. W. GFP and β-galactosidase transformation vectors for promoter/enhancer analysis in *Drosophila*. *Biotechniques* **29,** 726–732 (2000).
39. Rubin, G. M. & Spradling, A. C. Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218,** 348–353 (1982).
40. Furlong, E. E., Andersen, E. C., Null, B., White, K. P. & Scott, M. P. Patterns of gene expression during *Drosophila* mesoderm development. *Science* **293,** 1629–1633 (2001).
41. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300 (1995).
42. Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481,** 389–393 (2012).
43. Marygold, S. J. *et al.* FlyBase: improvements to the bibliography. *Nucleic Acids Res.* **41,** D751–D757 (2013).
44. Hebenstreit, D. *et al.* RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Mol. Syst. Biol.* **7,** 497 (2011).
45. Wilczynski, B., Liu, Y.-H., Yeo, Z. X. & Furlong, E. E. M. Predicting spatial and temporal gene expression using an integrative model of transcription factor occupancy and chromatin state. *PLOS Comput. Biol.* **8,** e1002798 (2012).

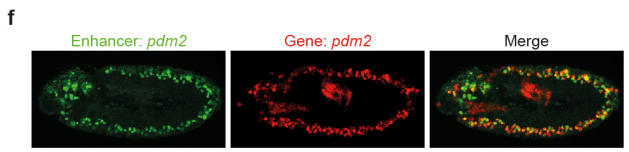
46. Gallo, S. M. *et al.* REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in *Drosophila*. *Nucleic Acids Res.* **39,** D118–D123 (2011).
47. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).
48. Chen, K. *et al.* A global change in RNA polymerase II pausing during the Drosophila midblastula transition. *eLife* **2,** e00861 (2013).
49. Nègre, N. *et al.* A cis-regulatory map of the *Drosophila* genome. *Nature* **471,** 527–531 (2011).
50. Lai, W. K. M., Bard, J. E. & Buck, M. J. ArchTEx: accurate extraction and visualization of next-generation sequence data. *Bioinformatics* **28,** 1021–1023 (2012).
51. Berman, B. P. *et al.* Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in *Drosophila melanogaster* and *Drosophila pseudoobscura*. *Genome Biol.* **5,** R61 (2004).
52. Keleman, K. & Dickson, B. J. Short- and long-range repulsion by the *Drosophila* Unc5 netrin receptor. *Neuron* **32,** 605–617 (2001).
53. Albrecht, S., Altenhein, B. & Paululat, A. The transmembrane receptor Uncoordinated5 (Unc5) is essential for heart lumen formation in *Drosophila melanogaster*. *Dev. Biol.* **350,** 89–100 (2011).
54. Gaertner, B. *et al.* Poised RNA polymerase II changes over developmental time and prepares genes for future expression. *Cell Rep.* **2,** 1670–1683 (2012).
55. Tomancak, P. *et al.* Global analysis of patterns of gene expression during Drosophila embryogenesis. *Genome Biol.* **8,** R145 (2007).
56. Core, L. J. & Lis, J. T. Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science* **319,** 1791–1792 (2008).

