



Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early *Drosophila* development

Sergio Martin Espinola^{ID 1,5}, Markus Götz^{ID 1,5}, Maelle Bellec^{ID 2}, Olivier Messina^{1,2}, Jean-Bernard Fiche¹, Christophe Houbron¹, Matthieu Dejean², Ingolf Reim^{ID 3}, Andrés M. Cardozo Gizzi^{ID 4}, Mounia Lagha^{ID 2}✉ and Marcelo Nollmann^{ID 1}✉

Acquisition of cell fate is thought to rely on the specific interaction of remote cis-regulatory modules (CRMs), for example, enhancers and target promoters. However, the precise interplay between chromatin structure and gene expression is still unclear, particularly within multicellular developing organisms. In the present study, we employ Hi-M, a single-cell spatial genomics approach, to detect CRM-promoter looping interactions within topologically associating domains (TADs) during early *Drosophila* development. By comparing cis-regulatory loops in alternate cell types, we show that physical proximity does not necessarily instruct transcriptional states. Moreover, multi-way analyses reveal that multiple CRMs spatially coalesce to form hubs. Loops and CRM hubs are established early during development, before the emergence of TADs. Moreover, CRM hubs are formed, in part, via the action of the pioneer transcription factor Zelda and precede transcriptional activation. Our approach provides insight into the role of CRM-promoter interactions in defining transcriptional states, as well as distinct cell types.

Chromosomes are organized at different levels—nucleosomes, chromatin loops, TADs and chromosome territories—and each of these layers contributes to the regulation of transcription^{1,2}. In particular, loops between enhancers and promoters are critical for the precise regulation of transcriptional activation^{3–7}. In addition, organization of chromosomes into TADs plays a role in transcriptional regulation⁸, primarily by facilitating communication between enhancers and promoters through enhancer–promoter (E–P) loops within a TAD and restricting contacts from enhancers of neighboring TADs^{5,9–13}. However, the interplay of formation of E–P loops, emergence of TADs and transcriptional output is still poorly understood¹⁴.

Tissue-specific enhancers have been shown to be in proximity⁸ to their cognate promoters, indicating that E–P contacts are needed for precise gene regulation^{15–18}. Indeed, introduction of ectopically enforced E–P contacts can lead to transcriptional activation of a reporter gene during *Drosophila* development¹⁹. In some cases, enhancers can increase transcriptional output by modulating transcriptional bursting^{6,20–23}. However, in other cases, E–P contacts seem to be dissociated from gene activation^{24,25}, suggesting that an enhancer may not necessarily need to be in continuous physical contact with a promoter to influence transcription. The mechanisms by which E–P contacts may regulate transcription are currently under intense debate^{14,26–28}.

Promoters can contact several distant enhancers^{15–17}, raising the possibility that more than one enhancer may contact a promoter at any given time. More recently, use of multi-way 3C and 4C methods showed that, indeed, enhancers can cluster together to form enhancer hubs^{29–32}. This is supported by evidence of nuclear microenvironments containing multiple enhancers and clusters of

transcription factors (TFs)^{33–39}. This model is consistent with multi-way interactions between distal enhancers to regulate promoter activity of single or multiple genes by sharing resources. Whether and how formation of multi-way interactions may be related to the emergence of TADs during development^{40,41} are still open questions.

To shed light on to these questions, we investigated the interplay of transcriptional state and physical proximity between promoters and large sets of CRMs (for example, enhancers and silencers) during the awakening of the zygotic genome in early *Drosophila* embryos. During the first hours of development, *Drosophila* embryos offer an ideal biological context to decipher how CRMs are employed to establish precise spatiotemporal patterns of gene expression. Decades of genetic and genomic studies have characterized CRMs on a large scale and their usage to interpret morphogen gradients^{42–44}. In particular, the pioneering activity of factors such as Zelda (Zld) establishes early accessibility of CRMs (reviewed in ref. 45) and is involved in the emergence of TAD organization^{40,46}.

