

Independence of 3D chromatin conformation and gene regulation during *Drosophila* dorsoventral patterning

Elizabeth Ing-Simmons¹, Roshan Vaid², Mattias Mannervik², and Juan M. Vaquerizas^{1,3}

1. Max Planck Institute for Molecular Biomedicine, Röntgenstraße 20, 48149 Münster, Germany

2. Stockholm University, Department of Molecular Biosciences, The Wenner-Gren Institute, SE-106 91 Stockholm, Sweden

3. MRC London Institute of Medical Sciences, Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, Du Cane Road, London W12 0NN, UK

Corresponding author: Juan M. Vaquerizas (jmv@mpi-muenster.mpg.de, Twitter: [@vaquerizasjm](https://twitter.com/vaquerizasjm))

Keywords: genome organisation, chromatin conformation, development, gene regulation, *Drosophila*

ABSTRACT

The relationship between the 3D organisation of chromatin inside the nucleus and the regulation of gene expression remains unclear. While disruption of domains and domain boundaries can lead to mis-expression of developmental genes, acute depletion of key regulators of genome organisation, such as CTCF and cohesin, and major reorganisation of genomic regions have relatively small effects on gene expression. Therefore, it is unclear whether changes in gene expression and chromatin state drive chromatin reorganisation, or whether changes in chromatin organisation facilitate cell type-specific activation of genes and their regulatory elements. Here, using the *Drosophila melanogaster* dorsoventral patterning system as a model, we demonstrate the independence of 3D chromatin organisation and developmental gene regulation. We define tissue-specific enhancers and link them to expression patterns at the single-cell level using single cell RNA-seq. Surprisingly, despite tissue-specific differences in chromatin state and gene expression, 3D chromatin organisation is maintained across tissues. Our results provide strong evidence that tissue-specific chromatin conformation is not required for tissue-specific gene expression, but rather acts as an architectural framework to facilitate proper gene regulation during development.

Chromatin is highly organised within the nucleus into chromosome territories, compartments of active and inactive chromatin, self-interacting domains, and loops between specific loci. However, the relationship between the 3D organisation of chromatin and the regulation of gene expression remains unclear. There is much evidence that chromatin conformation is important for gene regulation: disruption of domains and domain boundaries can lead to mis-expression of developmental genes, contributing to developmental defects and cancer (Flavahan et al., 2016; Franke et al., 2016; Hnisz et al., 2016; Ibn-Salem et al., 2014; Lupiáñez et al., 2015; Spielmann et al., 2018). In addition, the general principles

of 3D genome organisation are conserved across large evolutionary distances, as well as the chromatin conformation at specific loci (Fudenberg and Pollard, 2019; Harmston et al., 2017b; Krefting et al., 2018; Özdemir and Gambetta, 2019; Rowley et al., 2017; Vietri Rudan et al., 2015). Further evidence comes from the identification of interactions between promoters and their regulatory elements (Cruz-Molina et al., 2017; Ghavi-Helm et al., 2014; Hsieh et al., 2020; Jin et al., 2013; Krietenstein et al., 2020; Li et al., 2012; Sanyal et al., 2012; Weintraub et al., 2017), and the finding that forced enhancer-promoter looping is sufficient to activate transcription of some genes (Bartman et al., 2016; Deng et al.,

2012; Deng et al., 2014; Morgan et al., 2017). However, in other cases direct enhancer-promoter contacts may be neither strictly required nor sufficient for gene activation (Alexander et al., 2019; Benabdallah et al., 2019; Chen et al., 2018; Heist et al., 2019). Furthermore, depletion of key regulators of genome organisation such as CTCF and cohesin has relatively small effects on gene expression (Nora et al., 2017; Rao et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017), and genomic rearrangements are not always associated with changes in gene expression (Ghavi-Helm et al., 2019; Meadows et al., 2010; Williamson et al., 2019).

While multiple studies have shown differences in chromatin conformation between different cell types or tissues (Bonev et al., 2017; Chathoth and Zabet, 2019; Joshi et al., 2015; Kragestein et al., 2018; Le Dily et al., 2014; Oudelaar et al., 2020; Schmitt et al., 2016), whether these changes are the cause or consequence of changes in gene expression is unclear. Therefore, a fundamental question arises as to whether changes in gene expression and chromatin state drive chromatin reorganisation, or whether changes in chromatin organisation facilitate cell type-specific activation of genes and their regulatory elements.

Embryonic development requires precise regulation of gene expression, making it an ideal context in which to investigate the relationship between gene regulation and chromatin organisation (Hug and Vaquerizas, 2018a). In particular, *Drosophila melanogaster* has long been used as a model organism for the study of development, and the key principles and factors involved in embryonic patterning are well understood (Wieschaus, 2016; Wolpert, 2016). Early *Drosophila* development involves a series of thirteen rapid, synchronous nuclear divisions, before the embryo becomes cellularised and undergoes zygotic genome activation

(ZGA) at nuclear cycle (nc) 14 (Hamm and Harrison, 2018). We and others have previously shown that chromatin organisation in *Drosophila* is established at nc14, coincident with ZGA (Hug et al., 2017; Ogiyama et al., 2018). While a small number of genes are zygotically expressed prior to the major wave of ZGA (Lott et al., 2011), maternally provided cues are responsible for establishing the major anterior-posterior and dorsoventral axes (Ma et al., 2016; Stein and Stevens, 2014). Therefore, by ZGA, cells in different regions of the embryo contain different developmental transcription factors, have different patterns of chromatin accessibility (Cusanovich et al., 2018), and are primed to express different genes.

Cell fate along the dorsoventral (DV) axis is controlled by the nuclear concentration of the transcription factor Dorsal (Dl) (Hong et al., 2008; Stathopoulos and Levine, 2002), which peaks during nc14 (Reeves et al., 2012). Activation of the Toll signalling pathway on the ventral side of the embryo leads to high levels of Dl entering the nucleus, while Dl is excluded from the nucleus on the dorsal side (Stein and Stevens, 2014). Different levels of Dl concentration are responsible for the specification of different cell fates (Hong et al., 2008; Stathopoulos and Levine, 2002) (Fig. 1A). Maternal effect mutations in the Toll pathway lead to a uniform level of nuclear Dl across the embryo, making it possible to obtain females that produce a homogeneous population of embryos that consist entirely of presumptive mesoderm (*Toll^{l0B}*), neuroectoderm (*Toll^{rm9/rm10}*), or dorsal ectoderm (*gd⁷*) (Fig. 1A). These embryos provide an excellent model system to study tissue-specific regulation during development, which has led to the discovery of key transcription factors, regulatory elements, and processes required for embryo patterning (Boija and Mannervik, 2016; Hong et al., 2008; Koenecke et al., 2016; Koenecke et al., 2017; Stathopoulos et al., 2002; Zeitlinger et al., 2007).

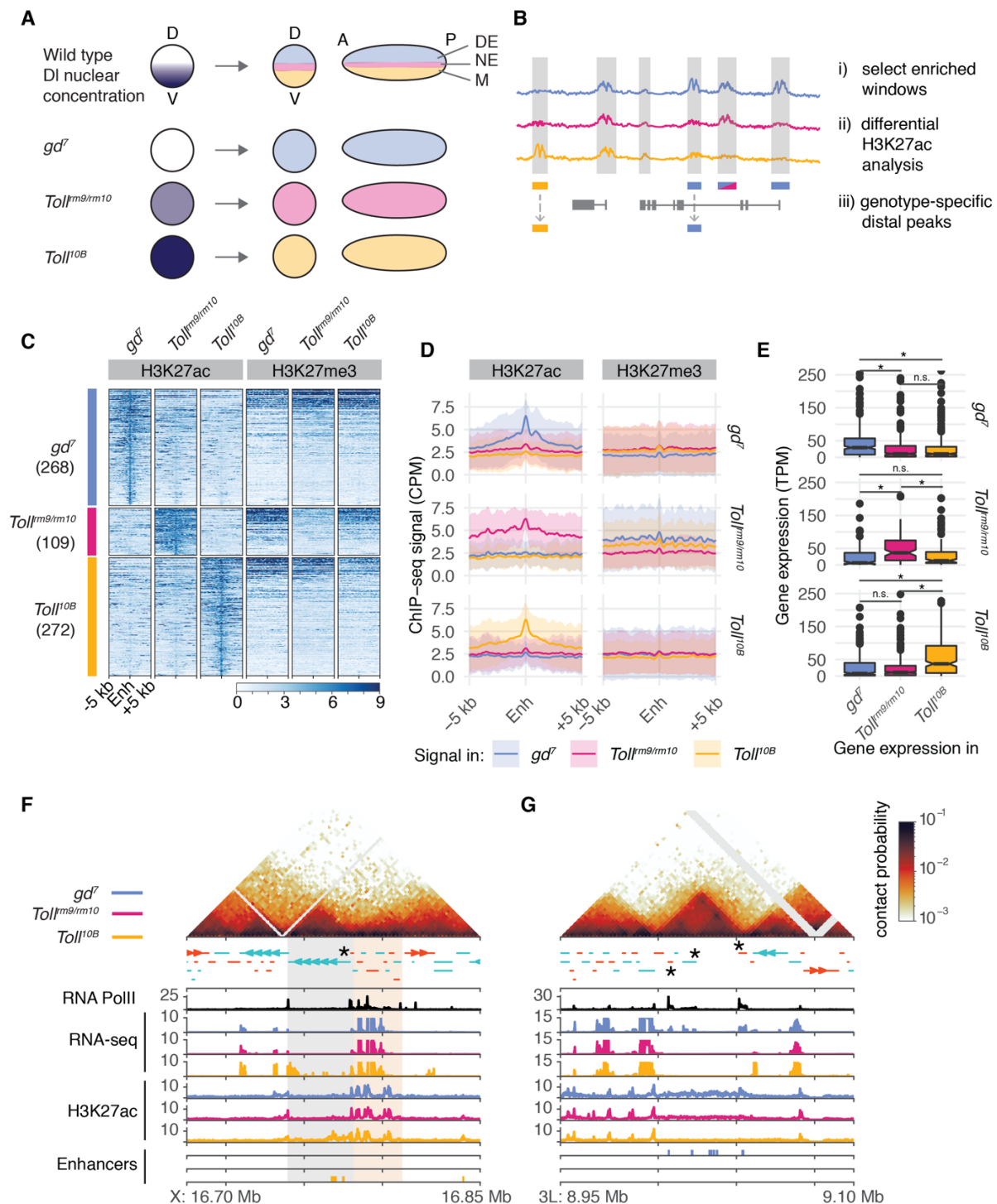


Figure 1. Identification of tissue-specific regulatory elements for dorsoventral patterning. A. Dorsoventral patterning of the *Drosophila* embryo is controlled by the nuclear concentration of Dl. High levels of nuclear Dl on the ventral side of the embryo produce mesoderm (ME, yellow), intermediate levels produce neuroectoderm (NE, pink), while nuclei without Dl will produce dorsal ectoderm (DE, blue). The *gd⁷*, *Toll^{m9/rm10}*, and *Toll^{10B}* maternal effect mutations lead to embryos with uniform levels of nuclear Dl, which produce only dorsal ectoderm, neuroectoderm, and mesoderm, respectively. D, dorsal; V, ventral; A, anterior; P, posterior. **B.** Schematic representation of the identification of putative tissue-specific enhancers. In order to identify regions with tissue-specific H3K27ac, regions with enriched H3K27ac ChIP-seq signal compared to global background (grey shaded areas) were selected and pairwise comparisons between genotypes (*gd⁷*, blue; *Toll^{m9/rm10}*, pink; *Toll^{10B}*, yellow) were performed using *csaw* (Lun and Smyth, 2014; Lun and Smyth, 2016). We selected regions that were enriched for H3K27ac in one genotype compared to both others, and

removed regions that overlapped a promoter. **C.** Heatmaps of H3K27ac and H3K27me3 ChIP-seq signal at putative tissue-specific enhancers in dorsoventral mutant embryos. **D.** Average H3K27ac and H3K27me3 ChIP-seq signal at putative tissue-specific enhancers in dorsoventral mutant embryos. Shaded area represents +/- one standard deviation from the mean. **E.** Expression of genes associated with putative tissue-specific enhancers in dorsoventral mutant embryos. Top, genes associated with *gd⁷*-specific H3K27ac peaks; middle, genes associated with *Toll^{rm9/rm10}*-specific H3K27ac peaks; bottom, genes associated with *Toll^{10B}*-specific H3K27ac peaks. Asterisks indicate Wilcoxon rank-sum test $p < 0.005$. **F, G.** Examples of chromatin organisation at dorsoventral patterning genes (*if*, **F**; *Doc1*, *Doc2*, *Doc3*, **G**; these genes are labelled with asterisks). Top, normalised Hi-C contact probability maps, 2kb resolution, from 3-4 hpf control embryos (Hug et al., 2017). Positive strand genes are shown in orange and negative strand genes are shown in blue. RNA PolII-Ser5p ChIP-seq from late nc14 control embryos is shown in black (Blythe and Wieschaus, 2015). RNA-seq and H3K27ac ChIP-seq from 2-4 hpf dorsoventral mutant embryos is shown in blue (*gd⁷*), pink (*Toll^{rm9/rm10}*), and yellow (*Toll^{10B}*) (Koenecke et al., 2016), this study). Tissue-specific putative enhancers identified as described above are shown as colour-coded bars. The grey shaded area in **F** represents a domain containing a developmentally regulated gene; the orange shaded area highlights a region containing housekeeping genes.

In this study, we use *Drosophila* DV patterning as a model system to investigate the relationship between tissue-specific gene regulation and 3D chromatin organisation. We focus on the cellular blastoderm stage, approximately 2-3 hours post fertilisation (hpf), which is coincident with nc14, establishment of chromatin organisation, and the onset of ZGA. We identify putative regulatory elements involved in DV patterning, and show that their target genes are developmentally regulated and have a distinct chromatin organisation compared to housekeeping genes. We find that while there are clear differences in chromatin state and overall gene expression between *gd⁷*, *Toll^{rm9/rm10}*, and *Toll^{10B}* embryos, there is still significant heterogeneity in gene expression at the single-cell level. However, these tissue-specific differences in chromatin state and gene expression are not associated with tissue-specific 3D chromatin organisation. Together, these results provide strong evidence that tissue-specific chromatin conformation is not required for tissue-specific gene expression. Rather, our findings indicate that the organisation of the genome into 3D chromatin domains acts as an architectural framework to facilitate correct regulation of gene expression during development.

RESULTS

Identification of regulatory elements and genes involved in dorsoventral patterning

In order to identify candidate tissue-specific regulatory elements involved in dorsoventral patterning, we carried out genome-wide differential peak identification for the active chromatin mark H3K27ac, using ChIP-seq data from 2-4 hpf embryos, including ChIP-seq data for *Toll^{rm9/rm10}* embryos generated in this study and for *gd⁷* and *Toll^{10B}* embryos from (Koenecke et al., 2016) (Fig. 1B). This identified 268 regions enriched for H3K27ac in *gd⁷* compared to both *Toll^{rm9/rm10}* and *Toll^{10B}*, 109 regions specifically enriched in *Toll^{rm9/rm10}*, and 272 regions specifically enriched in *Toll^{10B}* (Fig. 1C). These putative tissue-specific enhancers are depleted for the repressive chromatin mark H3K27me3 in the tissue in which they are enriched for H3K27ac (Fig. 1C, 1D), providing further evidence for their tissue specificity. By requiring significant enrichment in one genotype compared to both other genotypes, we select a highly stringent set of regions with tissue-specific increases in H3K27ac. These putative enhancers overlap with genomic regions that have been shown to drive expression in the

expected regions of the embryo (Fig. S1D, E). Next, we assigned putative enhancers to target genes using a combination of gene expression data, linear genomic proximity, and chromatin conformation data (see Methods), and verified that genes assigned to tissue-specific candidate enhancers have significantly higher expression in the tissue where the enhancer is active (Fig. 1E, all $p < 0.001$). We conclude that the identified regions represent a stringent set of candidate enhancers associated with regulation of dorsoventral patterning.