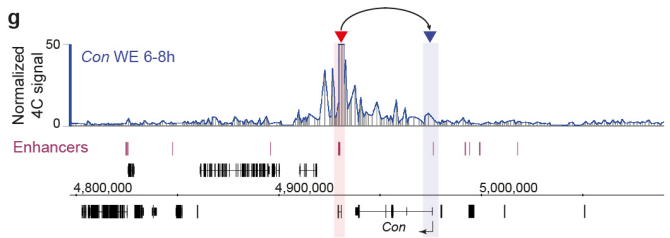


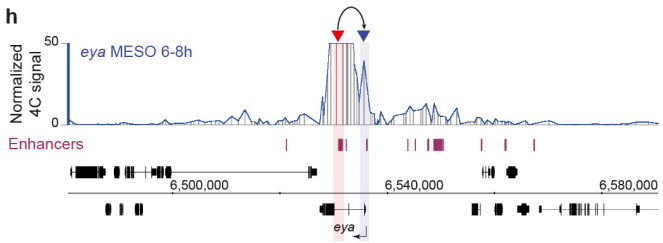


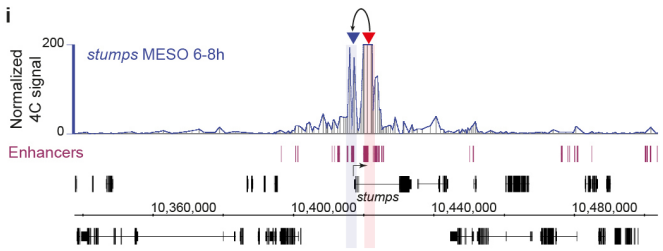


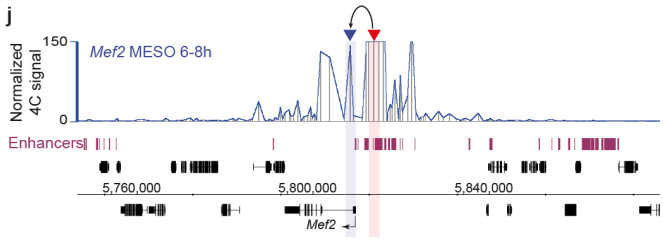


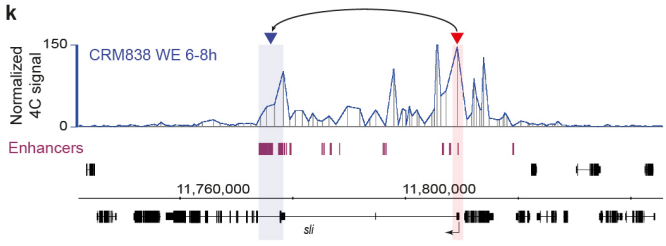


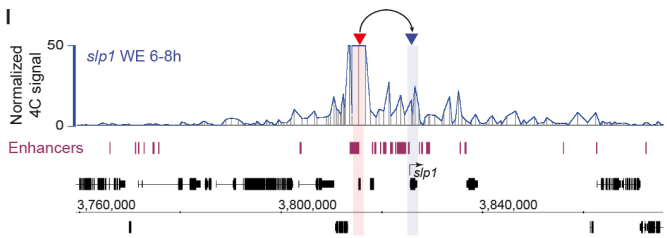




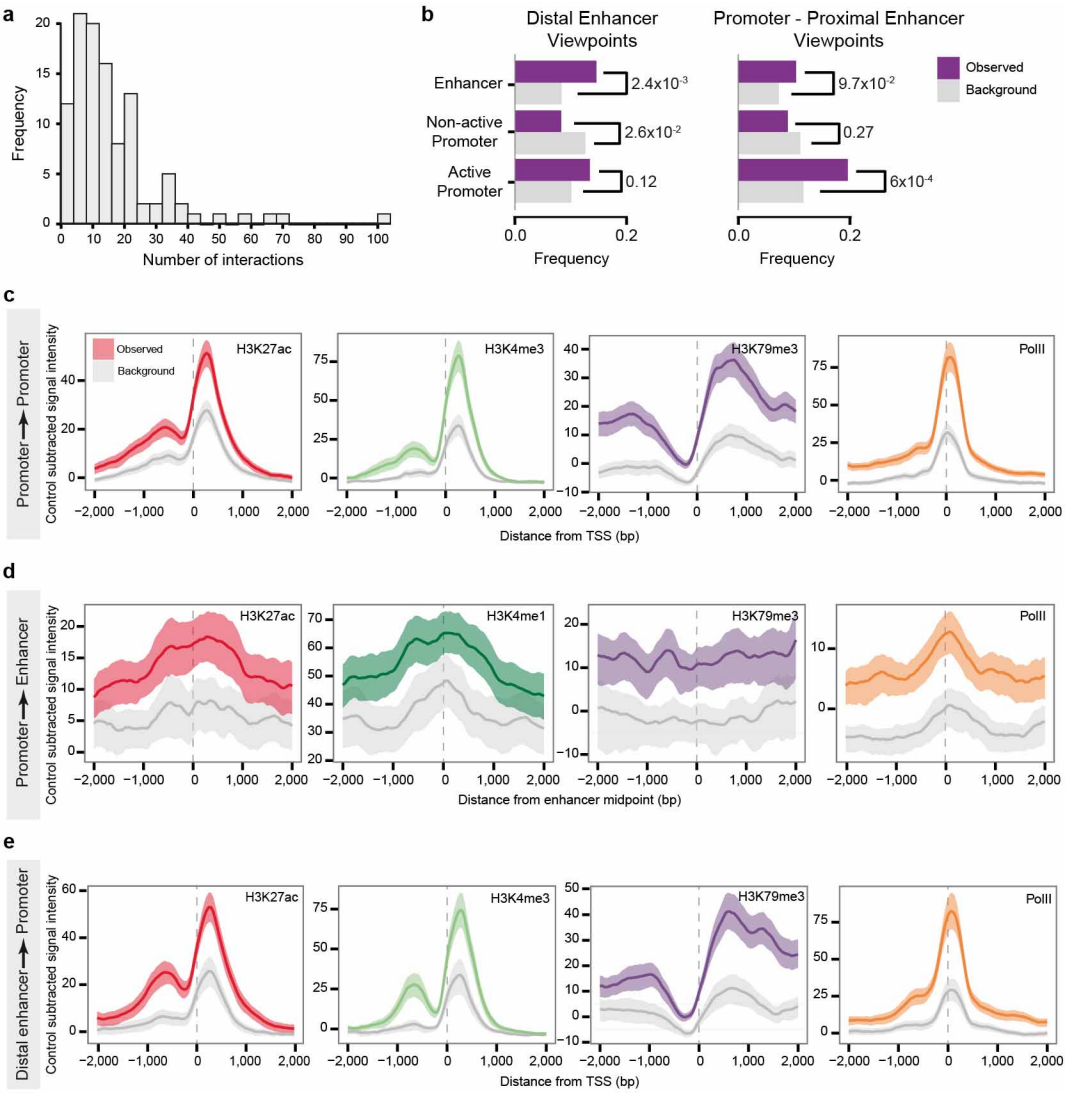




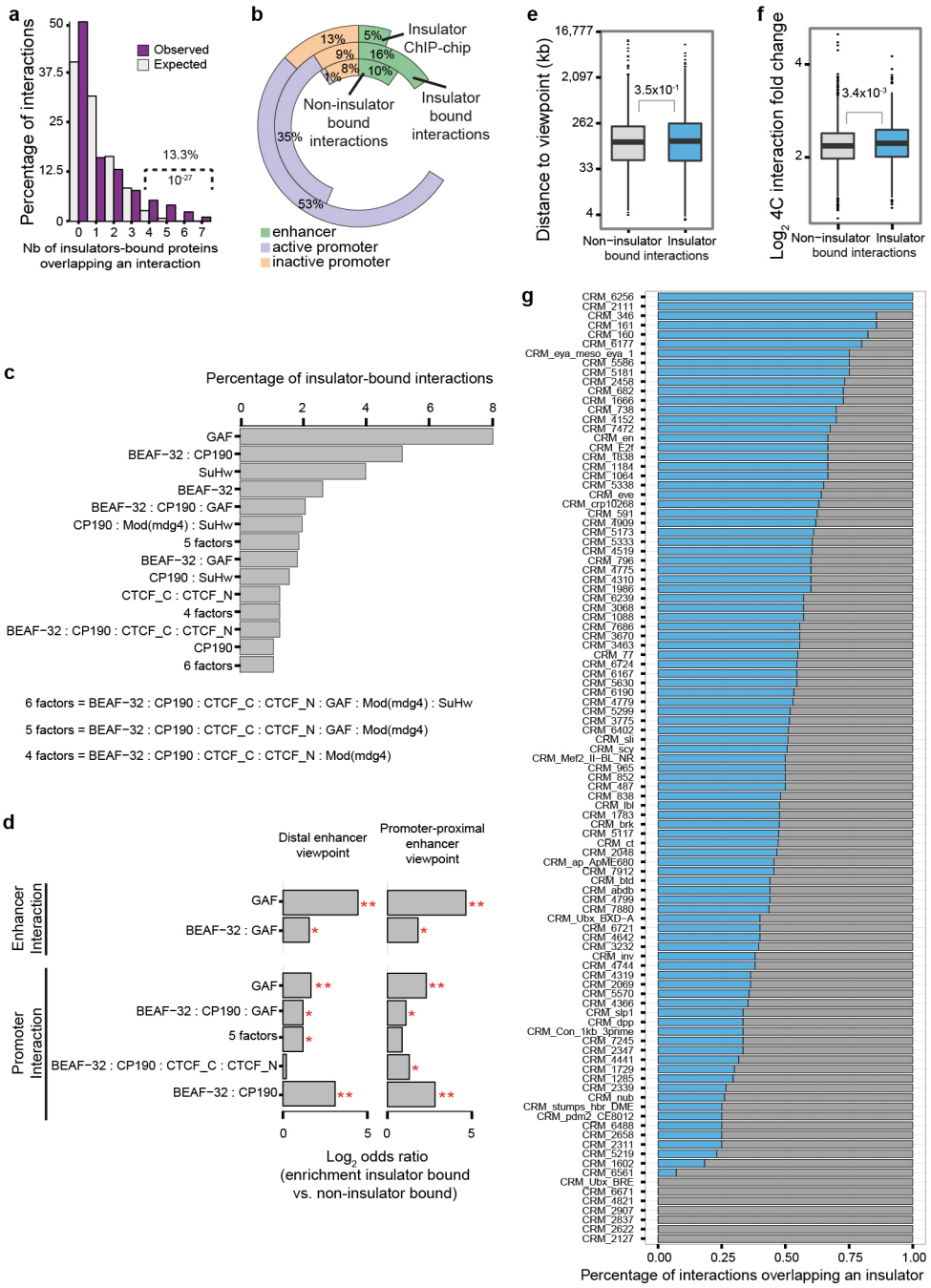




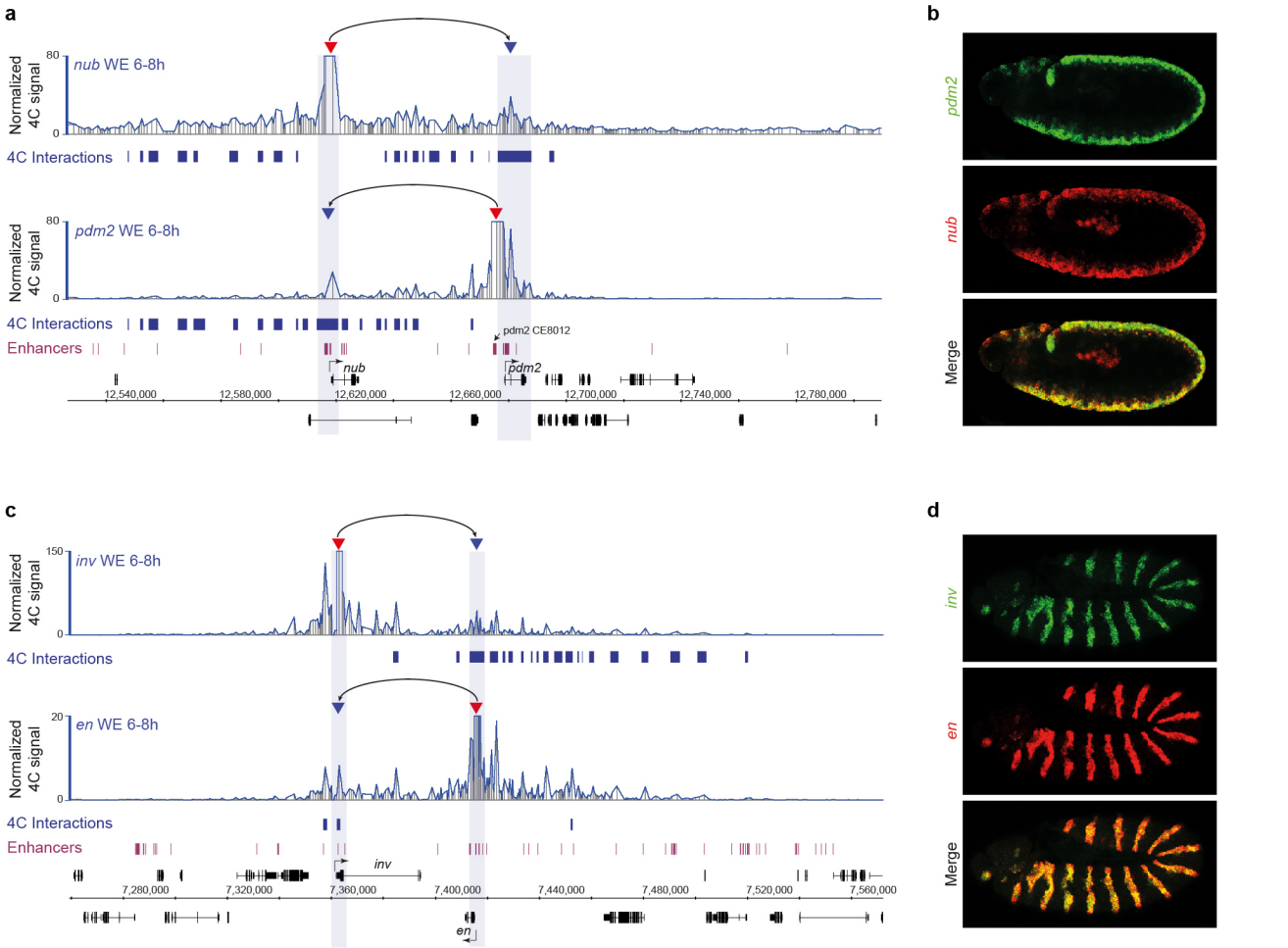
Extended Data Figure 1 | 4C-seq in *Drosophila* embryos. **a**, Example of classification into distal and promoter-proximal enhancer viewpoints. Distal enhancers are located >1 kb away from the closest TSS and devoid of H3K4me3 chromatin signature (green), a mark characteristic of gene promoters, while promoter-proximal enhancers are located within 1 kb of the closest TSS. WE, whole embryo. **b**, Outline of BiTS-4C-seq: genetically modified embryos expressing a tagged nuclear protein exclusively in the mesoderm were collected, aged and crosslinked. Fixed nuclei were extracted and digested with DpnII (whole embryo analysis) or sorted by FACS before DpnII digestion (tissue-specific analysis). After 4C template preparation, ligation junctions were amplified and sequenced. Count values for each DpnII fragment were transformed and fitted using a monotone fit (red line) facilitating an analysis of significant interactions (blue arrowheads). In addition to the 103 viewpoints used at both time-points, four additional viewpoints were used for reciprocal 4C experiments in whole embryos at 6–8 h. **c**, **e**, **g**–**l**, 4C interaction maps (viewpoint, red arrowhead) at the *E2f* (**c**), *pdm2* (ref. 51) (**e**), *Con* (**g**), *eya* (**h**), *stumps* (**i**), *Mef2* (**j**), *sli* (**k**) and *slp1* (**l**) loci. The expected interaction with the promoter (blue arrowhead) of the genes is observed. Known enhancers are indicated. MESO, mesoderm (generated by FACS sorting). **d**, **f**, Expression (double *in situ* hybridization) of the *E2f* (**d**) or *pdm2* (**f**) genes (red) and the expression driven by their interacting enhancer (green) at stage 11.