In the present study, we used Hi-M, an imaging-based technology enabling the detection of chromatin organization and transcriptional status in intact embryos^{47,48}. This technology allowed us to visualize where and when interactions between CRMs occur and investigate their impact on transcriptional states. We first used Hi-M to detect intra-TAD chromatin loops in *Drosophila* embryos. We show that most of these loops involve CRMs. In fact, we identified not only E–P loops but multiple CRM contacts (E–P, P–P and E–E) co-interacting locally in single nuclei and referred to as CRM hubs. Unexpectedly, these contacts were not found to be specific to transcriptionally active nuclei. Hence, tissues with different cell fates exhibit similar CRMs contacts and E–P loops. Moreover, networks

¹Centre de Biologie Structurale, CNRS UMR 5048, INSERM U1054, Univ Montpellier, Montpellier, France. ²IGMM, CNRS, Univ Montpellier, Montpellier, France. ³Department of Biology, Friedrich-Alexander University of Erlangen-Nürnberg, Erlangen, Germany. ⁴Centro de Investigación en Medicina Traslacional Severo Amuchastegui, Instituto Universitario de Ciencias Biomédicas de Córdoba, Consejo Nacional de Investigaciones Científicas y Técnicas, Córdoba, Argentina. ⁵These authors contributed equally: Sergio Martin Espinola, Markus Götz. ✉e-mail: mounia.lagha@igmm.cnrs.fr; marcelo.nollmann@cbs.cnrs.fr

of CRM loops are established at early stages, before both the emergence of TADs and transcriptional activation. Finally, the pioneer factor Zld is required for the establishment of subsets of CRM hubs.

Results

High-resolution Hi-M reveals preferential interactions between *cis*-regulatory modules. Functional characterization of specific chromatin loops between CRMs within TADs (Fig. 1a) requires the development of technologies adapted for the simultaneous detection of such looping interactions and of transcriptional output. Recently, we and others established a new family of imaging-based methods able to retrieve chromatin architecture and transcriptional status simultaneously in single cells (Hi-M and ORCA)^{47–49}. Hi-M relies on the labeling and imaging of the expression pattern of genes by direct detection of transcripts via RNA–FISH, followed by the sequential imaging of tens of distinct DNA loci by Oligopaint–FISH⁵⁰ in intact *Drosophila* embryos^{47,48}. First, we tested whether conventional Hi-M could detect intra-TAD chromatin loops in two genomic regions harboring early developmental genes expressed at different timings and regions of the embryo (*dorsocross* (*doc*)- and *snail* (*sna*)-TADs).

The doc-TAD contains a family of three genes, the *dorsocross* genes *doc1*, *doc2* and *doc3* encoding functionally redundant T-box TFs essential for the development of the amnioserosa and cardio-genesis⁵¹. These genes display similar expression patterns, particularly during the early stages of embryogenesis, in the blastoderm embryo (nuclear cycle (nc) 11–14), which is the focus of the present study (Extended Data Fig. 1a,b). In early embryos, the doc-TAD is flanked by insulator-binding sites (for example, CP190) and displays extensive H3K27me3 marks as well as several prominent Zld peaks (Fig. 1b)^{52–55}. At nc14, the Hi-M contact probability map of this genomic region displays two clear TADs, similar to those detected by Hi-C (TAD1 and doc-TAD; Fig. 1b and Extended Data Fig. 1a)⁴⁶. Inspection of assay for transposase-accessible chromatin using sequencing (ATAC-seq)⁵⁶, H3K4me3, H3K4me1 and H3K27ac profiles⁵⁷, as well as of enhancer databases⁵⁸, revealed that the doc-TAD contains several putative CRMs, including four potential enhancers (CRM_a, CRM_b, CRM_c and CRM_d) for the three *doc* promoters (Fig. 1c and Supplementary Table 1)^{52,53,58,59}. We note that only CRM_a displayed exhaustive binding by several chromatin insulators (Extended Data Fig. 1b). Conventional Hi-M/Hi-C did not exhibit clear specific looping interactions within the doc-TAD, most probably due to insufficient genomic resolution and coverage (Fig. 1b).