Developmentally regulated genes have a distinct regulatory landscape compared to housekeeping genes

We next assessed the chromatin conformation landscape around these tissue-specific regulatory elements and their target genes. Using Hi-C data from 3-4 hpf *Drosophila* embryos (Hug et al., 2017), we observed that dorsoventral patterning genes and their regulatory elements are located within the same self-interacting domains (Fig. 1F, grey shaded region; Fig. 1G). These domains are larger than domains not associated with developmentally regulated genes (Fig. S1A, B) (mean size 98 kb compared to 67 kb, $p = 3.47 \times 10^{-14}$). This is in contrast to housekeeping genes, which are enriched at the boundaries between domains and in small domains (Hug et al., 2017; Sexton et al., 2012; Ulianov et al., 2016) (Fig. 1F, orange shaded region). In addition, domains containing developmentally regulated genes are significantly more likely to overlap with the large regions of high non-coding sequence conservation known as Genomic Regulatory Blocks (Engström et al., 2007; Harmston et al., 2017b; Kikuta et al., 2007) (Fig. S1C). These results are robust to different definitions of tissue-specific enhancers (Fig. S2) and emphasize the distinct organisation of developmentally regulated and housekeeping genes in the *Drosophila* genome.

Single-cell expression analysis reveals heterogeneity in gene expression within dorsoventral mutant embryos

While the *gd⁷*, *Toll^{rm9/rm10}*, and *Toll^{10B}* maternal effect mutants have long been used as models to analyse tissue-specific regulation during dorsoventral patterning (Hong et al., 2008; Stathopoulos et al., 2002; Zeitlinger et al., 2007), the extent of cell fate conversion at the single-cell level within these embryos is unknown. Anterior-posterior patterning mechanisms are still active and RNA *in situ* hybridisation experiments suggest that cell fate conversion may be incomplete (Stathopoulos and Levine, 2002). In order to assess the heterogeneity of gene expression and cell identities within these embryos, we carried out single-cell gene expression analysis using the 10X Genomics Chromium platform. We used embryos at 2.5-3.5 hpf to target the late cellular blastoderm stage when dorsoventral patterning has been established (Reeves et al., 2012; Sandler and Stathopoulos, 2016). Clustering of single-cell expression profiles from wild type and mutant embryos showed good concordance with bulk RNA-seq (Fig. S3A) and identified 15 clusters representing different cell identities within the embryo (Fig. 2A, Fig. S3C). Two of these represent cells from embryos in the earlier stages of cellularisation, due to the timed collection. Other clusters represent mesoderm, ectoderm, amnioserosa, and terminal regions of the embryo, in addition to more specific cell populations such as pole cells, haemocytes, and trachea precursors. A full list of clusters and cluster marker genes is available in Table S2. Visualisation of single-cell gene expression profiles from *gd⁷*, *Toll^{rm9/rm10}*, and *Toll^{10B}* embryos revealed that specific clusters are depleted in these mutant embryos (Fig. 2B, Fig. S3B-C). Cells from clusters representing mesoderm cell fates are almost completely absent in *gd⁷* and *Toll^{rm9/rm10}* embryos, and subsets of ectoderm cells are missing in each of the mutants (Fig. 2B, Fig. S3B-C).

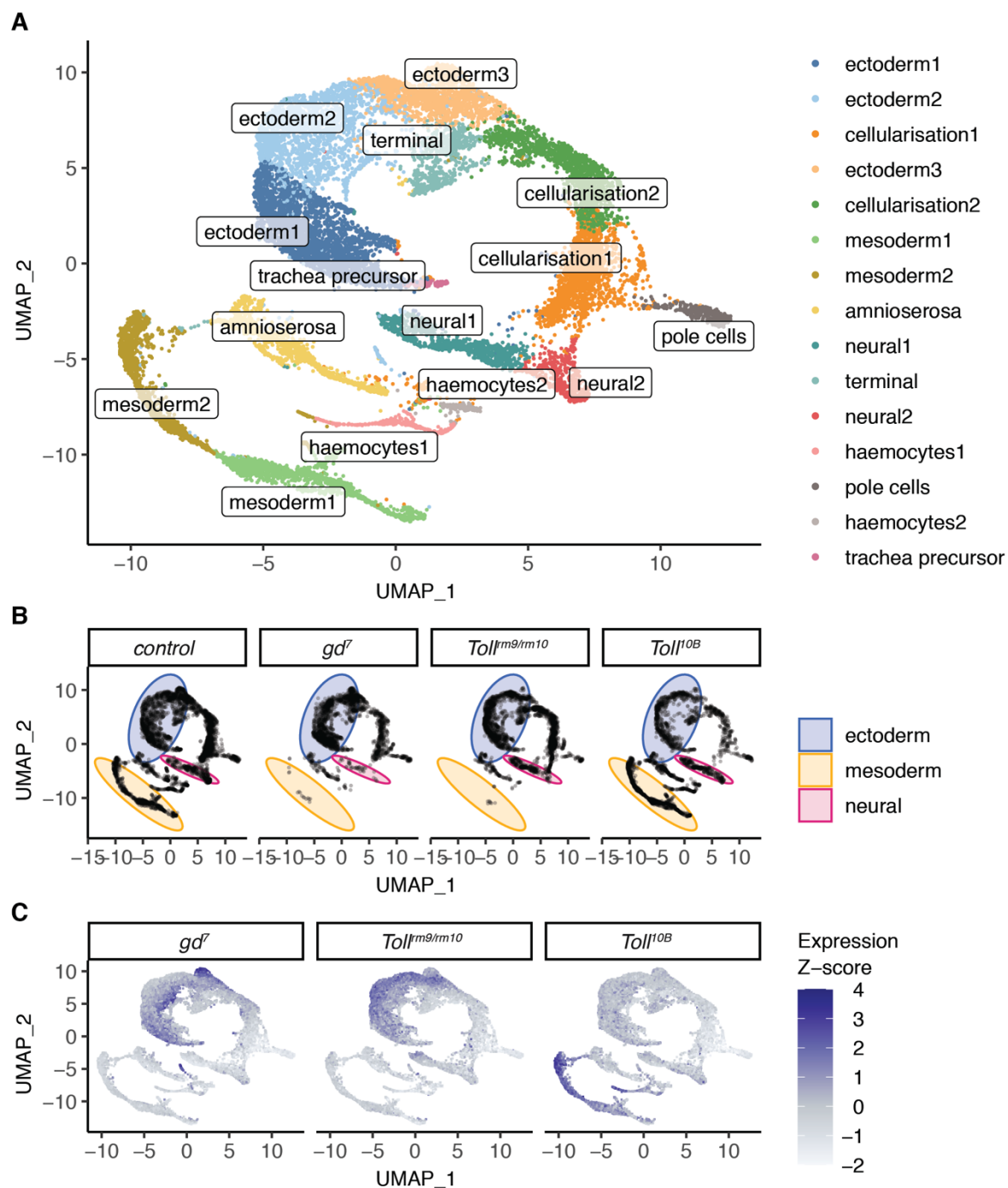


Figure 2. Single-cell RNA-seq analysis of gene expression during dorsoventral patterning. A. Clustering of single-cell gene expression profiles from 2.5-3.5 hpf embryos reveals clusters corresponding to distinct cell populations. **B.** Single cell gene-expression profiles separated by cell origin (control, *gd⁷*, *Toll^{rm9/rm10}*, or *Toll^{10B}* embryos). Certain clusters are depleted in the dorsoventral mutant embryos. Blue, area of graph corresponding to ectoderm clusters; pink, neural; yellow, mesoderm. **C.** Single-cell expression of genes associated with putative tissue-specific enhancers. Colour is mapped to the Z-score of average expression of genes assigned to the groups of tissue-specific enhancers identified in Figure 1.

In order to further dissect the ectoderm clusters and identify cells corresponding to dorsal ectoderm and neuroectoderm, we visualised expression of genes assigned to tissue-specific enhancers (Fig. 2C) and known dorsal ectoderm, neuroectoderm, and mesoderm marker genes (Fig. S3C, (Karaikos et al., 2017)). This revealed that the ectoderm clusters contain distinct subpopulations of cells expressing dorsal ectoderm markers and neuroectoderm markers. These subpopulations correspond to the regions of the cell distribution which are depleted in *Toll^{rm9/rm10}* and *gd⁷* respectively (Fig. 2B, compare distributions in the ‘ectoderm’ region). Therefore, while the mutant embryos evidently still consist of a mixture of cell identities, certain cell fates are lost. Overall, these results highlight that these embryos have a significant level of cell-to-cell heterogeneity. Importantly, the loss of specific cell fates combined with the tissue-specific enhancer usage shown above supports the use of these embryos to model dorsoventral patterning perturbations.

Major features of chromatin organisation are conserved across tissues

We next asked how differential usage of regulatory elements and differential gene expression relates to chromatin conformation during dorsoventral patterning. To do so, we generated Hi-C datasets for *gd⁷*, *Toll^{rm9/rm10}*, *Toll^{l0B}*, and control embryos at the cellular blastoderm stage, at 2kb resolution (Table S3).

Systematic comparison of the Hi-C datasets across genotypes revealed that on average, chromatin conformation is similar across datasets (Fig. 3). Saddle plots reveal similar strength of compartmentalisation in control and *gd⁷*, *Toll^{rm9/rm10}*, and *Toll^{l0B}* embryos (Fig. 3A, B). We next analysed overall self-interacting domain strength using domains identified in control 3-4 hpf embryos as a reference (Fig. 3C). While domain strength is weaker in cellular blastoderm embryos than at 3-4 hpf, the strength is similar across

all genotypes, suggesting that the vast majority of domains and domain boundaries are present in all tissues. We obtained similar conclusions when we examined chromatin loop strengths, using loops from Kc167 cells (Cubéñas-Potts et al., 2017) as a reference (Fig. 3D, E), indicating that loops are conserved across tissues. Finally, we analysed genome-wide contact probability decay with distance ($P(s)$). A shallow slope at distances < 100 kb reflects local chromatin compaction into domains, while the flattening of the slope around separation distances of 1 Mb indicates compartment formation ((Gassler et al., 2017; Naumova et al., 2013; Schwarzer et al., 2017), Fig. 3F). We also examined the derivative of $P(s)$, since this can highlight differences in the strength of domain formation (Fig. 3G) (Gassler et al., 2017). These analyses reveal differences in these profiles at distances > 5 Mb, which correspond to genomic rearrangements on balancer chromosomes present in a subset of *Toll^{rm9/rm10}* and *Toll^{l0B}* embryos (Fig. S4). Combined, our results demonstrate that overall genome organisation at the level of compartments, domains, and chromatin loops is similar across tissues in cellular blastoderm embryos.

Chromatin conformation at developmentally regulated genes is similar across tissues despite differences in gene expression and chromatin state

In order to systematically assess chromatin conformation across the genome and identify regions with differences, we used CHES (Galan, Machnik et al., in revision), a novel approach for differential chromatin conformation detection based on computer vision techniques. Briefly, Hi-C submatrices are compared genome-wide between pairs of datasets to produce a similarity score and a signal to noise ratio for each pair of genomic windows (see Methods). Using this approach, we compared control and mutant embryos at the cellular blastoderm stage at 5kb resolution and with a 500 kb window

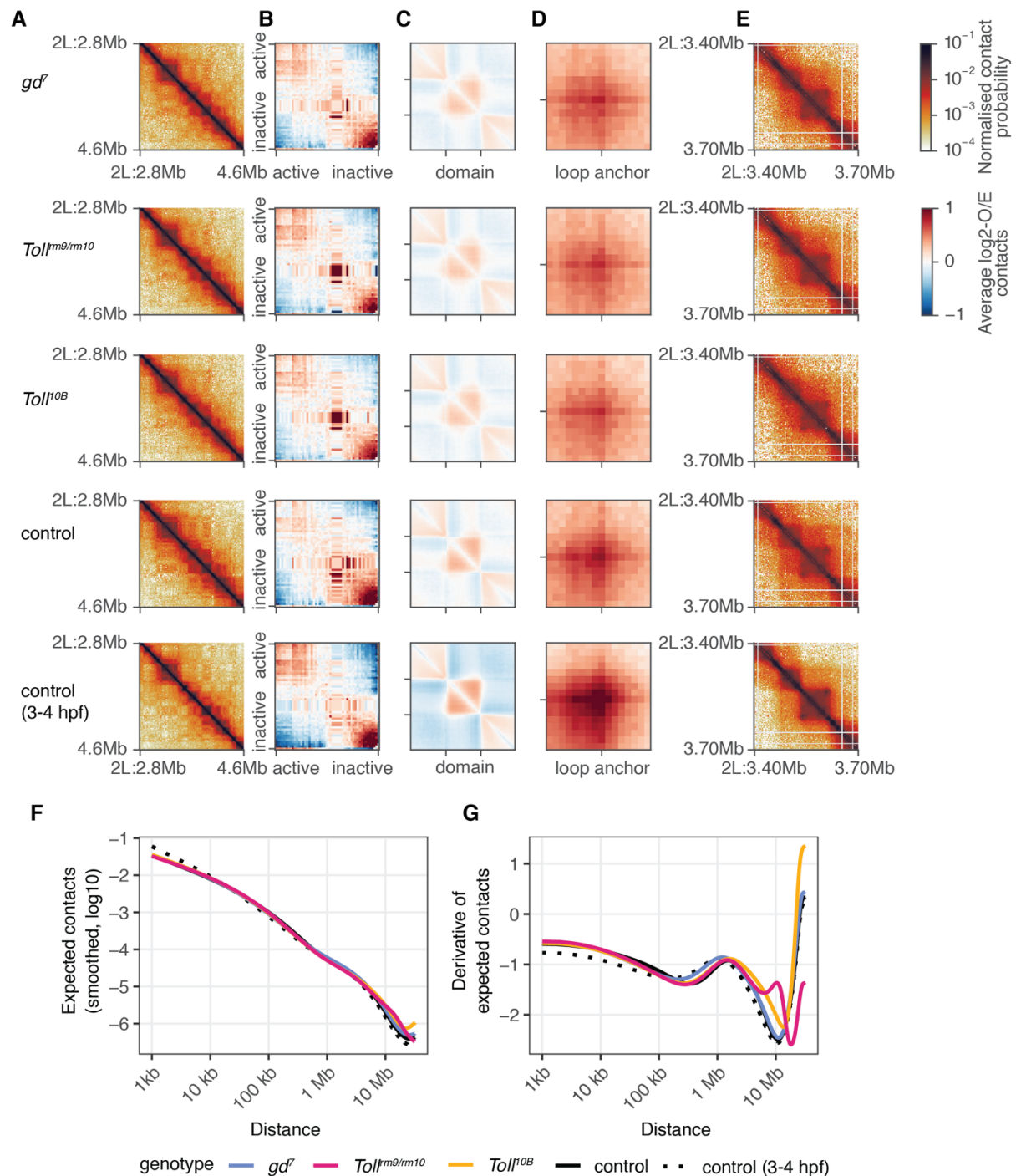


Figure 3. Global chromatin conformation along the dorsoventral axis. A. Chromatin conformation for a 1.8 Mb region of chromosome 2L in *gd⁷*, *Toll^{m9/rm10}*, *Toll^{10B}*, and control embryos at the cellular blastoderm stage, and for control embryos at 3-4 hpf using data from (Hug et al., 2017). **B.** ‘Saddle-plot’ representing genome-wide average chromatin compartmentalisation. Active regions tend to interact with other active regions (top left), while inactive regions interact with other inactive regions (bottom right). **C.** Aggregate analysis of domains identified using Hi-C data from 3-4 hpf

embryos at 2kb resolution. **D.** Aggregate analysis of chromatin loops identified in (Cubebñas-Potts et al., 2017). **E.** Chromatin conformation for a 300 kb region of chromosome 2L, showing domains and a loop. **F.** Average contact probability at different distances for control (black), *gd⁷* (blue), *Toll^{rm9/rm10}* (pink), and *Toll^{10B}* (yellow) embryos. **G.** The derivative of the expected contact probability by distance, highlighting differences between samples at far-cis distances due to the presence of balancer chromosomes in the *Toll^{rm9/rm10}* and *Toll^{10B}* embryos.

size. As a reference, we compared control cellular blastoderm stage data from this study with Hi-C data from nc14 embryos from (Hug et al., 2017). We subtracted this reference score from the score for each control-mutant comparison in order to identify regions with differences in genome organisation between control and mutant embryos. This analysis revealed that most regions across the genome do not display significant differences in 3D chromatin organisation between *gd⁷*, *Toll^{rm9/rm10}*, and *Toll^{10B}* embryos (Fig. 4A, Fig. S4, Fig. S5). This agreed with visual examinations of control-mutant difference matrices (Fig. 4E). The subset of regions that do display strong changes in chromatin organisation between genotypes can be attributed to genomic rearrangements present on balancer chromosomes in a subset of the *Toll^{rm9/rm10}* and *Toll^{10B}* embryos, rather than correlating with the locations of genes that are differentially expressed (Fig. 4A-B, Fig. S4-S6). To further investigate the relationship between changes in chromatin conformation and gene expression, we analysed CHES similarity scores in windows containing genes that are differentially expressed in pairwise comparisons between the mutant embryos (Koencke et al., 2016) compared to other genomic windows (Fig. 4C). This revealed a lack of association between differential gene expression and differential chromatin structure genome wide. To further validate these observations, we visually assessed genome organisation in *gd⁷*, *Toll^{rm9/rm10}*, and *Toll^{10B}* embryos at known differentially expressed dorsoventral patterning genes (Fig. 4E, Fig. S6). Examination of chromatin conformation and chromatin data at these regions does not indicate any

differences in domain organisation, boundary formation, or loop formation. For example, the *Doc1*, *Doc2* and *Doc3* genes, which are highly expressed in *gd⁷* and required for amnioserosa differentiation and dorsolateral ectoderm patterning (Reim, 2003), lie in a well-insulated domain that contains multiple *gd⁷*-specific putative enhancers and is enriched for H3K27me3 in all three mutants at 2-4 hpf, although to a lesser extent in *gd⁷* (Fig. 4E). There is no evidence of changes in the insulation of this domain in *gd⁷*, where the genes are active, compared to the other datasets, nor is there any change in its internal structure such as changes in interactions between enhancers and the target gene promoters. Similar conclusions were obtained from examining additional loci, including the *pnr* locus (expressed in *gd⁷*) and the *NetA/NetB*, *if*, and *sna* loci which are active in *Toll^{10B}* (Fig. S6). Together, these results suggest that tissue-specific gene expression and enhancer activity do not necessarily involve changes in domain organisation.