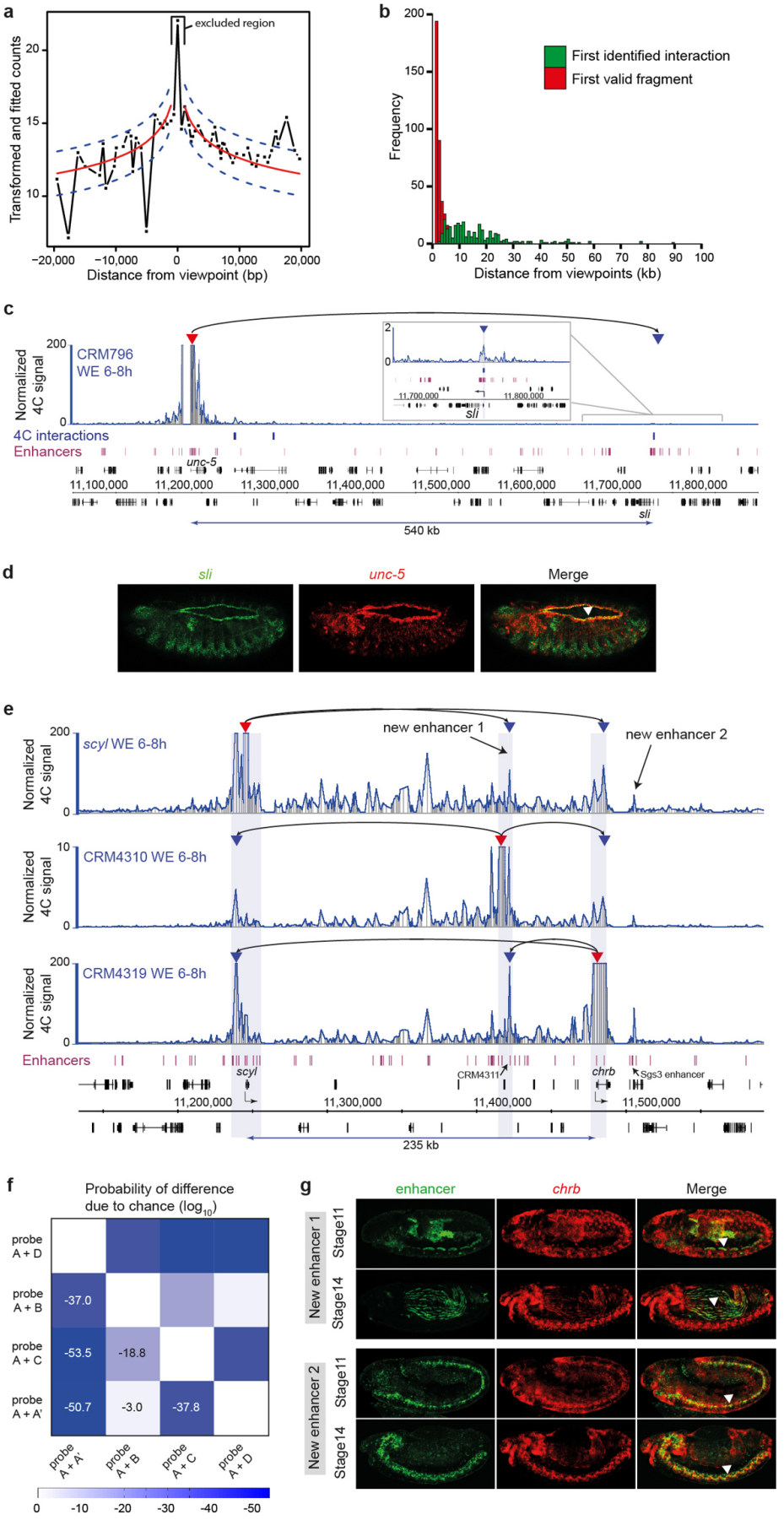
Extended Data Figure 2 | Enhancers have a complex interaction network with multiple active enhancers and promoters. **a**, Histogram showing the frequency of viewpoint interactions, averaged over all conditions per viewpoint ($n = 107$). **b**, Frequency of 4C unique interacting regions that overlap enhancers or promoters when using distal enhancers (left) or promoter-proximal elements (right) as viewpoints (defined in Extended Data Fig. 1a). Whole-embryo RNA-seq[9] was used to define active and inactive promoters. Enrichments over a background set of interactions (Fisher's Exact Test) are indicated. The background set (or expectation regions) was based on random sampling of all DpnII fragments throughout the genome that have matched mappability, G+C content and size to the observed interactions. **c**–**e**, Mesoderm-specific chromatin signatures[6] at whole embryo 6–8 h 4C interacting regions. **c**, H3K27ac, H3K4me3, H3K79me3 and Pol II signals[6] are enriched at promoters interacting with promoter-proximal viewpoints. Note, we only consider interactions between promoter-proximal viewpoints and distal promoters (>1 kb away; Extended Data Fig. 1a). **d**, H3K27ac, H3K4me1, H3K79me3 and Pol II signals[6] are enriched at enhancers interacting with promoter-proximal viewpoints. **e**, H3K27ac, H3K4me3, H3K79me3 and Pol II signals[6] are enriched at promoters interacting with distal enhancer viewpoints. Shading indicates 95% confidence intervals estimated by non-parametric bootstrapping, grey lines indicate background signal.