To overcome these limitations and probe communications between CRMs and promoters within TADs in an unbiased manner, we improved the genomic resolution and coverage of Hi-M by threefold (from ~8–10 kb to ~3 kb) and painted the entire doc-TAD with contiguous barcodes (Extended Data Fig. 1a,c–i), particularly targeting promoters and predicted CRMs (Fig. 1c and Extended Data Fig. 1a). We first focused on enhancers already validated by transgenic assays (CRM_{b–d}; Supplementary Table 1). The three *doc* genes within the doc-TAD exhibit a shared spatiotemporal profile of expression in late nc14 (Extended Data Fig. 2a). The frequency of activation, estimated by the number of alleles transcribing per nucleus^{20,60}, was elevated for both *doc1* and *doc2* (~90%; Extended Data Fig. 2b). In addition, *doc* genes displayed a high degree of co-activation (~70%; Extended Data Fig. 2c). Thus, we hypothesized that multiple putative CRMs are likely to contact *doc* promoters to regulate their common expression pattern.

To test this hypothesis, we selected nuclei displaying at least one nascent messenger RNA–FISH *doc1* spot and built a high-resolution Hi-M contact map containing only these nuclei (Fig. 1d–f). Remarkably, the improvement in genomic coverage in Hi-M now enabled the detection of specific looping interactions between genetic elements within the doc-TAD in intact embryos (Fig. 1f and Extended Data Fig. 1g). The strongest contacts represented, in

all cases, interactions between CRMs (Fig. 1f, yellow arrows), but there was a considerable internuclear variation (see single-nucleus snapshots in Fig. 1f). Contact frequencies did not vary considerably when only nuclei displaying the strongest *doc1* RNA–FISH signals were used to construct the matrix (Extended Data Fig. 2d), suggesting that stronger transcriptional activity did not involve different interactions. To quantify the strength of looping interactions, we calculated the intensity of the Hi-M map across an anchor to generate virtual interaction profiles (hereafter referred to as 4M plots). We observed that CRM_c predominantly interacts with CRM_a and CRM_b with similar probabilities (Extended Data Fig. 2e,vii). In contrast, we did not observe specific loops between CRMs and barcodes not containing early CRMs (for example, ctrl barcode; Fig. 1c and Extended Data Fig. 2e,iv). Interactions between CRMs and promoter regions (for example, P1, P2 and P3) were present but displayed lower frequencies than interactions between CRMs (red arrow, Fig. 1f). Interactions between CRMs and the *doc1* promoter did not depend on the activation level of *doc1* (Extended Data Fig. 2f).

Next, we investigated whether all putative CRMs displayed chromatin loops. It is interesting that a CRM predicted from epigenetic profiling but not present in enhancer databases (for example, CRM_a; see complete list of reported enhancers in Supplementary Table 1) displayed extensive interactions with reported enhancers (for example, CRM_b, CRM_c and CRM_d), as well as with the promoters of *doc* genes (Extended Data Fig. 2e,v). In contrast, a subset of barcodes harboring previously described enhancers or displaying enhancer marks (see barcodes 2, 12, 13 and 15 in Fig. 1c) failed to exhibit looping interactions with other CRMs (for example, black circle in Fig. 1f). We observed similar results at the *sna*-TAD (Supplementary Note 1 and Extended Data Fig. 2g,h). Thus, high-resolution Hi-M reveals unforeseen interactions between CRMs and other regulatory regions within the doc-TAD, and permits the quantification of the frequencies with which putative enhancers actively contact cognate target promoters in a specific tissue and developmental timing. Collectively, these data suggest that promoters interact with a panoply of enhancers that can be shared between different genes within a TAD.