Finally, traditional models of gene regulation by enhancers predict that interactions between regulatory elements and their target promoters increase upon tissue-specific gene expression (Alberts et al., 2014). While such changes were not apparent upon visual examination (Fig. 4D, Fig. S6), we performed aggregate analysis to systematically determine whether subtle changes in interaction strength manifest at these loci. This revealed that there is no significant increase in interaction frequency between enhancers and their target promoters in the tissue in which the enhancer is active (Fig. S7).

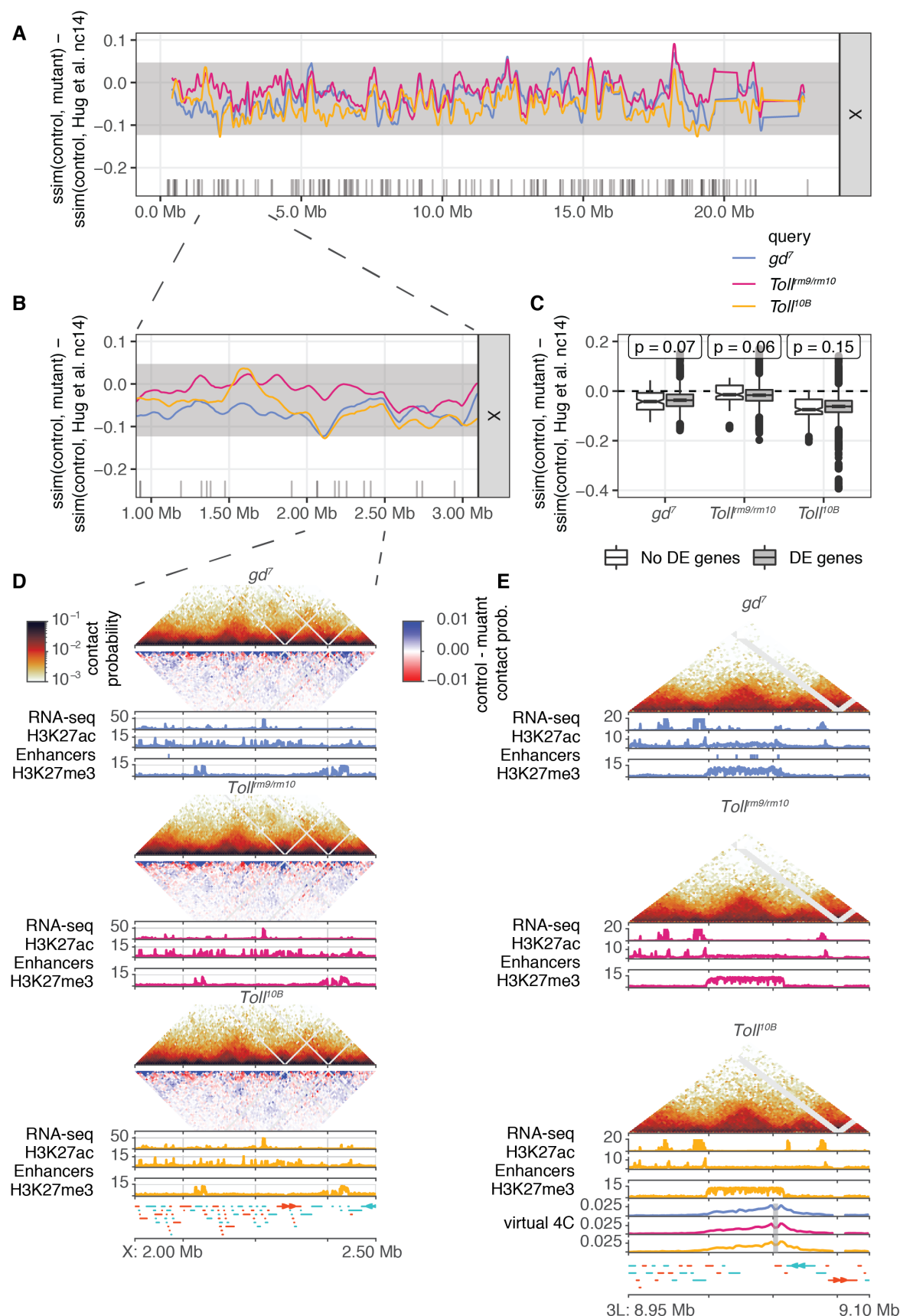


Figure 4. Chromatin conformation is not affected by tissue-specific gene expression. A. CHES (Galan, Machnik et al., in revision) similarity scores were calculated between mutant and control

embryo Hi-C datasets, using 5kb resolution and a 500 kb window size. As a reference, similarity scores were calculated between control embryo Hi-C data and the nc14 Hi-C data from (Hug et al., 2017). The difference between this reference similarity score and the similarity scores between mutant and control embryos for chromosome X is shown (blue, *gd⁷*; pink, *Toll^{rm9/rm10}*; yellow, *Toll^{10B}*). Similarity score differences around zero represent regions where chromatin conformation is similar between control and mutant, while negative values represent regions where there are greater differences between control and mutant than between control and the Hug et al. data. Shaded area represents similarity score differences within two standard deviations of the genome-wide mean. Grey ticks represent the positions of genes that are differentially expressed between dorsoventral mutant embryos (Koenecke et al., 2016). Other chromosomes are shown in Figure S4. **B.** Example of differences in CHES similarity scores for a 2 Mb region on chromosome X. Differences in similarity score do not strongly correlate with the positions of differentially expressed genes. Hi-C data for a subset of this region is shown in **D.** **C.** Boxplots of differences in similarity score for genomic windows with or without genes that are differentially expressed (DE) in pairwise comparisons between dorsoventral mutant embryos (Koenecke et al., 2016). There are no significant differences between windows with and without DE genes (Wilcoxon rank sum test, all $p > 0.05$). **D.** Tissue-specific chromatin data for a subset of the region on chromosome X shown in **C** above. For each genotype, top, normalised Hi-C contact probability maps and Hi-C difference maps at 5kb resolution. Hi-C difference maps are calculated as *contact probability in control* – *contact probability in mutant*; red indicates regions with increased contact probability in embryos of the mutant genotype and blue indicates decreased contact probability. Bottom, RNA-seq (Koenecke et al., 2016), H3K27ac and H3K27me3 ChIP-seq data ((Koenecke et al., 2016), this study). Tissue-specific putative enhancers identified as described above are shown as colour-coded bars beneath the corresponding H3K27ac ChIP-seq track. Lower panel, gene annotations. Positive-strand genes are shown in orange and negative strand genes are shown in blue. See also additional example loci in Figure S5. **E.** Tissue-specific chromatin data for the region around the *Doc1*, *Doc2* and *Doc3* genes shown in Figure 1E. For each genotype, top, normalised Hi-C contact probability maps, at 2kb resolution; middle, RNA-seq data (Koenecke et al., 2016); bottom, H3K27ac and H3K27me3 ChIP-seq data ((Koenecke et al., 2016), this study). Tissue-specific putative enhancers identified as described above are shown as colour-coded bars beneath the corresponding H3K27ac ChIP-seq track. Bottom, “virtual 4C” tracks for each genotype representing interactions of a 2kb region around the promoter of *Doc1*; gene annotations. Positive-strand genes are shown in orange and negative strand genes are shown in blue. See also additional example loci in Figure S6.

Taken together, our results provide strong evidence for the independence of tissue-specific gene expression and chromatin conformation during dorsoventral patterning.

DISCUSSION

Our results indicate that despite significant differences in chromatin state and gene expression between tissues, cell type-specific gene regulation does not require cell type-specific chromatin conformation. Nevertheless, developmentally regulated genes and enhancers are organised into chromatin domains. We suggest that this organisation plays a permissive role to

facilitate the precise regulation of developmental genes.

Maternal effect mutations in the Toll signalling pathway lead to embryos that lack the usual patterning of the DV axis (Stein and Stevens, 2014). These embryos have long been used as a system to study the specification of mesoderm (*Toll^{10B}*), neuroectoderm (*Toll^{rm9/rm10}*), and dorsal ectoderm (*gd⁷*) cell fates and the regulation of tissue-specific gene expression (Boija and Mannervik, 2016; Koenecke et al., 2016; Koenecke et al., 2017; Stathopoulos et al., 2002; Zeitlinger et al., 2007). However, these embryos are still under the influence of anterior-posterior patterning signals and

do not show completely uniform cell identities (Stathopoulos et al., 2002). We sought to investigate heterogeneity of cell identity at the single-cell level by using single-cell gene expression profiling. This revealed that certain cell types are indeed maintained in all three Toll pathway mutants, including pole cells and other terminal region cell identities, haemocytes, and trachea precursor cells (Fig. 2). However, heterogeneity of gene expression is reduced in the mutants, as shown by the loss of cells assigned to mesoderm clusters in *gd⁷* and *Toll^{rm9/rm10}* embryos, and the depletion of ectoderm subsets in each of the mutants. These datasets showcase the advantages of measuring cellular heterogeneity at the single cell level and will provide a useful resource for further characterisation of these embryos and investigation of the regulation of DV patterning.

Although the *gd⁷*, *Toll^{rm9/rm10}*, and *Toll^{l^{OB}}* embryos still have heterogeneous gene expression profiles, nevertheless there are clear differences in chromatin state and overall gene expression between them (Boija and Mannervik, 2016; Koenecke et al., 2016; Stathopoulos et al., 2002; Zeitlinger et al., 2007). We expanded on previous studies by identifying putative enhancers specific to neuroectoderm in addition to dorsal ectoderm and mesoderm. This allowed the identification of tissue-specific putative enhancer-gene pairs, which correspond well with known dorsoventral patterning enhancers and genes that are differentially expressed across the dorsoventral axis. These regulatory elements and their target genes are located inside chromatin domains, distinct from the enrichment of housekeeping genes at domain boundaries (Cubéñas-Potts et al., 2017; Hou et al., 2012; Hug et al., 2017; Ramírez et al., 2018; Sexton et al., 2012; Ulianov et al., 2016). This is in line with previous results that suggest that 3D chromatin domains act as regulatory

domains (Despang et al., 2019; Harmston et al., 2017b; Ibrahim and Mundlos, 2020; Le Dily and Beato, 2015; Symmons et al., 2014).

Previous studies have produced conflicting results about whether tissue-specific enhancer-promoter interactions are correlated with tissue-specific activation of gene expression. Studies using 3C approaches have found evidence of enriched enhancer-promoter interactions (Hsieh et al., 2020; Jin et al., 2013; Krietenstein et al., 2020; Li et al., 2012; Sanyal et al., 2012), which may precede (Cruz-Molina et al., 2017; Ghavi-Helm et al., 2014) or correlate with (Bonev et al., 2017; Oudelaar et al., 2020) transcriptional activation. However, several recent studies in mammals as well as *Drosophila* provide evidence that stable enhancer-promoter contacts are not required for gene activation (Alexander et al., 2019; Benabdallah et al., 2019; Chen et al., 2018; Heist et al., 2019; Mir et al., 2018). Our results indicate stable chromatin conformation across tissues during *Drosophila* ZGA, and we found no evidence for widespread enrichment of interactions between enhancers and their target promoters. This is in line with “transcriptional hub” models, in which transient or indirect contacts with a regulatory element are sufficient to activate transcription (Furlong and Levine, 2018; Lim et al., 2018; Mir et al., 2018; Tsai et al., 2017; Yokoshi and Fukaya, 2019), such as through the formation of phase-separated condensates (Cho et al., 2018; Chong et al., 2018; Sabari et al., 2018). While a subset of long-range enhancer-promoter pairs do form stable interactions that are enriched above local background (Ghavi-Helm et al., 2014), these are not likely to be the primary mechanism of promoter regulation during *Drosophila* development. Many stable loops in the *Drosophila* genome are instead associated with Polycomb repression (Eagen et al., 2017; Ogiyama et al., 2018).

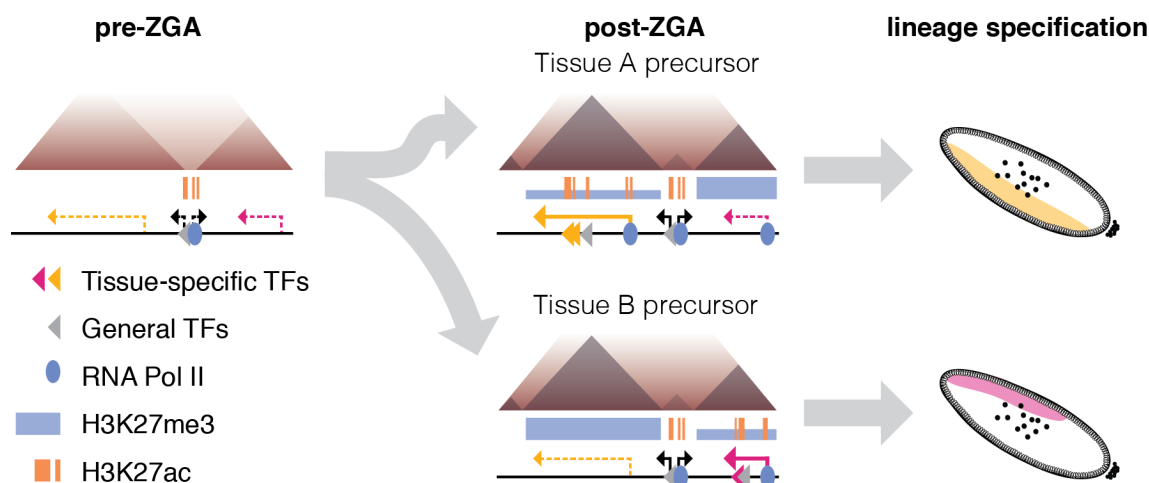


Figure 5. Model of the relationship between chromatin conformation and developmentally regulated gene expression. Before ZGA, the genome is unstructured (left), with domain boundaries appearing at a subset of regions associated with binding of RNA PolII and Zelda. Chromatin domains are established at ZGA, and domain structure is the same across tissues with different gene expression and transcription factor binding (middle). Differential activity of regulatory elements in the context of the same chromatin conformation leads to different patterns of gene expression in the developing embryo (right). Thick and thin blue bars represent high and low levels of H3K27me3 respectively; dashed lines represent inactive genes while solid lines represent actively transcribed genes.