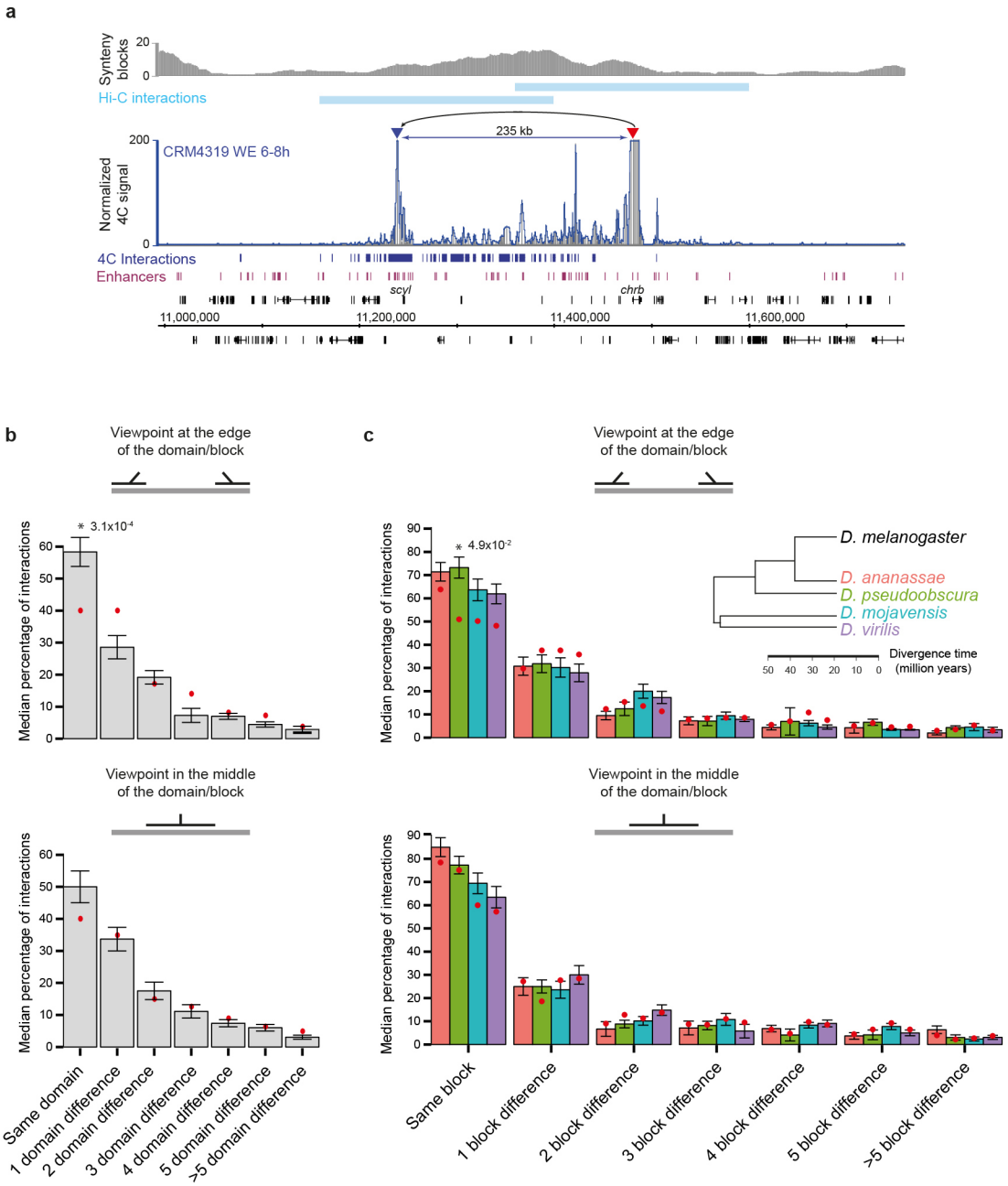
Extended Data Figure 3 | 4C interactions at insulator protein-bound and unbound regions. **a**, Histogram showing the observed and expected percentage of 4C interacting regions (shown for whole embryo 6–8 h) that overlap with insulator-binding protein occupancy[17] (6 factors, including N- and C-term CTCF). A significantly higher percentage of interactions overlap the occupancy of at least 4 insulator-binding proteins (Fisher's Exact Test). **b**, Percentage of genomic regions occupied by insulator-binding proteins (outer circle), insulator-bound 4C-interacting regions (middle circle) and non-insulator-bound 4C-interacting regions (inner circle) that overlap enhancers or promoters. **c**, Histogram representing the percentage of insulator-bound interacting regions for all combinations of insulator proteins that overlap with at least 20 interacting regions. **d**, Log_2 odds ratio for an interacting region bound by a given combination of insulator proteins to be more enriched on enhancers or promoters than non-insulator bound interacting regions (Fisher's Exact Test). Red asterisks in **d**: *FDR <10%; **FDR <0.1%. **e**, Box plot showing the genomic distance to the viewpoint of non-insulator-bound and insulator-bound interacting regions (*P* value from a two-sided Wilcoxon test). **f**, Box plot showing the interaction strength (log_2 fold change of the 4C signal) on non-insulator-bound and insulator-bound 4C interacting regions (*P* value from a two-sided Wilcoxon test). **g**, Histogram showing, for each of the 107 viewpoints (103 plus 4 reciprocal viewpoints), the percentage of 4C interacting regions (blue) overlapping with the binding of insulator proteins[17].



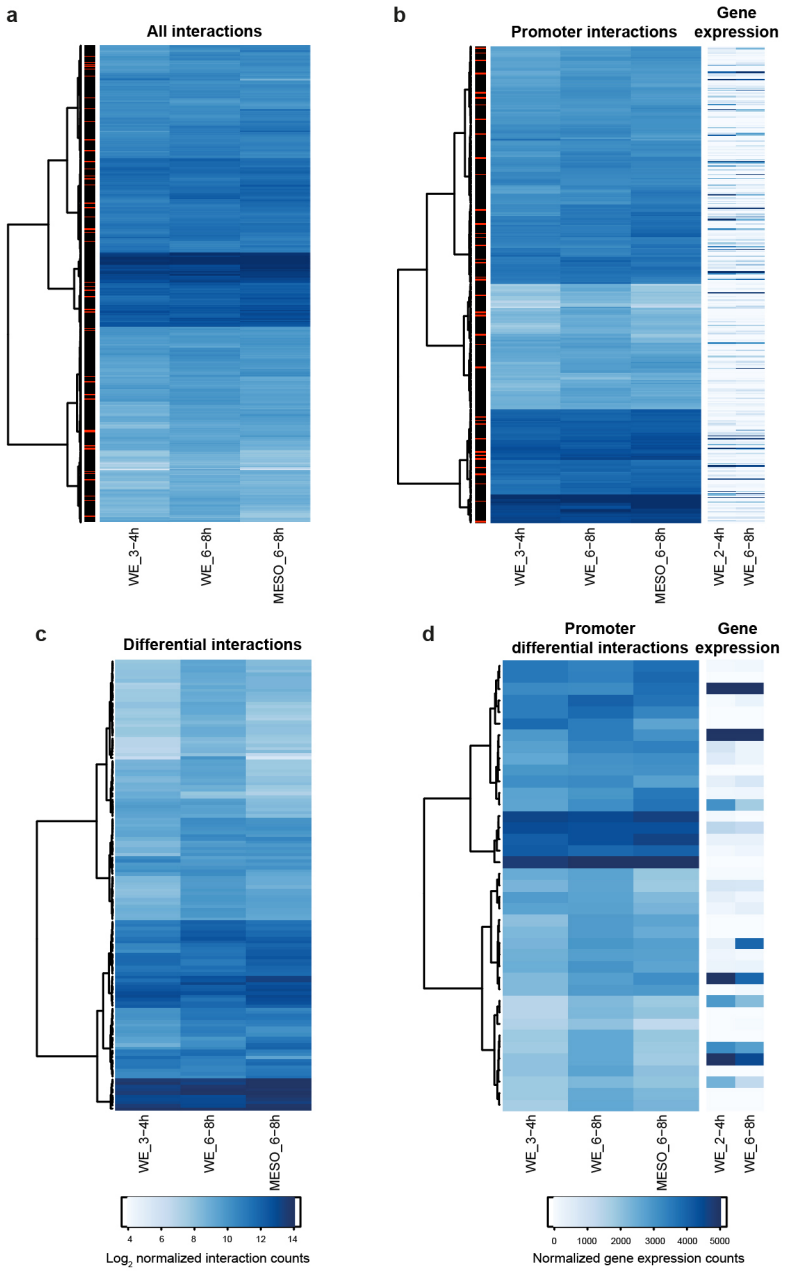
Extended Data Figure 4 | Chromatin 3D interactions at the *pdm2* and *en* loci. **a**, **c**, Reciprocal 4C interaction map (viewpoint, red arrowhead) around the *nub* and *pdm2* (**a**) or *en* and *inv* (**c**) loci. Top 2 lanes: 4C interactions with a region surrounding the *pdm2* or *nub* (**a**) and *en* or *inv* (**c**) genes (blue arrowhead). Top lane: 4C using the *nub* (**a**) or *inv* (**c**) promoter as the viewpoint; bottom lane: using a promoter-proximal region in the vicinity of the *pmd2* (**a**) or *en* (**c**) promoter as the viewpoint. Significant 4C interactions and known enhancers are indicated. The *pdm2* CE8012 enhancer is indicated. WE, whole embryo. **b**, **d**, Expression (double *in-situ* hybridization) of the *pdm2* or *inv* (green) and *nub* or *en* (red) at stage 11, the same stage shown for the 4C interaction in **a** and **c** (6–8 h).