Shared enhancers, promoter competition and CRM hubs. The existence of multiple pairwise interactions between CRMs within the doc-TAD and the naturally occurring overlapping expression patterns of *doc* genes (Extended Data Fig. 2a) suggest that multiple CRMs may compete or cooperate for gene activation in single cells⁶¹. To discriminate between these two hypotheses, we tested whether multi-way interactions are formed by excluding an anchor of interest and plotting the frequencies with which two barcodes interact with this given anchor³⁰. First, we selected promoters as anchors. We observed that three-way interactions with multiple promoters were infrequent (green arrows, Fig. 1g,i,ii,iii), consistent with our previous observations from Hi-M contact maps (Fig. 1f and Extended Data Fig. 2e). Instead, the three *doc* promoters preferentially looped to multiple CRMs in single nuclei (yellow arrows, Fig. 1g,i,ii,iii). A control locus with no promoter marks failed to display specific looping interactions (Fig. 1g,iv).

Genomic methods revealed the spatial clustering of multiple enhancers in cultured mammalian cells^{29,30}. To test whether spatial clustering of multiple enhancers could be directly visualized in intact embryos, we mapped three-way interactions using CRMs as anchors. It is interesting that we observed CRM_{a–d} displaying high frequencies of multi-way interactions (Fig. 1h, see examples labeled by yellow arrows). By analogy with previous studies^{29,30}, we termed these CRM interaction networks ‘CRM hubs’. CRM hubs can contain promoters (green arrows, Fig. 1h), but most often contained known or putative enhancers. Analysis of the *sna*-TAD reveals a similar scenario, where CRMs are involved in most three-way interactions (Extended Data Fig. 2i,iv,v,vi).