Previous studies have also not provided conclusive evidence whether changes in domain and compartment organisation are a cause or consequence of changes in chromatin and transcriptional activity. Some studies suggested that domains are widely conserved across different tissues and even different species (Dixon et al., 2012; Schmitt et al., 2016; Vietri Rudan et al., 2015). However, there are also numerous examples of changes in domain structures across cell types and differentiation time points (Bonev et al., 2017; Joshi et al., 2015; Kragesteen et al., 2018; Kruse et al., 2019; Le Dily et al., 2014). However, it is important to note that detection of domains and boundaries in different samples may be confounded by sequencing depth, analysis resolution, noise, and the presence of so-called “sub-domain” structure. In *Drosophila*, Hi-C maps from anterior and posterior embryo halves showed no differences (Stadler et al., 2017), and it has been proposed that active chromatin, especially at broadly expressed genes, is responsible for partitioning the genome into

domains (Rowley et al., 2017; Ulianov et al., 2016). Rowley et al. (Rowley et al., 2017) proposed that compartmentalisation of active and inactive chromatin, at the level of individual genes, underlies the formation of insulated chromatin domains. This model predicts that when a developmentally regulated gene is active, its domain would merge with, or have increased interactions with, neighbouring domains containing active genes, for example broadly expressed housekeeping genes. In contrast to these predictions, we find that domain organisation is conserved across tissues, even in cases where there are significant changes in the local chromatin state and gene expression (Fig. 3 and Fig. S4). This suggests that, similarly to mammalian domain architecture, additional factors such as insulator proteins modulate domain organisation in *Drosophila* (Mateo et al., 2019; Moretti et al., 2020).

Together, our results indicate that differential chromatin organisation is not a necessary feature of cell-type specific gene

expression. We propose that chromatin organisation into domains instead provides a scaffold or framework for the regulation of developmental genes during and after the activation of zygotic gene expression (Fig. 5, left and middle). This increases the frequency of contacts between regulatory elements and their target genes, as in models where enhancers scan the domain for their target promoter, and allows for cell type-specific activation of gene expression (Fig. 5, right). Feedback effects such as downstream modification of chromatin state and additional mechanisms including looping between Polycomb-bound elements and segregation of active and inactive chromatin then act as layers on top of the initially established domain structure.

Author Contributions

Conceptualisation: E.I.-S. and J.M.V.; Formal analysis: E.I.-S.; Funding acquisition: E.I.-S., M.M., and J.M.V.; Investigation: E.I.-S. and R.V.; Resources: R.V. and M.M.; Supervision: M.M. and J.M.V.; Visualisation: E.I.-S.; Writing – original draft: E.I.-S.; Writing – review and editing: E.I.-S., R.V., M.M., and J.M.V.

REFERENCES

- Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K. and Walter, P. (2014). *Molecular Biology of the Cell*. W. Norton & Company.
- Alexander, J. M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., Huang, B., Lomvardas, S. and Weiner, O. D. (2019). Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *Elife* 8, 1–42.
- Alles, J., Karaikos, N., Praktiknjo, S. D., Grosswendt, S., Wahle, P., Ruffault, P. L., Ayoub, S., Schreyer, L., Boltengagen, A., Birchmeier, C., et al. (2017). Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol.* 15, 1–14.
- Amezquita, R. A., Lun, A. T. L., Becht, E., Carey, V. J., Carpp, L. N., Geistlinger, L., Martini, F., Rue-Albrecht, K., Risso, D., Soneson, C., et al. (2019). Orchestrating single-cell analysis with Bioconductor. *Nat. Methods* 1–32.
- Bartman, C. R., Hsu, S. C., Hsiung, C. C.-S., Raj, A. and Blobel, G. A. (2016). Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Mol. Cell* 62, 237–247.
- Benabdallah, N. S., Williamson, I., Illingworth, R. S., Kane, L., Boyle, S., Sengupta, D., Grimes, G. R., Therizols, P. and Bickmore, W. A. (2019). Decreased Enhancer-Promoter Proximity

Funding

Work in the Vaquerizas lab is supported by the Max Planck Society, the Deutsche Forschungsgemeinschaft (DFG) Priority Programme SPP2202 Spatial Genome Architecture in Development and Disease (Project Ref VA 1456/1) and the Medical Research Council, UK. E.I.-S. was supported by a postdoctoral fellowship from the Alexander von Humboldt-Stiftung. Work in the Mannervik lab is supported by the Swedish Research Council (Vetenskapsrådet) and the Swedish Cancer Society (Cancerfonden).

Acknowledgements

We thank the core facility at Novum, BEA – Bioinformatics and Expression Analysis, which is supported by the board of research at the Karolinska Institute and the research committee at the Karolinska hospital for help with sequencing. We are grateful to Christine Rushlow for critical reading of the manuscript, and members of the Vaquerizas lab for helpful discussions and feedback.

Competing Interests

The authors declare no competing or financial interests.

- Accompanying Enhancer Activation. *Mol. Cell* **76**, 473–484.e7.
- Blythe, S. A. and Wieschaus, E. F.** (2015). Zygotic genome activation triggers the DNA replication checkpoint at the midblastula transition. *Cell* **160**, 1169–1181.
- Blythe, S. A. and Wieschaus, E. F.** (2016). Establishment and maintenance of heritable chromatin structure during early *Drosophila* embryogenesis. *Elife* **5**, 1–21.
- Boija, A. and Mannervik, M.** (2016). Initiation of diverse epigenetic states during nuclear programming of the *Drosophila* body plan. *Proc. Natl. Acad. Sci.* **113**, 8735–8740.
- Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., Xu, X., Lv, X., Hugnot, J. P., Tanay, A., et al.** (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557–572.e24.
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R.** (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420.
- Chathoth, K. T. and Zabet, N. R.** (2019). Chromatin architecture reorganization during neuronal cell differentiation in *Drosophila* genome. *Genome Res.* **29**, 613–625.
- Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J. B. and Gregor, T.** (2018). Dynamic interplay between enhancer–promoter topology and gene activity. *Nat. Genet.* **50**, 1296–1303.
- Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V. and Cisse, I. I.** (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **4199**, eaar4199.
- Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G. M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., et al.** (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **2555**, eaar2555.
- Conway, J. R., Lex, A. and Gehlenborg, N.** (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940.
- Crane, E., Bian, Q., McCord, R. P., Lajoie, B. R., Wheeler, B. S., Ralston, E. J., Uzawa, S., Dekker, J. and Meyer, B. J.** (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244.
- Cruz-Molina, S., Respuela, P., Tebartz, C., Kolovos, P., Nikolic, M., Fueyo, R., van Ijcken, W. F. J., Grosveld, F., Frommolt, P., Bazzi, H., et al.** (2017). PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell Stem Cell* **20**, 689–705.e9.
- Cubeñas-Potts, C., Rowley, M. J., Lyu, X., Li, G., Lei, E. P. and Corces, V. G.** (2017). Different enhancer classes in *Drosophila* bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic Acids Res.* **45**, 1714–1730.
- Cusanovich, D. A., Reddington, J. P., Garfield, D. A., Daza, R. M., Aghamirzaie, D., Marco-Ferreres, R., Pliner, H. A., Christiansen, L., Qiu, X., Steemers, F. J., et al.** (2018). The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature* **555**, 538–542.
- Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., Dean, A. and Blobel, G. A.** (2012). Controlling Long-Range Genomic Interactions at a Native Locus by Targeted Tethering of a Looping Factor. *Cell* **149**, 1233–1244.
- Deng, W., Rupon, J. W., Krivega, I., Breda, L., Motta, I., Jahn, K. S., Reik, A., Gregory, P. D., Rivella, S., Dean, A., et al.** (2014). Reactivation of Developmentally Silenced Globin Genes by Forced Chromatin Looping. *Cell* **158**, 849–860.
- Despang, A., Schöpflin, R., Franke, M., Ali, S., Jerković, I., Paliou, C., Chan, W.-L., Timmermann, B., Wittler, L., Vingron, M., et al.** (2019). Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat. Genet.* **51**, 1263–1271.
- Díaz, N., Kruse, K., Erdmann, T., Staiger, A. M., Ott, G., Lenz, G. and Vaquerizas, J. M.** (2018). Chromatin

- conformation analysis of primary patient tissue using a low input Hi-C method. *Nat. Commun.* **9**, 4938.
- Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S. and Ren, B.** (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380.
- Eagen, K. P., Aiden, E. L. and Kornberg, R. D.** (2017). Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *Proc. Natl. Acad. Sci.* **114**, 8764–8769.
- Engström, P. G., Sui, S. J. H., Drivenes, Ø., Becker, T. S. and Lenhard, B.** (2007). Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Res.* **17**, 1898–1908.
- Feng, J., Liu, T., Qin, B., Zhang, Y. and Liu, X. S.** (2012). Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **7**, 1728–40.
- Flavahan, W. A., Drier, Y., Liao, B. B., Gillespie, S. M., Venteicher, A. S., Stemmer-Rachamimov, A. O., Suvà, M. L. and Bernstein, B. E.** (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114.
- Flyamer, I. M., Gassler, J., Imakaev, M., Brandão, H. B., Ulianov, S. V., Abdennur, N., Razin, S. V., Mirny, L. A. and Tachibana-Konwalski, K.** (2017). Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114.
- Franke, M., Ibrahim, D. M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., Kraft, K., Kempfer, R., Jerković, I., Chan, W.-L., et al.** (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265–269.
- Fudenberg, G. and Pollard, K. S.** (2019). Chromatin features constrain structural variation across evolutionary timescales. *Proc. Natl. Acad. Sci.* **116**, 2175–2180.
- Furlong, E. E. M. and Levine, M.** (2018). Developmental enhancers and chromosome topology. *Science* **361**, 1341–1345.
- Galan, S., Machnik, N., Kruse, K., Díaz, N., Marti-Renom, M. A. and Vaquerizas, J. M.** Quantitative comparison and automatic feature extraction for chromatin contact data.
- Gassler, J., Brandão, H. B., Imakaev, M., Flyamer, I. M., Ladstätter, S., Bickmore, W. A., Peters, J., Mirny, L. A. and Tachibana, K.** (2017). A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *EMBO J.* **36**, 3600–3618.
- Germain, P.-L., Sonrel, A. and Robinson, M. D.** (2020). pipeComp, a general framework for the evaluation of computational pipelines, reveals performant single-cell RNA-seq preprocessing tools. *bioRxiv* 2020.02.02.930578.
- Ghavi-Helm, Y., Klein, F. A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W. and Furlong, E. E. M.** (2014). Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100.
- Ghavi-Helm, Y., Jankowski, A., Meiers, S., Viales, R. R., Korbel, J. O. and Furlong, E. E. M.** (2019). Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* **51**, 1272–1282.
- Griffiths, J. A., Richard, A. C., Bach, K., Lun, A. T. L. and Marioni, J. C.** (2018). Detection and removal of barcode swapping in single-cell RNA-seq data. *Nat. Commun.* **9**, 2667.
- Hamm, D. C. and Harrison, M. M.** (2018). Regulatory principles governing the maternal-to-zygotic transition: insights from *Drosophila melanogaster*. *Open Biol.* **8**, 180183.
- Harmston, N., Ing-Simmons, E., Tan, G., Perry, M., Merkenschlager, M. and Lenhard, B.** (2017a). Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nat. Commun.* **8**.
- Harmston, N., Ing-Simmons, E., Tan, G., Perry, M., Merkenschlager, M. and Lenhard, B.** (2017b). Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nat. Commun.* **8**, 441.

- Heist, T., Fukaya, T. and Levine, M.** (2019). Large distances separate coregulated genes in living *Drosophila* embryos. *Proc. Natl. Acad. Sci.* **116**, 15062–15067.
- Hnisz, D., Weintraub, A. S., Day, D. S., Valton, A.-L., Bak, R. O., Li, C. H., Goldmann, J., Lajoie, B. R., Fan, Z. P., Sigova, A. A., et al.** (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458.
- Hong, J. W., Hendrix, D. A., Papatsenko, D. and Levine, M. S.** (2008). How the Dorsal gradient works: Insights from postgenome technologies. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20072–20076.
- Hou, C., Li, L., Qin, Z. S. and Corces, V. G.** (2012). Gene Density, Transcription, and Insulators Contribute to the Partition of the *Drosophila* Genome into Physical Domains. *Mol. Cell* **48**, 471–484.
- Hsieh, T. H. S., Cattoglio, C., Slobodyanyuk, E., Hansen, A. S., Rando, O. J., Tjian, R. and Darzacq, X.** (2020). Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol. Cell* **78**, 539–553.e8.
- Hug, C. B. and Vaquerizas, J. M.** (2018a). The Birth of the 3D Genome during Early Embryonic Development. *Trends Genet.* **34**, 903–914.
- Hug, C. B. and Vaquerizas, J. M.** (2018b). Generation of genome-wide chromatin conformation capture libraries from tightly staged early *drosophila* embryos. *J. Vis. Exp.* **2018**, 1–11.
- Hug, C. B., Grimaldi, A. G., Kruse, K. and Vaquerizas, J. M.** (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell* **169**, 216–228.e19.
- Ibn-Salem, J., Köhler, S., Love, M. I., Chung, H.-R., Huang, N., Hurles, M. E., Haendel, M., Washington, N. L., Smedley, D., Mungall, C. J., et al.** (2014). Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol.* **15**, 423.
- Ibrahim, D. M. and Mundlos, S.** (2020). The role of 3D chromatin domains in gene regulation: a multi-faceted view on genome organization. *Curr. Opin. Genet. Dev.* **61**, 1–8.
- Imakaev, M., Fudenberg, G., McCord, R. R. P., Naumova, N., Goloborodko, A., Lajoie, B. R., Dekker, J. and Mirny, L. a** (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003.
- Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., Yen, C.-A., Schmitt, A. D., Espinoza, C. a. and Ren, B.** (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290–294.
- Joshi, O., Wang, S. Y., Kuznetsova, T., Atlasi, Y., Peng, T., Fabre, P. J., Habibi, E., Shaik, J., Saeed, S., Handoko, L., et al.** (2015). Dynamic Reorganization of Extremely Long-Range Promoter-Promoter Interactions between Two States of Pluripotency. *Cell Stem Cell* **17**, 748–757.
- Karaïskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N. and Zinzen, R. P.** (2017). The *Drosophila* embryo at single-cell transcriptome resolution. *Science* **358**, 194–199.
- Kerpedjiev, P., Abdennur, N., Lekschas, F., Mccallum, C., Dinkla, K., Strobel, H., Luber, J. M., Ouellette, S. B., Azhir, A., Kumar, N., et al.** (2018). HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* **19**, 125.
- Kikuta, H., Laplante, M., Navratilova, P., Komisarczuk, A. Z., Engström, P. G., Fredman, D., Akalin, A., Caccamo, M., Sealy, I., Howe, K., et al.** (2007). Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res.* **17**, 545–555.
- Kim, D., Paggi, J. M., Park, C., Bennett, C. and Salzberg, S. L.** (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.*
- Knight, P. A. and Ruiz, D.** (2013). A fast algorithm for matrix balancing. *IMA J. Numer. Anal.* **33**, 1029–1047.
- Koenecke, N., Johnston, J., Gaertner, B., Natarajan, M. and Zeitlinger, J.** (2016). Genome-wide identification of *Drosophila* dorso-ventral enhancers by differential histone acetylation analysis. *Genome Biol.* **17**, 196.