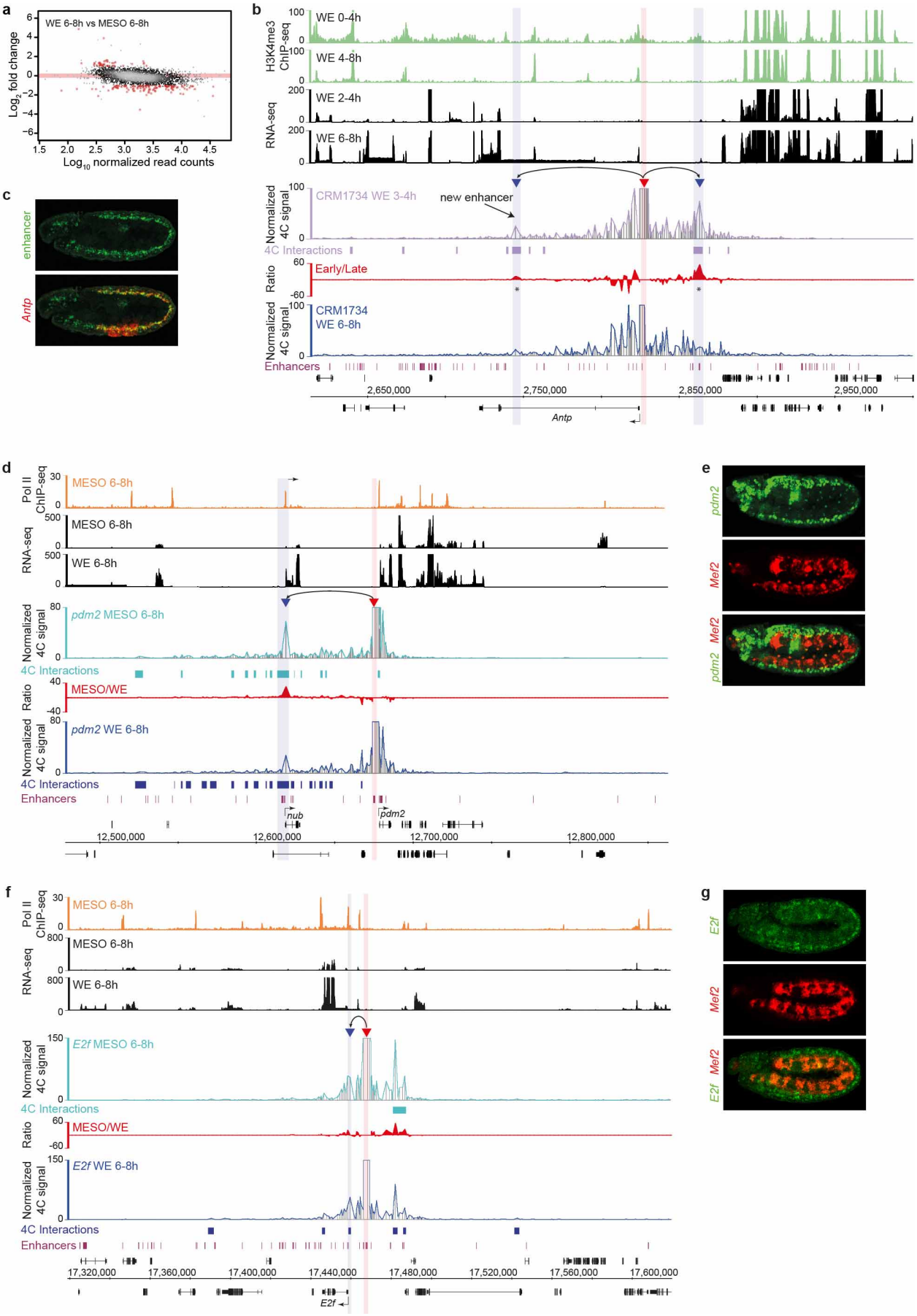
Extended Data Figure 5 | Long-range interactions in *Drosophila*. **a**, Zoom-in of the 4C-seq transformed and fitted counts around the *apterous* (*ap*) viewpoint (mesoderm 6–8 h). The first fragments around the viewpoint are not included in the monotone fit (red line). **b**, Frequency of the first identified interaction (green) and the first valid fragment (red) as a function of distance to their respective viewpoints. **c**, 4C interaction map of the *unc-5* and *sli* loci[52,53]. An interaction (blue arrowhead) between the viewpoint (red arrowhead) and the promoter of *sli*, over half a megabase away, is observed. Inset shows a zoomed-in view of the *sli* promoter. The location of significant 4C interactions and known enhancers is represented below. WE, whole embryo. **d**, Double *in situ* hybridization showing the overlap between *sli* (green) and *unc-5* (red) heart expression (arrowhead) at stage 14. **e**, 4C interaction map at the *scyl* and *chrb* loci[23]. Independent of the location of the viewpoint (red arrowhead), the same interacting regions are recovered (blue arrowheads). The location of known enhancers is represented below. The two cloned regions are indicated as 'new enhancer' 1 and 2, respectively. WE, whole embryo. **f**, A non-parametric two-sample Kolmogorov–Smirnov test was used to assess the significance of differences between DNA FISH distance distributions. **g**, Double *in situ* hybridization of the *chrb* gene (red) and the expression driven by two of its interacting regions (green) at stage 11 and 14, showing that both regions recapitulate part of the expression of *chrb*. The overlap (shown by a white arrowhead) is located in the trunk visceral mesoderm and later in the longitudinal visceral mesoderm for enhancer 1 and in the central nervous system for enhancer 2. New enhancer 1 is overlapping a ChIP-defined CRM (CRM4311), whereas new enhancer 2 is partially overlapping (<10%) an enhancer in the *sgs3* locus. Neither region was previously identified as *charybde* enhancers, and their spatio-temporal activity during development was unknown.