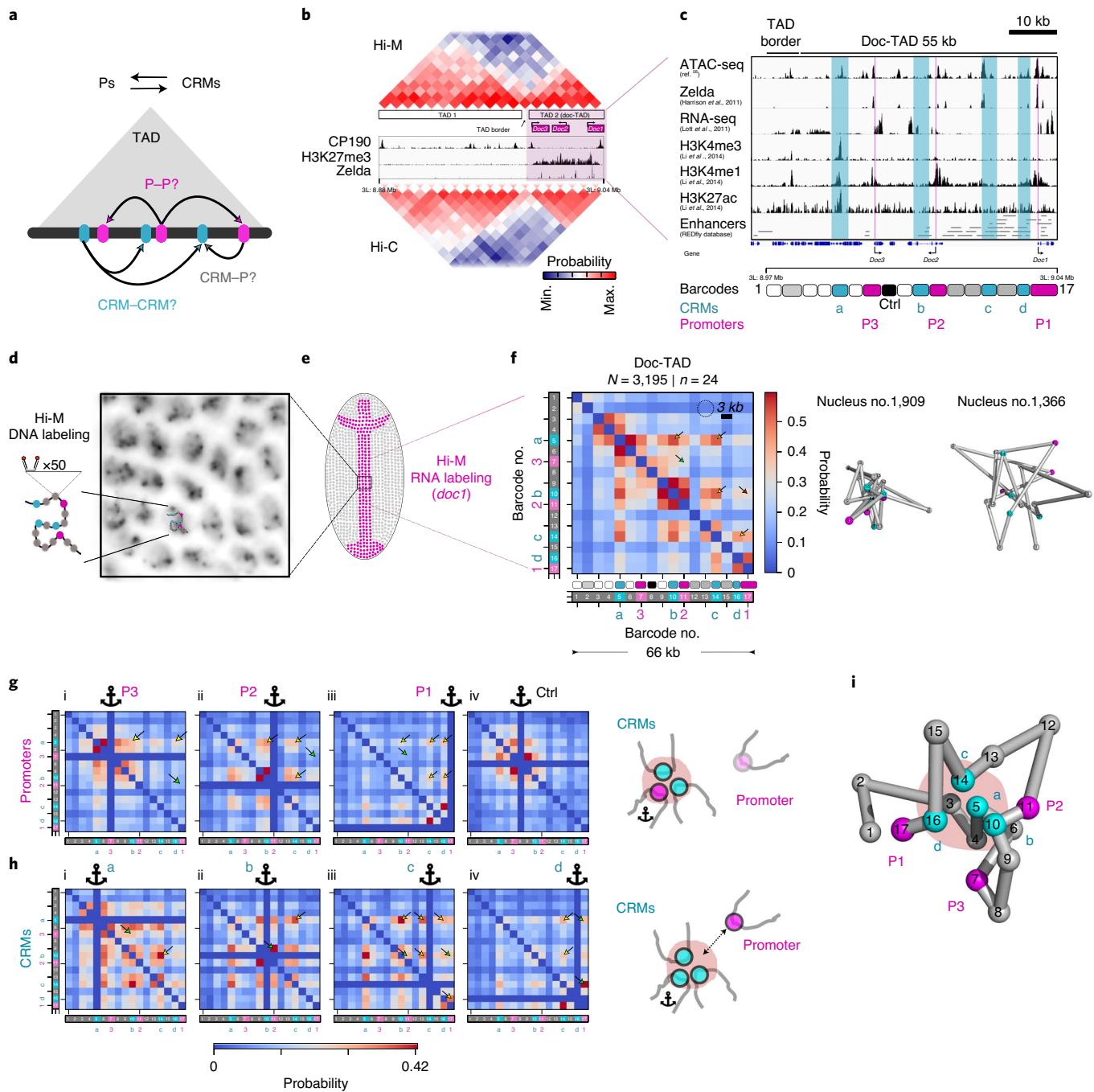


Fig. 1 | Hi-M reveals widespread cis-regulatory chromatin loops and hubs within TADs. **a**, Schematic of the networks of contacts between CRMs and promoters within TADs. **b**, The doc locus (Chr3L:8.88..9.03Mb) in *Drosophila melanogaster*. Low-resolution Hi-M and Hi-C⁴⁶ contact probability maps are shown at the top and bottom, respectively. Blue and red indicate low and high contact probabilities, respectively. **c**, Tracks for chromatin accessibility (ATAC-seq), Zld binding, transcriptional activity (RNA-seq), chromatin marks for active promoters (H3K4me3) and active enhancers (H3K4me1, H3K27ac) from nc14 embryos, and RedFly enhancers are shown. Barcodes used for high-resolution Hi-M included: regions with enhancer marks (CRM_{a-d}), doc promoters (P1-P3), intervening regions with no mark (for example, 'Ctrl') and other regions also documented as enhancers (gray). See Supplementary Table 1 for assignment of CRM_{b-d}. Accession codes for the datasets associated with the cited papers are listed in Supplementary Table 9. **d**, Schematic diagram of the labeling strategy. **e**, Schematic representation of a dorsally oriented *Drosophila* embryo. Segmentation of actively transcribing nuclei (magenta) is based on nascent RNA-FISH labeling. **f**, The high-resolution Hi-M contact probability map of doc-TAD in nuclei displaying doc1 expression in nc14 embryos. Boxes with irregular sizes above barcodes represent the relative genomic lengths of each barcode. Arrows: strong looping interactions of CRMs (yellow), expected P-P interactions (green) and CRM-P interactions (red). Single-allele example reconstructions of spatially clustered and open CRM conformations. Statistics (**f-h**), $N=3,195$ (number of nuclei with doc1 expression), $n=24$ (number of embryos with doc1 expression), $N_T=37,129$ (total number of nuclei), $n_T=29$ (total number of embryos). **g**, Multi-way interactions between promoter regions. Anchoring barcodes are highlighted by a pictogram. A control barcode is depicted in iv. Prominent peaks (yellow) comprise one promoter and two CRMs but not multiple promoters (green). The scheme illustrates the spatial arrangement of CRMs and promoter regions when the anchor is placed at a promoter. **h**, Multi-way interactions between CRMs. Yellow arrows highlight prominent peaks involving three CRMs. The scheme illustrates the spatial arrangement of CRMs and promoter regions when the anchor is placed at a CRM. **i**, The 3D topological representation of the doc-TAD.