- Koenecke, N., Johnston, J., He, Q., Meier, S. and Zeitlinger, J.** (2017). Drosophila poised enhancers are generated during tissue patterning with the help of repression. *Genome Res.* **27**, 64–74.
- Krasteen, B. K., Spielmann, M., Paliou, C., Heinrich, V., Schöpflin, R., Esposito, A., Annunziatella, C., Bianco, S., Chiariello, A. M., Jerković, I., et al.** (2018). Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. *Nat. Genet.* **50**, 1463–1473.
- Krefting, J., Andrade-Navarro, M. A. and Ibn-Salem, J.** (2018). Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol.* **16**, 87.
- Krietenstein, N., Abraham, S., Venev, S. V., Abdennur, N., Gibcus, J., Hsieh, T.-H. S., Parsi, K. M., Yang, L., Mahr, R., Mirny, L. A., et al.** (2020). Ultrastructural Details of Mammalian Chromosome Architecture. *Mol. Cell* **78**, 554–565.e7.
- Kruse, K., Diaz, N., Enriquez-Gasca, R., Gaume, X., Torres-Padilla, M.-E. and Vaquerizas, J. M.** (2019). Transposable elements drive reorganisation of 3D chromatin during early embryogenesis. *bioRxiv* 523712.
- Kruse, K., Hug, C. B. and Vaquerizas, J. M.** (2020). FAN-C: a feature-rich framework for the analysis and visualisation of C data. *bioRxiv* 10.1101/2020.02.03.932517.
- Kvon, E. Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J. O., Pagani, M., Schernhuber, K., Dickson, B. J. and Stark, A.** (2014). Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* **512**, 91–5.
- Langmead, B. and Salzberg, S. L.** (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–9.
- Le Dily, F. and Beato, M.** (2015). TADs as modular and dynamic units for gene regulation by hormones. *FEBS Lett.* **589**, 2885–2892.
- Le Dily, F., Baù, D., Pohl, A., Vicent, G. P., Serra, F., Soronellas, D., Castellano, G., Wright, R. H. G., Ballare, C., Filion, G., et al.** (2014). Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev.* **28**, 2151–2162.
- Lex, A., Gehlenborg, N., Strobel, H., Vuilleumot, R. and Pfister, H.** (2014). UpSet: Visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph.* **20**, 1983–1992.
- Li, H. and Durbin, R.** (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760.
- Li, G., Ruan, X., Auerbach, R. K., Sandhu, K. S., Zheng, M., Wang, P., Poh, H. M., Goh, Y., Lim, J., Zhang, J., et al.** (2012). Extensive Promoter-Centered Chromatin Interactions Provide a Topological Basis for Transcription Regulation. *Cell* **148**, 84–98.
- Lim, B., Heist, T., Levine, M. and Fukaya, T.** (2018). Visualization of Transvection in Living Drosophila Embryos. *Mol. Cell* **70**, 287–296.e6.
- Lott, S. E., Villalta, J. E., Schroth, G. P., Luo, S., Tonkin, L. A. and Eisen, M. B.** (2011). Noncanonical Compensation of Zygotic X Transcription in Early Drosophila melanogaster Development Revealed through Single-Embryo RNA-Seq. *PLoS Biol.* **9**, e1000590.
- Love, M. I., Anders, S. and Huber, W.** (2014). *Differential analysis of count data - the DESeq2 package.*
- Lun, A. T. L. and Smyth, G. K.** (2014). De novo detection of differentially bound regions for ChIP-seq data using peaks and windows: controlling error rates correctly. *Nucleic Acids Res.* **42**, e95–e95.
- Lun, A. T. L. and Smyth, G. K.** (2016). csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res.* **44**, e45–e45.
- Lun, A. T. L., McCarthy, D. J. and Marioni, J. C.** (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research* **5**, 2122.
- Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J. M., Laxova, R., et al.** (2015). Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161**, 1012–1025.

- Ma, J., He, F., Xie, G. and Deng, W. M.** (2016). Maternal AP determinants in the *Drosophila* oocyte and embryo. *Wiley Interdiscip. Rev. Dev. Biol.* **5**, 562–581.
- Mateo, L. J., Murphy, S. E., Hafner, A., Cinquini, I. S., Walker, C. A. and Boettiger, A. N.** (2019). Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54.
- McCarthy, D. J., Campbell, K. R., Lun, A. T. L. and Wills, Q. F.** (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 777–780.
- Meadows, L. A., Chan, Y. S., Roote, J. and Russell, S.** (2010). Neighbourhood Continuity Is Not Required for Correct Testis Gene Expression in *Drosophila*. *PLoS Biol.* **8**, e1000552.
- Mir, M., Stadler, M. R., Ortiz, S. A., Hannon, C. E., Harrison, M. M., Darzacq, X. and Eisen, M. B.** (2018). Dynamic multifactor hubs interact transiently with sites of active transcription in *Drosophila* embryos. *Elife* **7**, e40497.
- Moretti, C., Stévant, I. and Ghavi-Helm, Y.** (2020). 3D genome organisation in *Drosophila*. *Brief. Funct. Genomics* **19**, 92–100.
- Morgan, S. L., Mariano, N. C., Bermudez, A., Arruda, N. L., Wu, F., Luo, Y., Shankar, G., Jia, L., Chen, H., Hu, J.-F., et al.** (2017). Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nat. Commun.* **8**, 15993.
- Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B. R., Mirny, L. A. and Dekker, J.** (2013). Organization of the Mitotic Chromosome. *Science* **342**, 948–953.
- Nora, E. P., Goloborodko, A., Valton, A.-L., Gibcus, J. H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L. A. and Bruneau, B. G.** (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930–944.e22.
- Ogiyama, Y., Schuettengruber, B., Papadopoulos, G. L., Chang, J.-M. and Cavalli, G.** (2018). Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during *Drosophila* Development. *Mol. Cell* **71**, 73–88.e5.
- Oudelaar, A. M., Beagrie, R. A., Gosden, M., de Ornellas, S., Georgiades, E., Kerry, J., Hidalgo, D., Carrelha, J., Shivalingam, A., El-Sagheer, A. H., et al.** (2020). Dynamics of the 4D genome during in vivo lineage specification and differentiation. *Nat. Commun.* **11**, 2722.
- Özdemir, I. and Gambetta, M. C.** (2019). The Role of Insulation in Patterning Gene Expression. *Genes (Basel)*. **10**, 767.
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. and Kingsford, C.** (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419.
- Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. and Manke, T.** (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191.
- Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K. C., Grüning, B. A., Villavecies, J., Habermann, B., Akhtar, A. and Manke, T.** (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189.
- Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., et al.** (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680.
- Rao, S. S. P., Huang, S.-C., Hilaire, B. G. S., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K.-R., Sanborn, A. L., Johnstone, S. E., Bascom, G. D., Bochkov, I. D., et al.** (2017). Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305–320.e24.
- Reeves, G. T., Trisnadi, N., Truong, T. V., Nahmad, M., Katz, S. and Stathopoulos, A.** (2012). Dorsal-Ventral Gene Expression in the *Drosophila* Embryo Reflects the Dynamics and Precision of the Dorsal Nuclear Gradient. *Dev. Cell* **22**, 544–557.
- Reim, I.** (2003). The T-box-encoding Dorsocross genes function in amnioserosa development and the patterning of the dorsolateral germ band downstream of

- Dpp. *Development* **130**, 3187–3204.
- Robinson, M. D., McCarthy, D. J. and Smyth, G. K.** (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140.
- Rowley, M. J., Nichols, M. H., Lyu, X., Ando-Kuri, M., Rivera, I. S. M., Hermetz, K., Wang, P., Ruan, Y. and Corces, V. G.** (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol. Cell* **67**, 837–852.e7.
- Sabari, B. R., Dall’Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., Abraham, B. J., Hannett, N. M., Zamudio, A. V, Manteiga, J. C., et al.** (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **3958**, eaar3958.
- Sandler, J. E. and Stathopoulos, A.** (2016). Stepwise Progression of Embryonic Patterning. *Trends Genet.* **32**, 432–443.
- Sanyal, A., Lajoie, B. R., Jain, G. and Dekker, J.** (2012). The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113.
- Schmitt, A. D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C. L., Li, Y., Lin, S., Lin, Y., Barr, C. L., et al.** (2016). A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep* **17**, 2042–2059.
- Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N. A., Huber, W., H. Haering, C., Mirny, L., et al.** (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56.
- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A. and Cavalli, G.** (2012). Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell* **148**, 458–472.
- Soneson, C., Love, M. I. and Robinson, M. D.** (2016). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* **4**, 1521.
- Spielmann, M., Lupiáñez, D. G. and Mundlos, S.** (2018). Structural variation in the 3D genome. *Nat. Rev. Genet.* **19**, 453–467.
- Stadler, M. R., Haines, J. E. and Eisen, M. B.** (2017). Convergence of topological domain boundaries, insulators, and polytene interbands revealed by high-resolution mapping of chromatin contacts in the early Drosophila melanogaster embryo. *Elife* **6**, e29550.
- Stathopoulos, A. and Levine, M.** (2002). Dorsal Gradient Networks in the Drosophila Embryo. *Dev. Biol.* **246**, 57–67.
- Stathopoulos, A., Van Drenth, M., Erives, A., Markstein, M. and Levine, M.** (2002). Whole-genome analysis of dorsal-ventral patterning in the Drosophila embryo. *Cell* **111**, 687–701.
- Stein, D. S. and Stevens, L. M.** (2014). Maternal control of the Drosophila dorsal-ventral body axis. *Wiley Interdiscip. Rev. Dev. Biol.* **3**, 301–330.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P. and Satija, R.** (2019). Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21.
- Symmons, O., Uslu, V. V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L. and Spitz, F.** (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400.
- Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. and Prins, P.** (2015). Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034.
- Tsai, A., Muthusamy, A. K., Alves, M. R., Lavis, L. D., Singer, R. H., Stern, D. L. and Crocker, J.** (2017). Nuclear microenvironments modulate transcription from low-affinity enhancers. *Elife* **6**, e28975.
- Ulianov, S. V., Khrameeva, E. E., Gavrilov, A. A., Flyamer, I. M., Kos, P., Mikhaleva, E. A., Penin, A. A., Logacheva, M. D., Imakaev, M. V, Chertovich, A., et al.** (2016). Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains.

- Genome Res.* **26**, 70–84.
- Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D. T., Tanay, A. and Hadjur, S.** (2015). Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. *Cell Rep.* **10**, 1297–1309.
- Weintraub, A. S., Li, C. H., Zamudio, A. V, Sigova, A. A., Hannett, N. M., Day, D. S., Abraham, B. J., Cohen, M. A., Nabet, B., Buckley, D. L., et al.** (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171**, 1573-1588.e28.
- Wickham, H.** (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wieschaus, E.** (2016). Positional Information and Cell Fate Determination in the Early Drosophila Embryo. *Curr. Top. Dev. Biol.* **117**, 567–579.
- Williamson, I., Kane, L., Devenney, P. S., Flyamer, I. M., Anderson, E., Kilanowski, F., Hill, R. E., Bickmore, W. A. and Lettice, L. A.** (2019). Developmentally regulated Shh expression is robust to TAD perturbations. *Development* **146**, dev179523.
- Wolpert, L.** (2016). Positional Information and Pattern Formation. *Curr. Top. Dev. Biol.* **117**, 597–608.
- Wutz, G., Várnai, C., Nagasaka, K., Cisneros, D. A., Stocsits, R. R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, M. J., et al.** (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36**, 3573–3599.
- Yokoshi, M. and Fukaya, T.** (2019). Dynamics of transcriptional enhancers and chromosome topology in gene regulation. *Dev. Growth Differ.* **61**, 343–352.
- Yu, G., Wang, L.-G., Han, Y. and He, Q.-Y.** (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omi. A J. Integr. Biol.* **16**, 284–287.
- Zeitlinger, J., Zinzen, R. P., Stark, A., Kellis, M., Zhang, H., Young, R. a and Levine, M.** (2007). Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo. *Genes Dev.* **21**, 385–390.

METHODS

Drosophila stock maintenance and embryo collection

yw; eGFP-PCNA flies used as controls for Hi-C and the first scRNA-seq control experiment were kindly provided by S.A. Blythe and E. Wieschaus (Blythe and Wieschaus, 2016) and maintained on standard cornmeal-agar food. The *w¹¹¹⁸* flies used for the second scRNA-seq control experiment and the Toll mutant fly stocks *gd⁷/winscy hs-hid*, *Toll^{10B}/TM3 e Sb Ser/OR60* and *Toll^{rm9/rm10}/TM6 e Tb Sb* were grown on potato mash-agar food. All fly stocks were incubated at 25 °C with a 12-hour light/dark cycle.

The embryos representing presumptive dorsal ectoderm were collected from *gd⁷* homozygous flies. One day old larvae laid by *gd⁷/winscy hs-hid* were heat shocked for 1.5 hr at 37 °C twice with 24hr interval to eliminate *gd⁷* heterozygous animals. Embryos from *Toll^{10B}/TM3 e Sb Ser* or *Toll^{10B}/OR60* heterozygous females represented presumptive mesoderm. *Toll^{rm9}/Toll^{rm10}* trans-heterozygous females were used for collecting presumptive neuroectoderm embryos.

We adapted the fixation and sorting procedure described in (Blythe and Wieschaus, 2015) for *in situ* Hi-C (Hug and Vaquerizas, 2018b; Rao et al., 2014). Following a pre-collection period of at least one hour, fly embryos were collected on yeasted 0.4 % acetic acid agar plates or apple juice plates at 25 °C. After 1 hour of collection, the embryos on the plate were incubated at 25 °C for 2 hours. Embryos were dechorionated for 2 min in 2.6 % sodium hypochlorite, rinsed with water, and transferred to vials containing 2 mL of PBS, 0.5 % Triton X-100 and 6 mL of heptane. Cross-linking was initiated by adding 100 µl of 37 % formaldehyde, followed by vigorous shaking. After 10 min, samples were spun at 500 g for 1 min and the upper heptane layer was removed. 15 min after the start of fixation, 5 mL of PBS, 0.5 % Triton X-100, 125 mM glycine was added to the embryos, followed by vigorous shaking for 1 min. The embryos were rinsed 3 times with PBS, 0.5 % Triton X-100. Embryos were sorted in small batches under a light microscope, based on morphology, to select embryos of the appropriate developmental stage and remove damaged embryos or embryos with abnormal morphology. Sorted embryos were aliquoted so that a single tube contained enough embryos for one experiment, then flash-frozen in liquid nitrogen and stored at -80 °C. We used 30-60 embryos for each *in situ* Hi-C experiment.

For single-cell RNA-seq, we adapted the collection and methanol fixation procedures described in (Alles et al., 2017; Karaikos et al., 2017). Following a pre-collection period of at least one hour, fly embryos were collected on yeasted apple juice plates at 25 °C. After 1 hour of collection, the embryos on the plate were incubated at 25 °C for 2.25 hours. Embryos were dechorionated for 2 min in 2.6 % sodium hypochlorite, rinsed with water, and suspended in PBS, 0.5 % Triton X-100. Embryos were rinsed with cell culture grade DPBS without Ca²⁺ and Mg²⁺ to remove residual detergent, and placed on ice at precisely 3.5 hours after the start of collection. Embryos were resuspended in 500 µl ice cold dissociation buffer (cell culture grade DPBS without Ca²⁺ and Mg²⁺, 0.04 % BSA) and dissociated with a clean metal pestle. Cells and tissue fragments were pelleted at 500 x g for 5 min at 4 °C, then gently resuspended in 100 µl Trypsin-EDTA 0.25 % and incubated for 3 minutes. After 3 minutes Trypsin was quenched by adding 1 mL cell culture grade DPBS without Ca²⁺ and Mg²⁺, 10 % FCS. Cells were pelleted at 1000 g for 5 min at 4 °C, then resuspended in 500 µl dissociation buffer, pelleted again and resuspended in 100 µl dissociation buffer. A 10 µl aliquot of cells was kept and counted using an improved Neubauer chamber or a Luna2 cell counter. To fix cells, 4 volumes of 100 % methanol, pre-chilled at -20 °C, were slowly added to the cells. Fixed cells were stored at -80 °C and used within 3 days.

ChIP-seq

2-4 hr old *Toll^{rm9}/Toll^{rm10}* embryos for ChIP sequencing were collected and fixed as described above for Hi-C. Fixed embryos were flash-frozen in liquid nitrogen and stored at -80 °C until further use. Frozen embryos were homogenized in sonication buffer (50 mM HEPES, 140 mM NaCl, 1 mM EDTA, 1 % Triton, 0.1 % sodium deoxycholate, 0.1 % SDS and protease inhibitor) using a Dounce homogenizer. The sample were spun at 4000 g for 5 min and the pellet containing the intact nuclei was resuspended in same buffer supplemented with 0.5 % *N*-Lauroylsarcosine and SDS to final concentration of 0.5 %. The chromatin was sheared to fragment size in the range of 200–500 bp using a Bioruptor (Diagenode). The solubilized chromatin fraction was cleared by centrifugation and used for immunoprecipitation after diluting 5 times with sonication buffer. Immunoprecipitation with either 2 µg of H3K27ac (Abcam, ab4729) or 5 µg of H3K27me3 (Abcam, ab6002) antibody was carried out on chromatin corresponding to 20-25 µl of embryos at 4 °C overnight. Chromatin-antibody complexes were captured for at least 3 hr using a mix of Protein A and G Dynabeads (Invitrogen). The captured immunoprecipitated complex were washed 10 min each with sonication buffer (50 mM HEPES, 140 mM NaCl, 1 mM EDTA, 1 % Triton, 0.1 % sodium deoxycholate, 0.1 % SDS), WashA (as sonication buffer, but with 500 mM NaCl), WashB (20 mM Tris pH 8, 1 mM EDTA, 250 mM LiCl, 0.5 % NP-40, 0.5 % sodium deoxycholate) and TE. After the washes Dynabeads with bound chromatin-antibody complexes were resuspended in 100 µl TE supplemented with 20 mg/ml RNase A and incubated at 50 °C for 30 min. Cross-linking was reversed by adding Tris pH 8.0 and SDS to a final concentration of 50 mM and 0.1% respectively and heating at 68 °C for at least 4 hr. Protein digestion was carried out by Proteinase K treatment at 55 °C for 2 hr, followed by purifying ChIP DNA using ChIP DNA Clean & Concentrator kit (Zymo research #D5205). ChIP-seq libraries were prepared on the ChIP DNA eluted in 60 µl of DNA elution buffer, using the NEBNext Ultra II DNA Library Prep Kit (NEB). ChIP samples were single-end (1 × 75 bp) sequenced on Illumina NextSeq platform at BEA core facility, Stockholm.

scRNA-seq

We performed scRNA-seq using the 10X Genomics Chromium Single Cell 3' Reagents v3, according to the manufacturer's instructions (Rev B). Methanol-fixed cells were spun at 3000 g at 4 °C for 5 min and resuspended in 500 µl DPBS + 0.04 % BSA to rehydrate. Rehydrated cells were counted using a Luna2 cell counter, and the volume used for library preparation was chosen for a targeted recovery of 5000 cells. Libraries were sequenced on an Illumina NextSeq 500, using paired-end sequencing with read 1 length 28 cycles, index read length 8 cycles, and read 2 length 91 cycles.