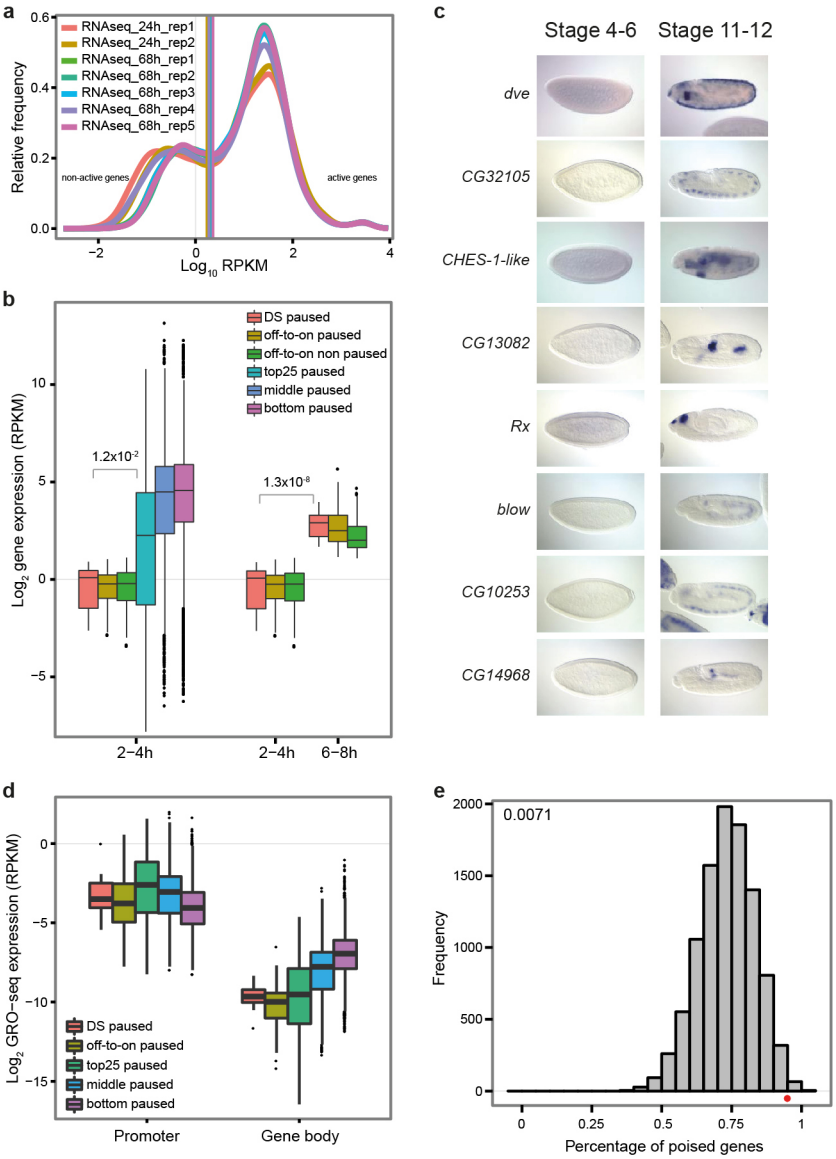
Extended Data Figure 6 | Location of 4C interactions with respect to topological domains and block of microsynteny. a, 900-kb region around the *scyl* and *chrb* loci. Top to bottom: regions of conserved synteny (grey)[24]; location of previously identified Hi-C interactions (light blue)[15]; 4C interaction map around the *scyl* and *chrb* loci (viewpoint location, red arrowhead); location of significant 4C interactions and known enhancers. The strong interaction with the *scyl* gene is highlighted (blue arrowhead). WE, whole embryo. **b**, Percentage of 4C interactions, averaged over all conditions per viewpoint, located in the same topological domain (TAD) as the viewpoint, or in adjacent TADs, if the viewpoint is on the edge of the domain (top) or in the middle (bottom). TADs were defined based on HiC data at 16–18 h of development[15]. Significant difference from the background (asterisk) and corresponding *P* values using a two-sided Mann–Whitney *U*-test are indicated. **c**, Percentage of interactions located in the same block of microsynteny as the viewpoint, or in adjacent blocks, if the viewpoint is on the edges of the block (top) or in the middle (bottom). Microsynteny was previously defined by Engström *et al.*[24] based on 5 species spanning ~50 million years of evolution. Significant differences between the observed and expected interactions (asterisk) were assessed using a two-sided Mann–Whitney *U*-test.



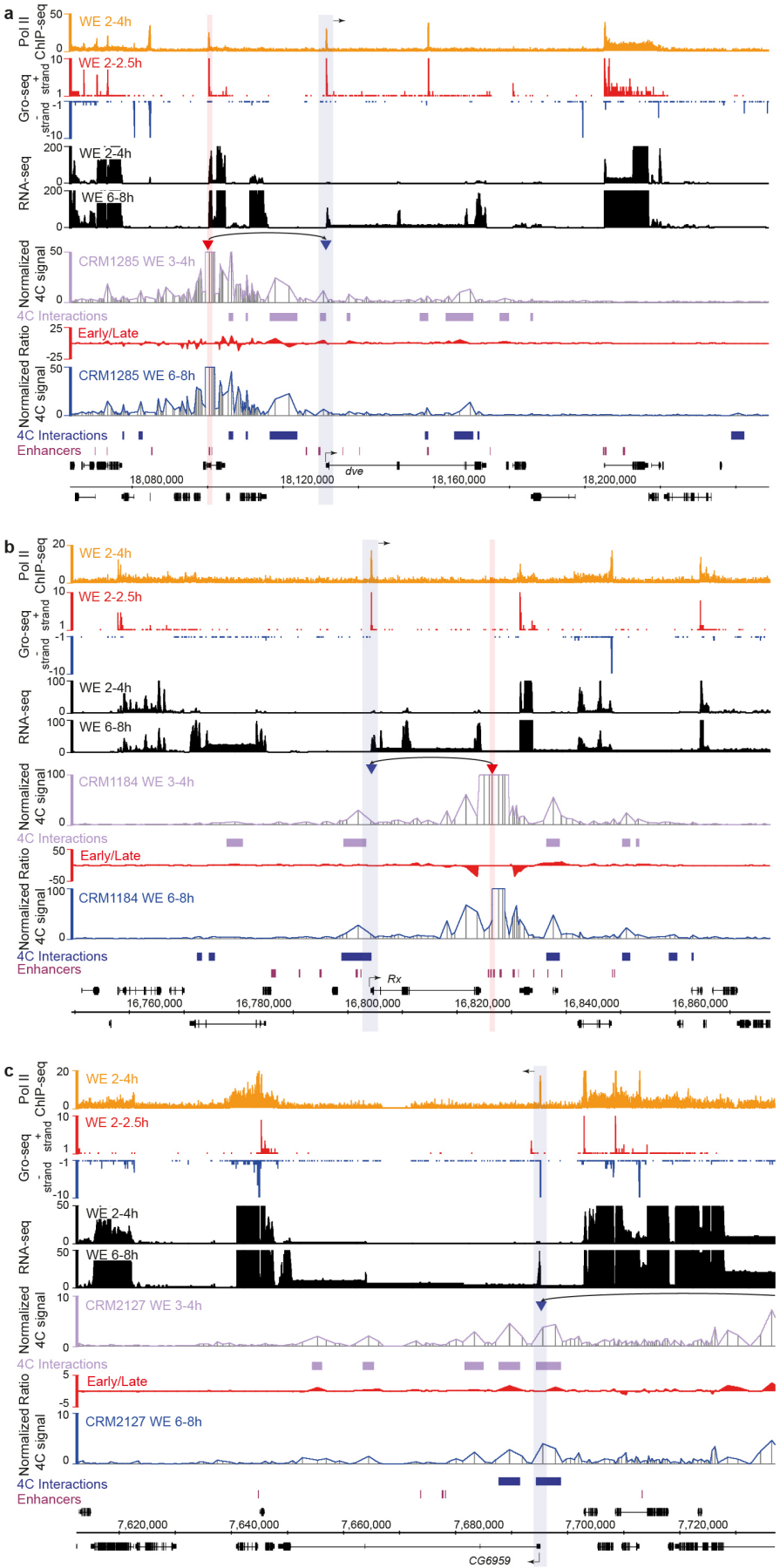
Extended Data Figure 7 | Hierarchical clustering of 4C interactions. **a**, **b**, Based on the 92 viewpoints used for 4C in all three conditions (whole embryo 3–4 h, whole embryo 6–8 h and mesoderm 6–8 h). Hierarchical clustering of the quantitative 4C signal of all interactions ($n = 1{,}389$) (**a**) and the subset of interactions with promoters ($n = 396$) (**b**), across 3 conditions (whole embryo 3–4 h, whole embryo 6–8 h and mesoderm 6–8 h). The expression level of the corresponding gene is shown on the right hand side in **b**, using RNA-seq data from Graveley *et al.*[9]. Clustering was performed by Euclidean distance and Ward agglomeration method on the significant interaction regions defined at whole embryo 6–8 h using *DESeq2* fit reference-normalized read counts. If the interacting region was associated with more than one promoter, expression for only one transcript is shown. Red horizontal bars indicate differential interactions. **c**, **d**, Hierarchical clustering of the quantitative 4C signal for all 140 differential interactions (**c**) and the subset of differential interactions at promoters ($n = 39$) with the expression level of the corresponding gene (**d**) across 3 conditions (whole embryo 3–4 h, whole embryo 6–8 h and mesoderm 6–8 h). Clustering was performed using Euclidean distance and Ward agglomeration method on the significant interaction regions defined at whole embryo 6–8 h. WE, whole embryo; MESO, mesoderm.