Finally, we used ShRec3D⁶² to obtain an ensemble topological reconstruction from the Hi-M matrix. In this reconstruction CRMs clustered at the center of the TAD, whereas promoter elements tended to be at the periphery (Fig. 1*i*). Similarly, we observed that CRMs within sna-TAD also tended to cluster together at the center of the TAD (Extended Data Fig. 2*j*). Collectively, three-way and topological analyses suggest that multiple enhancers physically interact to form CRM hubs. It is of interest that CRM hubs can but do not tend to contain multiple promoters (Supplementary Note 2).

Networks of CRM contacts are indistinguishable between cells of different cell fates. Next, we examined whether chromatin structure in this locus depended on transcriptional status (repression/activation). For this, we used Hi-M in three populations of nuclei established along the dorsoventral axis during nc14: mesoderm (M), neuroectoderm (NE) and dorsal ectoderm (DE)⁶³. To distinguish between these cell fates, we employed RNA-FISH labeling before Hi-M (with *sna* and *doc* probes directly labeling M and DE cells, respectively, Fig. 2*a*). Nuclei were classified as: (1) DE nuclei when an active *doc1* transcription hotspot could be visualized (Extended Data Fig. 2*a*); (2) M nuclei when located within the *sna* expression pattern (Fig. 2*a*); and (3) NE nuclei when located between the pattern of *sna* and the edge of the *doc1* pattern (Fig. 2*a*).

Unexpectedly, Hi-M interaction matrices for DE, NE and M displayed only minor differences (Fig. 2*b-d*), indicating that the same network of chromatin loops is present in nuclei that are actively transcribing and where *doc* gene expression is silent. In addition, 4M profiles were almost identical in nuclei with different cell fates and activation status, independently of whether promoters or CRMs were used as anchors (Fig. 2*e* and Extended Data Fig. 3*a,b*). For example, the *doc1* promoter (P1) showed identical interactions with the four CRMs (CRM_{a-d}) in the DE, the NE and the M (Fig. 2*e,iii,vi*). Likewise, CRM_a and CRM_c displayed patterns of interactions with other CRMs that were indistinguishable between tissues (Fig. 2*e,i,ii,iv,v*). Finally, to detect whether CRM hubs existed in tissues where *doc* genes are repressed, we performed single-nucleus three-way analyses. Indeed, comparison of three-way interaction matrices of NE and M with those of DE revealed the persistence of CRM hubs in nuclei where transcription is repressed (Fig. 2*f*), suggesting that CRM hubs also exist in these cell types. We observed similar results in the *sna* locus (Supplementary Note 3 and Extended Data Fig. 4*a-d*).

To search for a possible explanation of these results, we explored the TF-binding profiles of known activators and repressors in the *doc* locus^{53,64,65} (Supplementary Note 4). CRM_{a-d} are bound by activators in the DE, and tend to be occupied by transcriptional repressors in the M and NE (Fig. 2*g,h*). Thus, contacts between *doc* promoters and CRM hubs in the DE may promote activation and those in the M/NE may instead facilitate repression (Fig. 2*h*).

Cis-regulatory networks emerge before TADs and gene expression. Previous genome-wide and Hi-M studies have established that most *Drosophila* TADs emerge at nc14 during the major wave of zygotic gene activation^{40,46,47}. To explore whether the doc-TAD also emerges at this nuclear cycle, we performed low-resolution Hi-M experiments in nc11–nc12 and nc14 embryos (Fig. 3*a*). We used density of nuclei to unequivocally score developmental timing (Fig. 3*b*, insets). Hi-M contact maps revealed that the doc-TAD can be detected at nc14 but not at earlier stages (Fig. 3*a*), so emergence of this TAD coincides with the onset of *doc* expression (Fig. 3*c*).