Hi-C

We performed *in situ* Hi-C according to the protocol in (Hug and Vaquerizas, 2018b; Hug et al., 2017), using MboI as the restriction enzyme, with minor modifications for optimisation for low input according to (Díaz et al., 2018).

ChIP-seq analysis

ChIP-seq reads were mapped to the dm6 genome using Bowtie2 (version 2.3.3.1 (Langmead and Salzberg, 2012)). Mapped reads were filtered to remove alignments with quality scores less than 30, as well as secondary and supplementary alignments. PCR duplicates were marked using sambamba (version 0.6.8, (Tarasov et al., 2015)). Coverage tracks were generated using the bamCoverage tool from deepTools (version 3.2.0, (Ramírez et al., 2014)) with the following parameters: "-of bigwig --binSize 10 --normalizeUsing CPM --extendReads 200 --

ignoreDuplicates --minMappingQuality 30" and keeping only reads from chromosomes X, 2L, 2R, 3L, 4, and Y. ChIP-seq peaks were called using MACS2 (version 2.2.6, (Feng et al., 2012)) with the following parameters: "--nomodel --extsize 147 -g dm " or "--nomodel --extsize 147 -g dm --broad --min-length 500 --max-gap 200" for broad peaks. We used merged input samples for each genotype as the controls for all peak calling, due to a lack of sample-matching information for the published datasets that were re-analysed.

Bulk RNA-seq analysis

RNA-seq reads were quantified using Salmon (1.1.0, (Patro et al., 2017)) and the Flybase r6.30 transcripts. Salmon was used in mapping-based mode, with the following parameters: "-l A --validateMappings --seqBias". For visualisation purposes, RNA-seq reads were also aligned to the dm6 genome, using Hisat2 (version 2.1.0, (Kim et al., 2019)). Mapped reads were filtered to remove alignments with quality scores less than 30, as well as secondary and supplementary alignments. PCR duplicates were marked using sambamba (version 0.6.8, (Tarasov et al., 2015)). Coverage tracks were generated using the bamCoverage tool from deepTools (version 3.2.0, (Ramírez et al., 2014)) with the following parameters: "-of bigwig --binSize 10 --normalizeUsing CPM --extendReads 200 --ignoreDuplicates --minMappingQuality 30" and keeping only reads from chromosomes X, 2L, 2R, 3L, 4, and Y.

We used tximport (version 1.14.2, (Soneson et al., 2016)) to import quantifications from Salmon into R (3.6.3) and estimate transcripts per million values. We carried out pairwise differential expression analysis between *gd⁷*, *Toll^{rm9/rm10}*, and *Toll^{10B}* using DESeq2 (version 1.26.0, (Love et al., 2014)) with default parameters.

Identification of candidate tissue-specific enhancers

In order to identify tissue-specific enhancers, we first carried out pairwise differential H3K27ac signal analysis using csaw (version 1.20.0, (Lun and Smyth, 2014; Lun and Smyth, 2016)) and edgeR (version 3.28.1, (Robinson et al., 2010)). We used 2000 bp windows for the background calculations and selected 150 bp windows with a 3-fold enrichment over the background. Windows were merged using the parameters "tol = 100" and "max.width = 5000". Merged regions with an FDR < 0.05 and with a consistent direction of change across all windows were selected for downstream analysis. Candidate tissue-specific enhancers were defined by taking the intersection of regions identified as enriched for H3K27ac in each genotype compared to both others.

We validated the putative enhancers by comparing to enhancers identified in previous studies. 7 of 22 dorsal ectoderm enhancers identified from a literature search (Koenecke et al., 2016) overlap with our *gd⁷*-specific enhancers, while 13 of 37 mesoderm enhancers overlap our *Toll^{10B}*-specific enhancers (Fig. S1C). The relatively low overlap can be explained by the fact that many literature enhancers have H3K27ac signal in *Toll^{rm9/rm10}* as well as either *gd⁷* or *Toll^{10B}*. Putative enhancers were also overlapped with regions tested for enhancer activity in *Drosophila* embryos by (Kvon et al., 2014). Regions ("tiles") tested by Kvon et al. that were active in at least one tissue and timepoint were lifted over to dm6 from dm3. 114 putative enhancers overlapped a total of 124 tiles by at least 1bp. Out of 27 *gd⁷* enhancers which overlap tiles that are active in stage 4-6 or stage 7-8, 21 are active in either dorsal ectoderm or amnioserosa precursors/subsets. Out of 16 *Toll^{rm9/rm10}* enhancers that overlap tiles that are active in stage 4-6 or stage 7-8, 14 are active in brain or ventral nerve cord precursors, procephalic ectoderm, or ventral ectoderm. Out of the 30 *Toll^{10B}* enhancers that overlap tiles that are active in stage 4-6 or stage 7-8, 28 are active in mesoderm precursors/subsets.

Enhancer heatmaps were made using the plotHeatmap tool from deepTools (version 3.2.0, (Ramírez et al., 2014)). Overlaps between different enhancer sets were visualised using UpSetR (version 1.4.0, (Conway et al., 2017; Lex et al., 2014)).

Assignment of candidate enhancers to target genes

We defined “housekeeping genes” as genes that have at least ‘low’ expression in all stages and tissues according to Flybase RNA-seq data (1867 genes). We filtered the set of genes from the Flybase 6.30 transcripts to remove these housekeeping genes, as well as any genes with an average TPM < 1 in the *gd*⁷, *Toll*^{rm9/rm10}, and *Toll*^{10B} bulk RNA-seq. Candidate tissue-specific enhancers were assigned to target genes using the following rules: first, we assigned any enhancers that overlapped a single transcript to that gene. Next, we assigned enhancers to the closest promoter that was not separated from the enhancer by a domain boundary (using consensus boundaries from 3-4 hpf embryos, see below). The remaining enhancers were assigned to the closest promoter within the same domain, or, if they were not inside a domain, to the closest promoter.

scRNA-seq analysis

We used CellRanger (version 3.1.0) to produce fastq files for the scRNA-seq data and to align, filter, and quantify reads based on the BDGP6.22 genome release (Ensembl 98) to produce feature-barcode matrices. We imported the filtered matrices into R using DropletUtils (version 1.6.1, (Griffiths et al., 2018)), and performed additional quality control analysis using scater (version 1.14.6, (McCarthy et al., 2017)). Doublets were identified using scDblFinder (version 1.1.8, (Germain et al., 2020)), with an estimated doublet rate of 3.9%, and removed. Normalisation for library size across cells was performed using scater and scanr (version 1.14.6, (Lun et al., 2016)) using the “deconvolution” approach described in (Amezquita et al., 2019), in which cells are pre-clustered and size factors estimated using the calculateSumFactors() function.

Downstream analysis was carried out using Seurat (version 3.1.4, (Butler et al., 2018; Stuart et al., 2019)). The VST method was used to select the top 3000 variable features for each sample, then all datasets were integrated using the control dataset with the highest number of cells (replicate 1) as the reference dataset, and the first 30 dimensions. We performed clustering using the Shared Nearest Neighbour approach implemented in the Seurat functions FindNeighbors and FindClusters, using the first 12 dimensions from PCA, a k.param value of 60, and a clustering resolution of 0.5. These parameters were chosen because they produced clusters that were stable to small variations in the parameter values.

We carried out differential expression analysis using the Seurat function FindMarkers in order to identify genes with higher expression in a cluster compared to all other cells, and in pairwise comparisons. We carried out Gene Ontology enrichment analysis on the results marker gene sets using the enrichGO function from clusterProfiler (version 3.14.3, (Yu et al., 2012)), and simplified the results to remove semantically similar terms using the simplify function from clusterProfiler with the Wang method and a similarity threshold of 0.7. These marker gene sets and enriched GO terms, along with expression of known markers for embryonic cell populations, were used to assign putative cluster identities.

In order to quantify the average expression of particular gene sets in Figure 2C and Figure S3C, we calculated the sum of expression of those genes per cell, and then expressed this as a Z-score across all cells.

Pooled scRNA-seq reads from all barcodes were analysed using Salmon as described above.

Hi-C analysis

Hi-C data was analysed using FAN-C (version 0.8.28) (Kruse et al., 2020). Paired-end reads were scanned to identify ligation junctions, split at ligation junctions if any were present, and then aligned independently to the dm6 genome using BWA-MEM (version 0.7.17-r1188) (Li and Durbin, 2009). Aligned reads were filtered to retain only uniquely-aligned reads with a mapping quality of at least 3. Reads were then paired based on read names and assigned to restriction fragments. “Inward” and “outward” reads separated by less than 1 kb, representing likely unligated fragments and self-ligated fragments respectively, were removed. In addition, we removed PCR duplicates, reads mapping more than 500 bp from a restriction site, and self-ligations where both reads map to the same fragment.

We generated two biological replicate datasets for each genotype, which showed high similarity. Therefore, we pooled biological replicates to reach 2 kb resolution. Matrices created from merged biological replicates were binned and filtered using FAN-C default parameters to remove bins with coverage less than 10 % of the median coverage. Normalisation was performed using Knight-Ruiz matrix balancing (Knight and Ruiz, 2013). Expected contacts are calculated as the average contacts at each genomic distance separation. Hi-C data was visualised using plotting tools from FAN-C and using HiGlass (Kerpedjiev et al., 2018).

Domain and boundary identification

The insulation score was calculated as described in (Crane et al., 2015), using FAN-C, for 2 kb and 5 kb resolution matrices, each with window sizes of 4, 6, 8, and 10 bins. Domain boundaries were calculated from the insulation score with a delta parameter of 3 and filtered to keep only boundaries with a boundary score of at least 0.7. Consensus boundaries for each sample were created by overlapping boundaries called at the two different resolutions and four different window sizes, and keeping those boundaries which were identified using at least four out of the total eight parameter combinations. Domains were created by pairing boundaries, and domains less than 10 kb or more than 500 kb in size were removed.

Hi-C aggregate analysis

Aggregate compartment, domain, and loop plots were created using FAN-C. Compartments were identified using the first eigenvector of the correlation matrix of the normalised Hi-C data, using GC content to orient the eigenvector. The compartment eigenvector for the 3-4 hpf Hi-C data from Hug et al. 2017 was used as the reference for the aggregate compartment plots (“saddle plots”) (Flyamer et al., 2017; Gassler et al., 2017; Imakaev et al., 2012; Kruse et al., 2020). Domain aggregates were also created using the domains identified in the Hug et al. 3-4 hpf Hi-C data. Loop aggregates were created using the loops identified in Kc167 cells by (Cubenas-Potts et al., 2017). Similar results were obtained using loops from (Eagen et al., 2017; Stadler et al., 2017) (not shown).

We constructed a BEDPE file of putative enhancer-promoter interactions, considering all unique transcript start sites for assigned target genes. Interactions with a separation of at least 10 kb were used to create enhancer-promoter aggregate plots.

Hi-C similarity score analysis with CHESS

We used CHESS (Comparison of Hi-C Experiments using Structural Similarity) (Galan, Machnik et al., in revision) to compare Hi-C data from embryos of different genotypes. Briefly, CHESS treats Hi-C interaction matrices as images and applies the concept of the structural

similarity index (SSIM), which is widely used in image analysis. We applied CHESS to 5 kb resolution Hi-C matrices using windows of 500 kb and a step size of 5 kb to produce similarity scores for pairwise Hi-C comparisons. Hi-C data from stage 5 control embryos was used as the reference dataset, and compared to data from stage 5 dorsoventral mutant embryos and nc14 control embryos from Hug et al. 2017. In order to identify regions of the genome with significant changes between the reference and query datasets, we selected regions with a SSIM Z-score less than -2, and a signal-to-noise ratio Z-score of at least 1.

Statistics and visualisation

Statistical tests were carried out in R (version 3.6.3) and visualisation was performed using the ggplot2 package (Wickham, 2016). Box plots are defined with boxes spanning the first to third quartiles. The whiskers extend from the box to the smallest/largest values no further than 1.5 times the interquartile range (IQR) away from the box. The notches extend $1.58 * \text{IQR} / \sqrt{n}$ from the median.

Data availability

The Hi-C, scRNA-seq, and ChIP-seq data produced in this study have been submitted to ArrayExpress.

SUPPLEMENTARY FIGURES

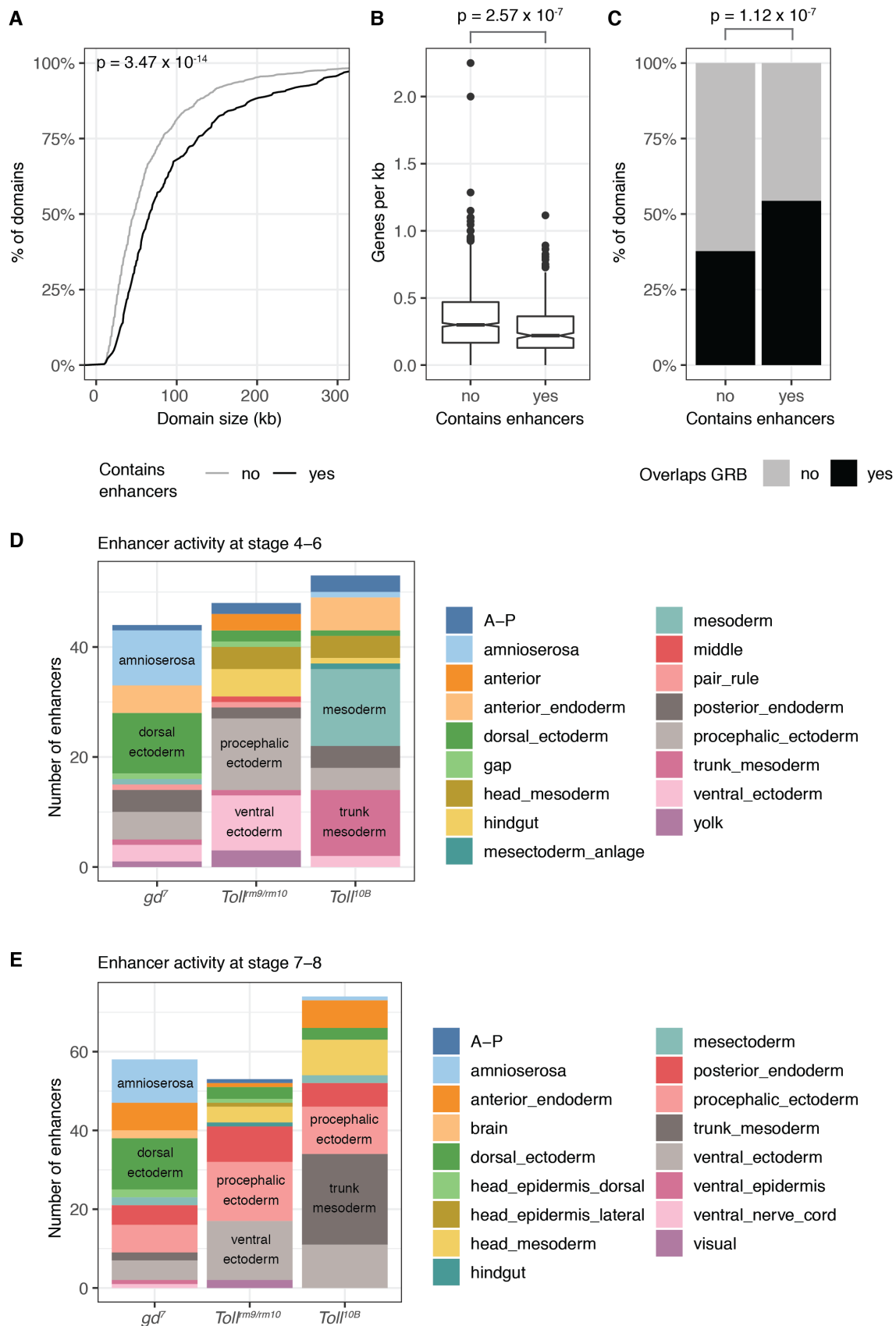


Figure S1. Properties of domains containing tissue-specific putative enhancers and validation of tissue-specific enhancer activity.