Extended Data Figure 8 | 4C interactions under different developmental conditions. **a**, MA plot of interaction signal between whole embryo 6–8 h and mesoderm 6–8 h. **b**, 4C interaction map at the *Antp* locus. Top to bottom: H3K4me3 ChIP-seq signal (RPGC) in whole embryo at 0–4 h and 4–8 h (ref. 49) (green), RNA-seq signal (RPKM, black) in whole embryo at 2–4 h and 6–8 h (ref. 9), 4C interaction map (viewpoint, red arrowhead) in whole embryo at 3–4 h (mauve) and 6–8 h (blue). The differential 4C signal is plotted in between in red with significant differential 4C interactions indicated (asterisk). WE, whole embryo. **c**, Expression (*in situ* hybridization) of *Antp* (red) and the expression driven by a new *Antp* enhancer (green) at stage 11 (6–8 h). The enhancer's activity overlaps expression of *Antp* at 6–8 h; however, the 4C contact between the enhancer and promoter is already present, and at even higher levels, at 3–4 h. **d**, **f**, Interaction map at the *pdm2* (**d**) and *E2f* (**f**) loci. Top to bottom: Pol II signal (RPGC) in mesoderm at 6–8 h (orange)[6], RNA-seq signal (RPKM) in mesoderm and whole embryo at 6–8 h (black)[9,54], 4C interaction map (viewpoint location, red arrowhead) in mesoderm (light blue) and whole embryo (dark blue) at 6–8 h. The differential 4C signal is plotted in between in red. Note that the 4C interaction is stronger in mesoderm compared to whole embryo, although those genes are not expressed in the mesoderm. Significant 4C interactions and known enhancers are indicated. WE, whole embryo, MESO, mesoderm (generated by FACS sorting). **e**, **g**, Expression (double *in situ* hybridization) of the *pdm2* (**e**) or *E2f* (**g**) genes (green) with a mesoderm-specific marker (*mef2*, red) at stage 11.



Extended Data Figure 9 | Characterization of DS genes. **a**, Defining off-to-on genes. Relative frequency of genes exhibiting a given RPKM value from RNA-seq in whole embryo at 2–4 h and 6–8 h (ref. 9). The threshold (vertical lines) between non-active (off) and active genes was selected based on the local minima in the log-RPKM distribution, as described previously[44]. *DESeq* was used to determine if non-active 2–4 h genes had a significant change in their expression at 6–8 h. **b**, DS-paused genes are expressed at very low levels at 2–4 h, or not at all. Log_2 gene expression signal (RPKM) in whole embryo at 2–4 h and 6–8 h (ref. 9) of different categories of genes. Paused DS genes (15 genes, using the stringent criteria for pausing) are significantly less expressed than the top 25% of paused genes (1,776 genes)[28] at 2–4 h and are also significantly less expressed at 2–4 h compared to 6–8 h (Mann–Whitney *U*-test). **c**, *In situ* hybridization showing available expression data for DS genes at stage 4–6 (2–4 h) and 11–12 (Berkeley *Drosophila* Genome Project[55]). Note, for all 8 genes there is no detectable specific expression at the early time point. **d**, Log_2 GRO-seq expression signal (RPKM) in whole embryo at 2–2.5 h (ref. 28) at the promoter and gene body of different categories of genes. RPKM was defined by Saunders *et al.*[28]. **e**, Histogram (grey bars) of the expected distribution for a gene to be paused (using 'all' paused genes defined by Saunders *et al.*[28]) by random sampling 20 genes 10,000 times from the 459 differentially expressed off-to-on genes. The red dot indicates the percentage of observed paused genes for all DS genes (differentially expressed but with stable loops). Using this test, the percentage of paused DS genes is highly significant (*P* 0.0071).



Extended Data Figure 10 | Stable enhancer loops prefigure gene expression and are associated with paused Pol II. Interaction map at three known Pol II paused genes: *dve* (**a**), *Rx* (**b**) and *CG6959* (**c**). Top to bottom: Pol II signal (RPGC) in whole embryo at 2–4 h (orange)[48], GRO-seq signal in whole embryo at 2–2.5 h (plus strand, red; minus strand, blue)[28], RNA-seq signal (RPKM) in whole embryo at 2–4 h and 6–8 h (black)[9], 4C interaction map (viewpoint, red arrowhead) in whole embryo at 3–4 h (purple) and 6–8h (dark blue). The ratio between whole embryo 3–4 h and whole embryo 6–8 h 4C signal is plotted in between in red. Significant 4C interactions and known enhancers are indicated. WE, whole embryo. Note, *dve* and *Rx* are transcribed on the plus strand, whereas *CG6959* is transcribed on the minus. In each case, the enhancer–promoter interactions, Pol II occupancy in the absence of full-length transcript production (indicative of Pol II pausing[56]) and short nascent RNA transcription are already present at the promoter at 2–4 h, whereas the gene becomes highly expressed at 6–8 h. As the viewpoint in **c** is located 80 kb away from the promoter of *CG6959*, a zoomed-in view of the promoter is shown for clarity.