- A.** Size of domains containing putative tissue-specific enhancers (black) and those without tissue-specific enhancers (grey). Wilcoxon rank sum test $p = 3.47 \times 10^{-14}$.
- B.** Genes per kilobase inside domains containing putative tissue-specific enhancers and those without tissue-specific enhancers. Wilcoxon rank sum test $p = 2.57 \times 10^{-7}$.
- C.** Overlap of chromatin domains with Genomic Regulatory Blocks (GRBs) from (Harmston et al., 2017a). Chi-squared test $p = 1.12 \times 10^{-7}$.
- D, E.** Putative enhancers active in *gd⁷*, *Toll^{rm9/rm10}*, and *Toll^{l0B}* embryos overlap with enhancers tested by (Kvon et al., 2014) and are enriched for enhancers that drive expression in relevant regions of the embryo at stages 4-6 (**D**) and 7-8 (**E**). The two largest categories are labelled for each stacked bar.

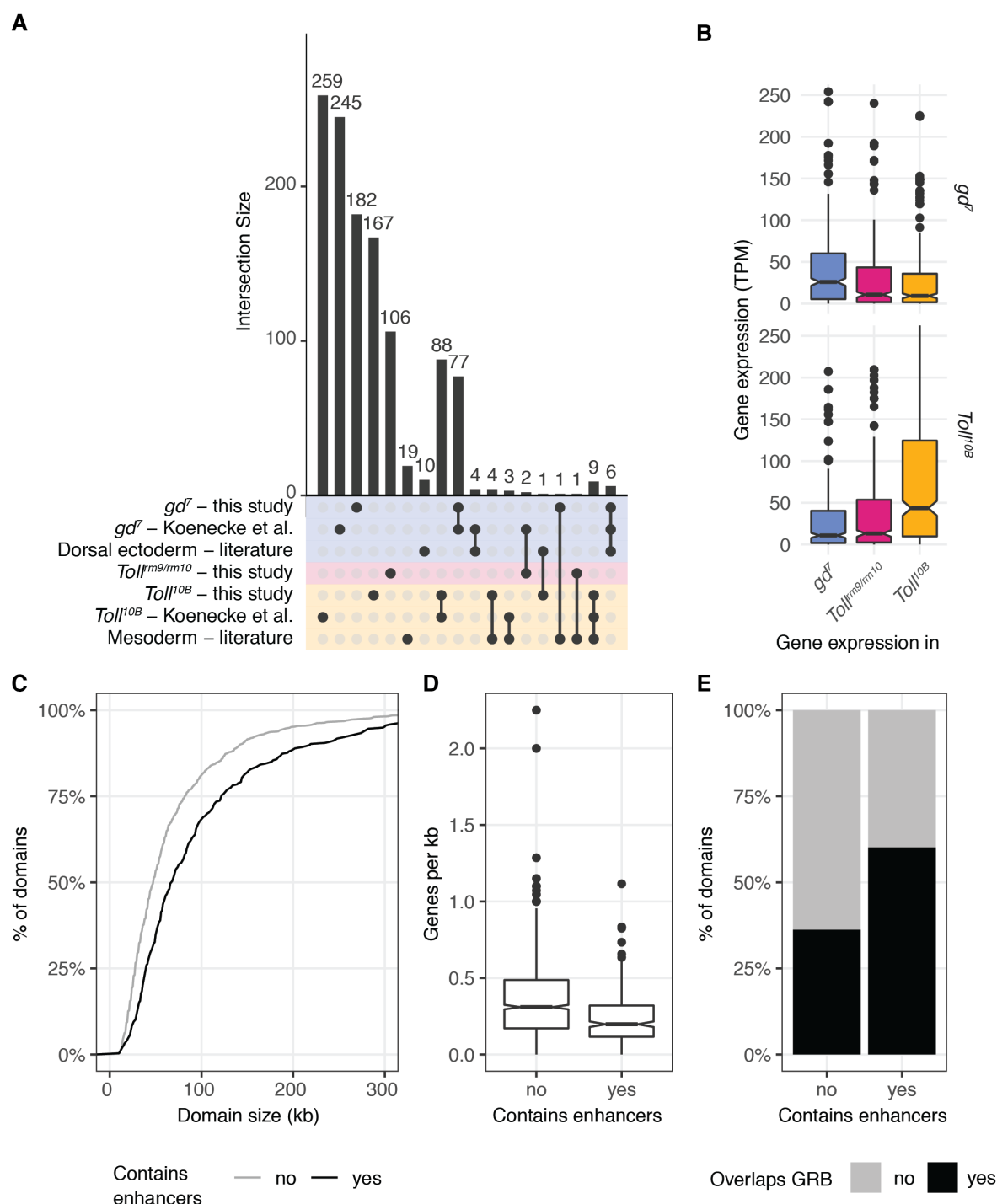


Figure S2. Properties of domains containing tissue-specific putative enhancers are robust to different enhancer definitions.

A. UpSet plot showing the overlap between the putative tissue specific enhancers identified in this study, those identified in (Koenecke et al., 2016), and mesoderm and dorsal ectoderm enhancers identified from a literature search by (Koenecke et al., 2016). The lower panel represents the groups of genomic regions considered (y axis), and intersections between these groups are shown by joined dots. The bars represent the sizes of these intersections. For example, there are 6 putative dorsal ectoderm enhancers that are identified in this study, identified by Koenecke et al., and previously identified from a literature search (rightmost column).

B. Expression of genes associated with putative enhancers identified by Koenecke et al. in *gd⁷* and *Toll^{10B}* mutant embryos. Top, genes associated with *gd⁷*-specific putative enhancers; bottom, genes associated with *Toll^{10B}*-specific putative enhancers. Asterisks indicate Wilcoxon rank sum test $p < 0.001$.

C. Size of domains containing putative tissue-specific enhancers identified by (Koenecke et al., 2016) (black) and those without tissue-specific enhancers (grey). Wilcoxon rank sum test $p = 7.12 \times 10^{-13}$.

D. Genes per kilobase inside domains containing putative tissue-specific enhancers identified by (Koenecke et al., 2016) and those without tissue-specific enhancers. Wilcoxon rank sum test $p = 7.67 \times 10^{-16}$.

E. Overlap of chromatin domains with Genomic Regulatory Blocks (GRBs) from (Harmston et al., 2017a). Chi-squared test $p = 3.14 \times 10^{-14}$.

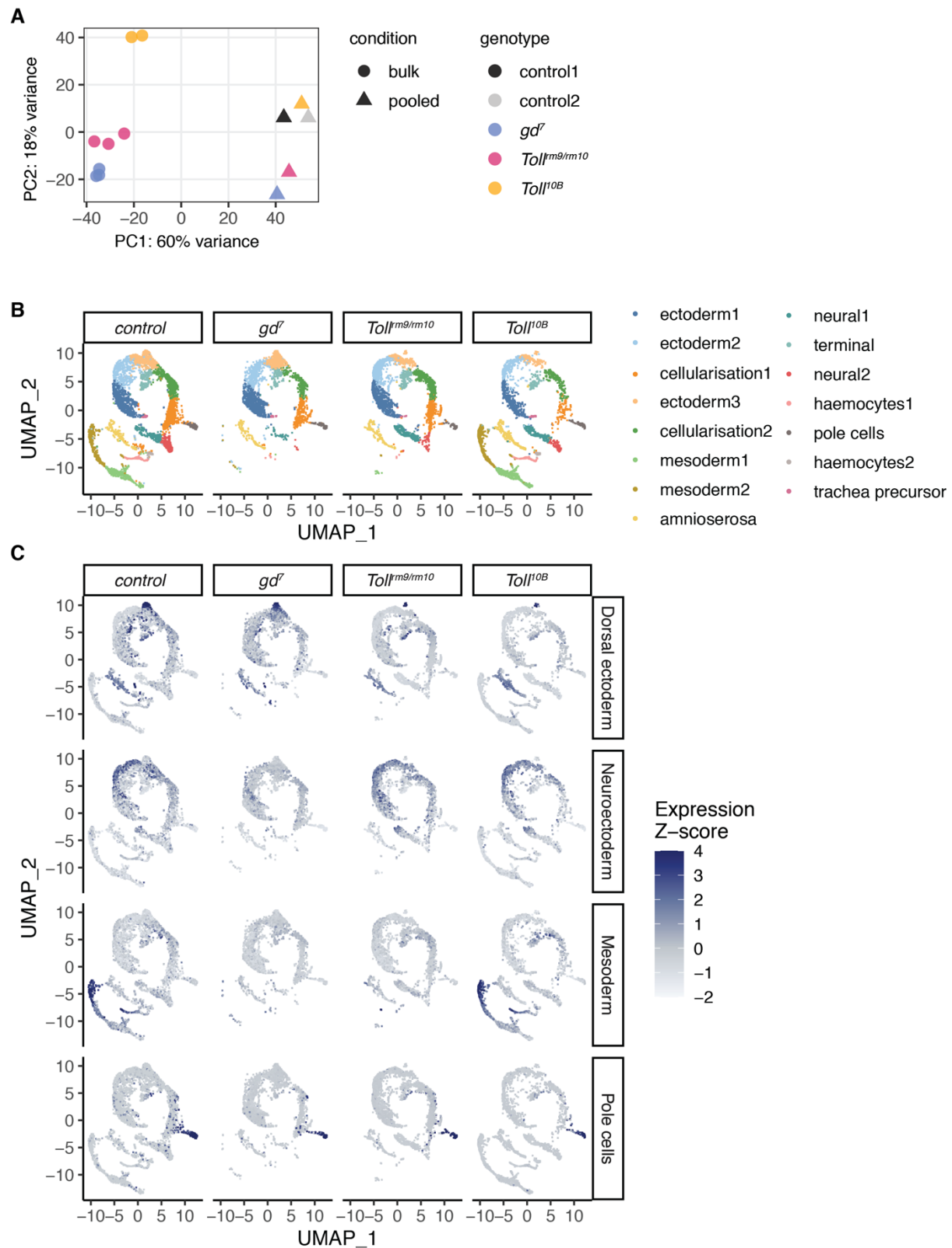


Figure S3. Validation of scRNA-seq data.

A. PCA of pooled single-cell RNA-seq data and bulk RNA-seq data. The first principal component separates the methods while the second principal component separates the genotypes. Replicate control single-cell RNA-seq experiments cluster together, demonstrating robustness.

B. Clustering of single-cell gene expression profiles from 2.5-3.5 hpf embryos reveals clusters corresponding to distinct cell populations, as in Fig 2A but separated by genotype of origin.

C. Expression of tissue-specific marker genes across single cells from different genotype origins. Marker genes for dorsal ectoderm (*Ance*, *CG2162*, *Doc1*, *Doc2*, *egr*, *peb*, *tok*, *ush*, *zen*), neuroectoderm (*ac*, *brk*, *CG8312*, *l(1)sc*, *mfas*, *Ptp4E*, *sog*, *SoxN*, *vnd*), mesoderm (*CG9005*, *Cyp310a1*, *GEFmeso*, *lth*, *Mdr49*, *Mes2*, *NetA*, *ry*, *sna*, *stumps*, *twi*, *wgn*, *zfh1*), and pole cells (*pgc*) were obtained from (Karaiskos et al., 2017).

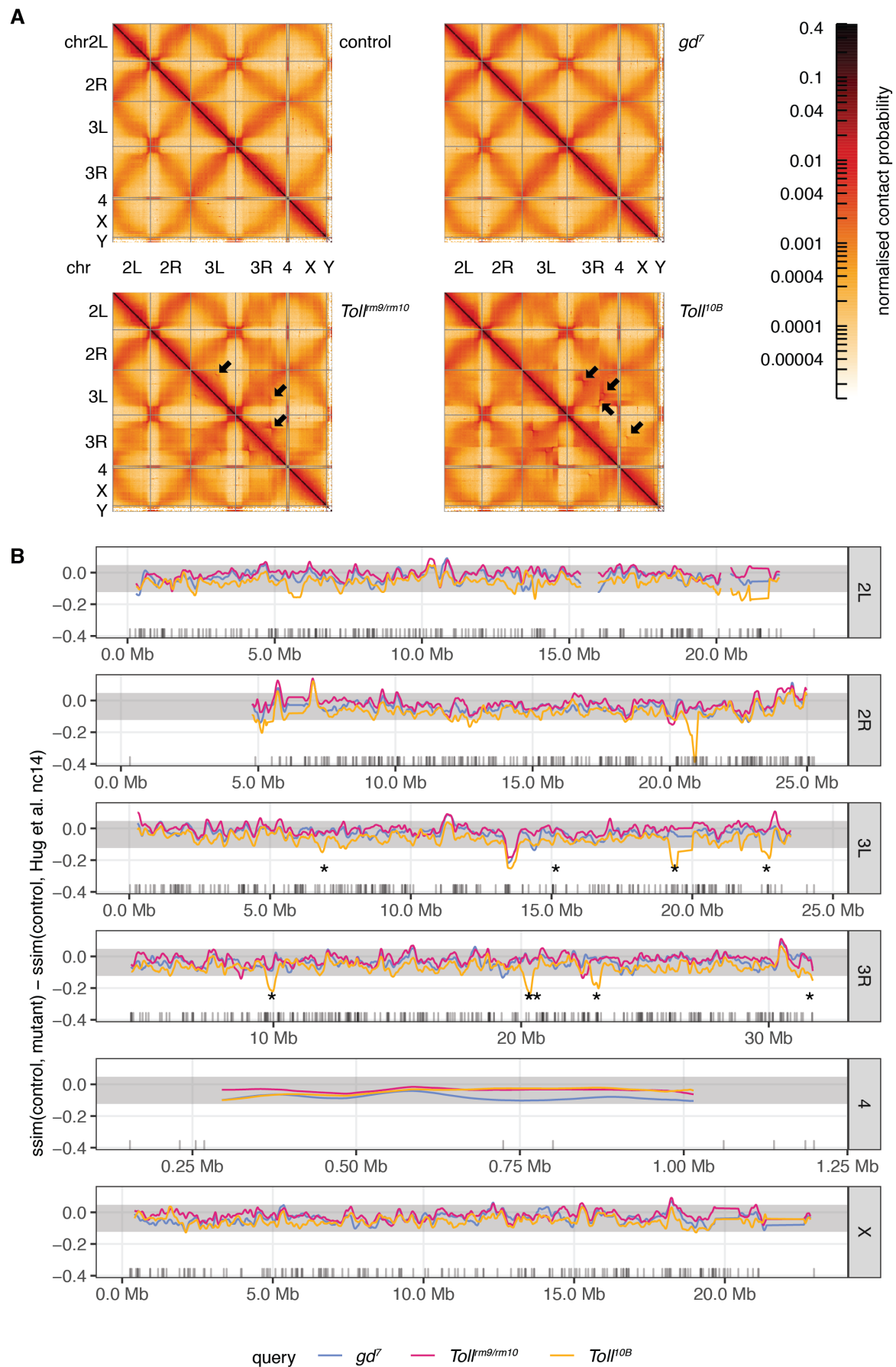


Figure S4. Balancer chromosomes contribute to differences between Hi-C matrices in embryos from different genotypes.

A. Whole genome contact probability maps for control, *gd⁷*, *Toll^{rm9rm/10}*, and *Toll^{10B}* embryos at the cellular blastoderm stage. The boundaries of assembled chromosomes and chromosome arms are marked by grey lines. Black arrows mark artefacts in the Hi-C data that indicate rearrangements on balancer chromosomes present in a subset of *Toll^{rm9rm/10}* (TM6) and *Toll^{10B}* (TM3 and OR60) embryos.

B. CHESS (Galan, Machnik et al., in revision; see companion manuscript) similarity scores were calculated between mutant and control embryo Hi-C datasets, using 5kb resolution and a 500 kb window size. As a reference, similarity scores were calculated between control embryo Hi-C data and the nc14 Hi-C data from Hug et al. 2017. The difference between this reference similarity score and the similarity scores between mutant and control embryos is shown for all chromosomes (blue, *gd⁷*; pink, *Toll^{rm9rm/10}*; yellow, *Toll^{10B}*). Similarity score differences around zero represent regions where chromatin conformation is similar between control and mutant, while negative values represent regions where there are greater differences between control and mutant than between control and the Hug et al. nc14 data. Shaded area represents values within two standard deviations of the genome-wide mean. Grey ticks represent the positions of genes that are differentially expressed between dorsoventral mutant embryos (Koennecke et al., 2016). Asterisks mark positions of known rearrangement breakpoints on the TM3 balancer chromosome (Ghavi-Helm et al., 2019).

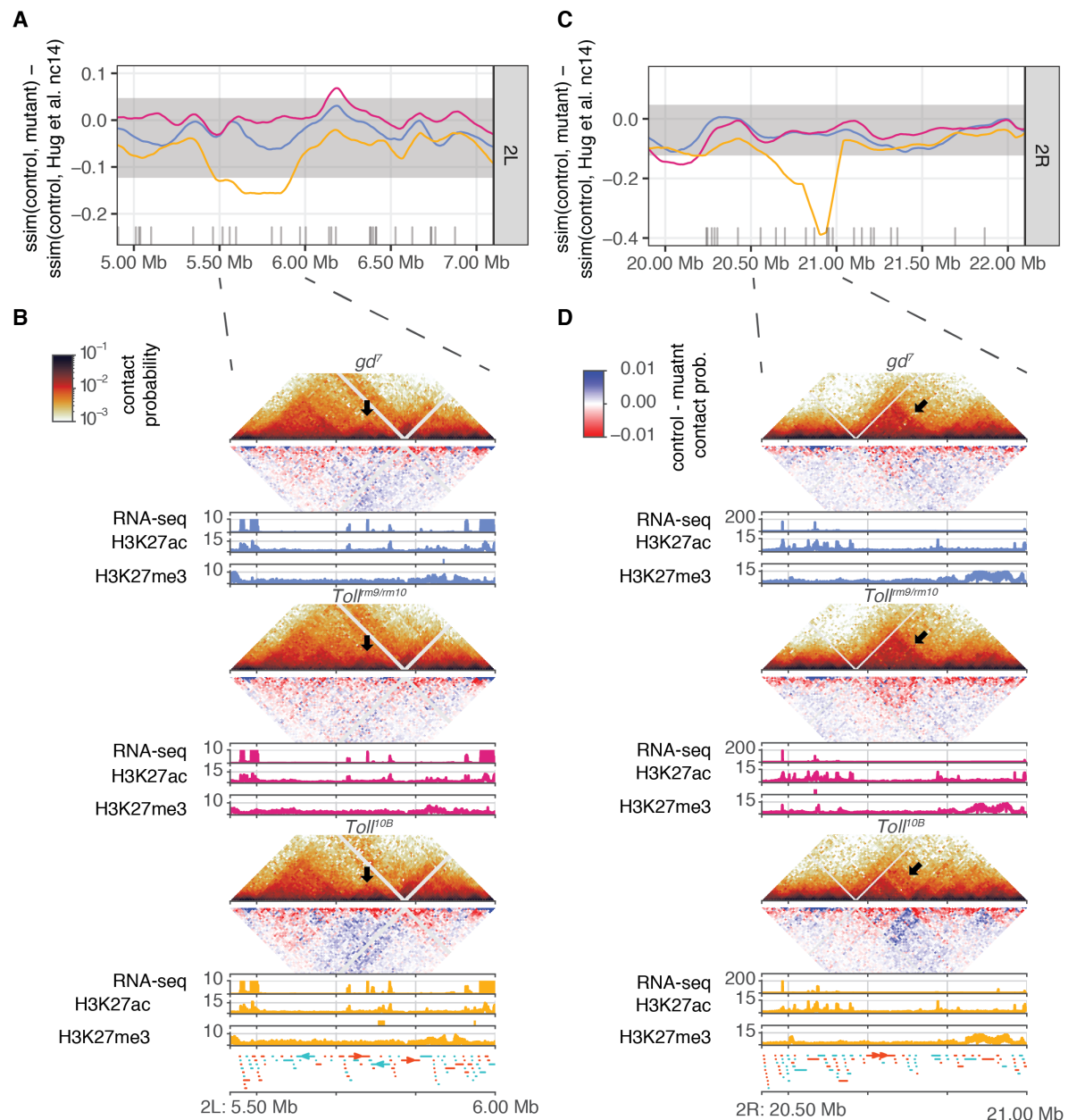


Figure S5. A small number of regions have changes in chromatin conformation detected by CHESS that are not associated with known genomic rearrangements.

A, C. CHESS (Galan, Machnik et al., in revision; see companion manuscript) similarity scores were calculated between mutant and control embryo Hi-C datasets, using 5kb resolution and a 500 kb window size. As a reference, similarity scores were calculated between control embryo Hi-C data and the nc14 Hi-C data from Hug et al. 2017. The difference between this reference similarity score and the similarity scores between mutant and control embryos is shown for all chromosomes (blue, *gd7*; pink, *Toll^{rm9/rm10}*; yellow, *Toll^{10B}*). Similarity score differences around zero represent regions where chromatin conformation is similar between control and mutant, while negative values represent regions where there are greater differences between control and mutant than between control and the Hug et al. data. Shaded area represents values within two standard deviations of the genome-wide mean. Grey ticks represent the positions of genes that are differentially expressed between dorsoventral mutant embryos (Koenecke et al., 2016). **B, D.** For each genotype, top, normalised Hi-C contact probability maps and Hi-C difference maps at 5kb resolution. Hi-C difference maps are calculated as *contact probability in control*

– *contact probability in mutant*; red indicates regions with increased contact probability in embryos of the mutant genotype and blue indicates decreased contact probability. The black arrows indicate a region with a change in contact probability in *Toll^{l0B}*. Bottom, RNA-seq (Koenecke et al., 2016), H3K27ac and H3K27me3 ChIP-seq data ((Koenecke et al., 2016), this study). Tissue-specific putative enhancers identified as described above are shown as colour-coded bars beneath the corresponding H3K27ac ChIP-seq track. Lower panel, gene annotations. Positive-strand genes are shown in orange and negative strand genes are shown in blue.

- A.** Example of differences in CHES similarity scores for a 2 Mb region on chromosome 2L.
- B.** Hi-C data for a 500 kb subset of the region shown in **A**.
- C.** Example of differences in CHES similarity scores for a 2 Mb region on chromosome 2R.
- D.** Hi-C data for a 500 kb subset of the region shown in **C**.

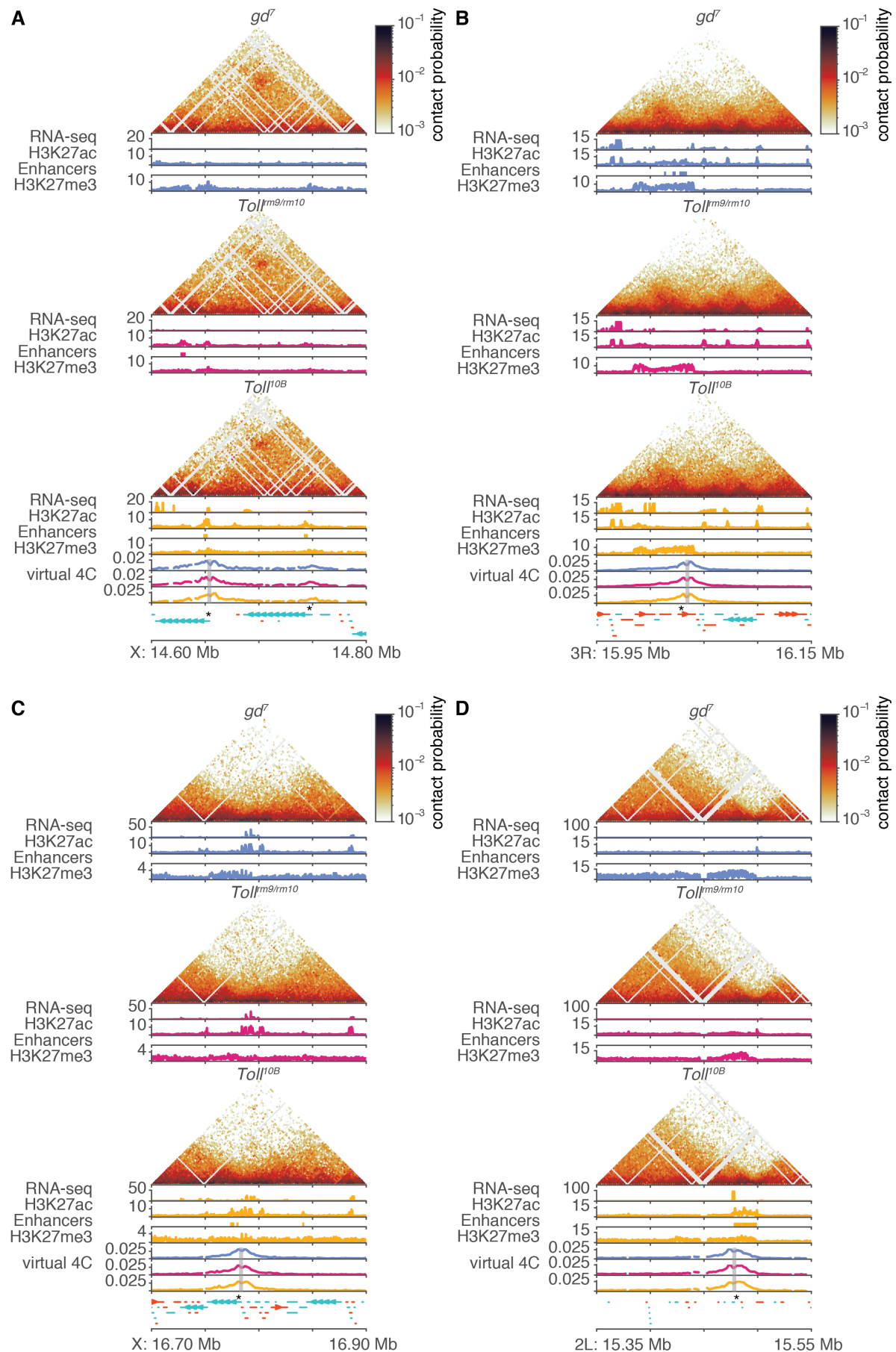


Figure S6. Chromatin conformation is not affected by tissue-specific gene expression.

Tissue-specific chromatin data for regions containing dorsoventral patterning genes. For each genotype, top, normalised Hi-C contact probability maps, at 2kb resolution; middle, RNA-seq data (Koencke et al., 2016); bottom, H3K27ac and H3K27me3 ChIP-seq data ((Koencke et al., 2016), this study). Tissue-specific putative enhancers identified as described above are shown as colour-coded bars beneath the corresponding H3K27ac ChIP-seq track. Lower panels, “Virtual 4C” tracks for each genotype representing interactions of a 2kb region around the promoters of genes of interest, as highlighted by the grey rectangle; gene annotations. Positive-strand genes are shown in orange and negative strand genes are shown in blue. Dorsoventral patterning genes of interest are marked with asterisks.

A. *NetA* and *NetB*. **B.** *pnr*. **C.** *if*. **D.** *sna*.

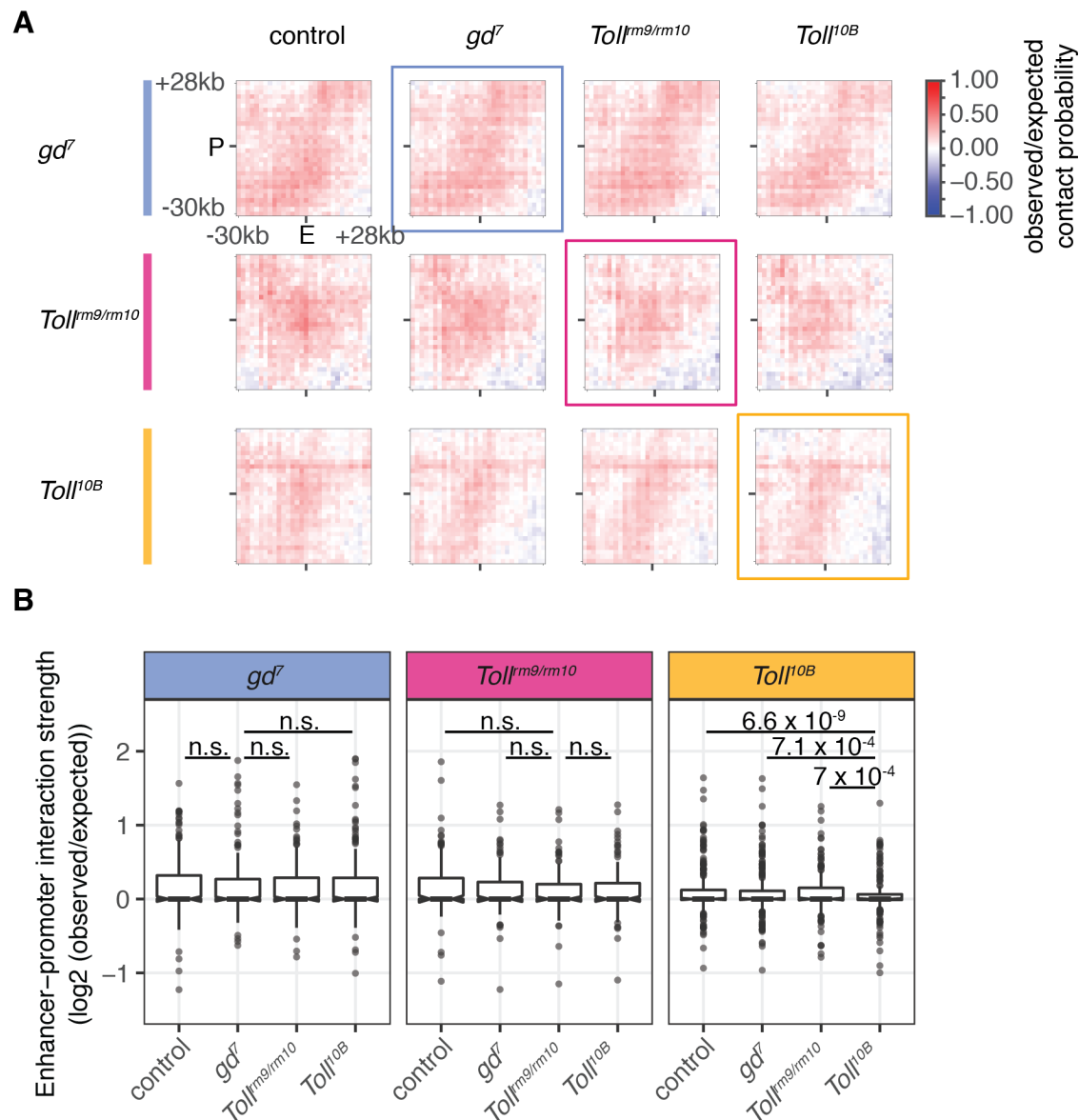


Figure S7. Interaction strength between putative tissue-specific enhancers and the promoters of their assigned genes does not change between tissues.

A. Aggregate contact analysis for putative tissue-specific enhancers (E) and the promoters (P) of their assigned genes. The average observed/expected contact probability is shown for Hi-C data at 2 kb resolution in a window of 60 kb around putative enhancer-promoter interactions. Rows represent different sets of putative tissue-specific enhancers, while columns represent Hi-C data from cellular blastoderm embryos of different genotypes. Coloured squares highlight datasets in which the enhancers in that row are active (blue, *gd⁷*; pink, *Toll^{rm9/rm10}*; yellow, *Toll^{10B}*).

B. Quantification of contact probability between putative enhancers and their assigned target promoters, in Hi-C datasets from embryos of different genotypes. Panels represent different sets of putative tissue-specific enhancers, while the x-axis represents Hi-C data from cellular blastoderm embryos of different genotypes. There are no significant differences in interaction strength across Hi-C datasets for *gd⁷* or *Toll^{rm9/rm10}* enhancers. *Toll^{10B}* enhancers have significantly lower contacts with their target promoters in Hi-C from *Toll^{10B}* mutant embryos.