To determine whether specific looping interactions between CRMs appear before the emergence of TADs, we performed high-resolution Hi-M between nc11 and nc14. As our previous data showed that Hi-M maps are similar in different presumptive tissues (Fig. 2), we built Hi-M maps for the different nuclear cycles using all detectable nuclei independently of their location in the embryo. Surprisingly, chromatin loops between CRMs were

observed very early in development (nc11) and remained almost unaffected at least until nc14 (Fig. 3*b*). For example, loops between CRM_c and CRM_a, and CRM_b and CRM_d were readily detected as early as nc11, and assumed their final contact frequencies at nc12 (Fig. 3*d*). Similar behaviors were observed when using other CRMs as anchors (Extended Data Fig. 5*a-c*).

These results are consistent with three-way analysis, where we observed that three-way interactions are almost indistinguishable from nc12–nc14 (Fig. 3*e* and Extended Data Fig. 5*d-f*). To gather further evidence for the formation of CRM hubs during early development, we obtained ShRec3D structures for each nuclear cycle. Notably, these structures show that CRMs cluster at the center of the TAD as early as nc11, with clusters becoming tighter as development progresses (Fig. 3*f*). We reached similar conclusions when analyzing the sna-TAD, which also emerges at nc14 (ref. ⁴⁷) (Supplementary Note 5 and Extended Data Fig. 6). Overall, these data indicate that pairwise looping interactions between CRMs in doc- and sna-TADs are established from nc11 (or before), whereas three-way interactions are progressively acquired during development. Importantly, both pairwise and multi-way looping interactions are formed before the emergence of TADs.

To investigate whether specific loops between *doc* promoters and CRMs displayed quantitative changes before the onset of *doc* gene expression (*doc1* is the first to be activated, followed by *doc3* and *doc2*; Fig. 3*c*), we plotted virtual 4M profiles with promoters as viewpoints. Notably, we observed that promoters contact CRMs as early as nc11, and that frequencies of interactions ceased to change after nc12 (see P1 in Fig. 3*g*, and P2–P3 in Extended Data Fig. 5*a-c*). Three-way interactions involving promoters could also be already detected at nc11, and became more frequent at later nuclear cycles (Fig. 3*h* and Extended Data Fig. 5*d-f*). Thus, our results indicate that loops involving promoters and one or several CRMs precede TAD formation and gene expression, and are equally frequent in pluripotent cells, which do not express *doc* genes.

Redundancy of CRM usage at the doc-TAD. To test whether enhancers were redundant in this locus, we searched the literature^{36,66,67} and performed enhancer reporter assays to identify the activation pattern of CRMs located around the *doc* genes (Methods). Several enhancers displayed partially overlapping patterns of activation at this stage of development, particularly in the region around CRM_c (Fig. 4*a,i,j*). This finding is consistent with multiple enhancers being able to activate the transcription of *doc* genes at this developmental stage.

To further test this hypothesis, we deleted CRM_c, which is located midway between two co-regulated promoters (*doc1* and *doc2*) (Methods, Fig. 4*b* and Extended Data Fig. 7*a*). Notably, deletion of CRM_c did not lead to detectable changes in *doc1* or *doc2* expression or on their co-activation frequency (Extended Data Fig. 7*b*), but produced large changes in the organization of the doc-TAD (Fig. 4*c,d*). Interactions involving the barcode originally containing CRM_c were considerably diminished, consistent with the binding of factors to this region being responsible for the formation of long-range CRM and promoter contacts (Fig. 4*c,d*, yellow arrows). Remarkably, we still observed interactions between the *doc1* promoter and the other enhancers (CRM_a, CRM_b and CRM_d), indicating that removal of CRM_c did not affect the ability of the *doc1* promoter to be frequently in proximity to the other CRMs within the doc-TAD (Fig. 4*c,d*, red arrows). Finally, three-way interactions between P1 and other CRMs (CRM_d, CRM_a and CRM_b), as well as with P2, persisted despite deletion of CRM_c (Fig. 4*e*, red arrows). Thus, CRM–CRM and CRM–P interactions still occur in the absence of CRM_c.

Formation of CRM hubs requires the pioneer factor Zld. Having shown that interactions between multiple CRMs do not depend on transcriptional state or developmental timing, we searched for

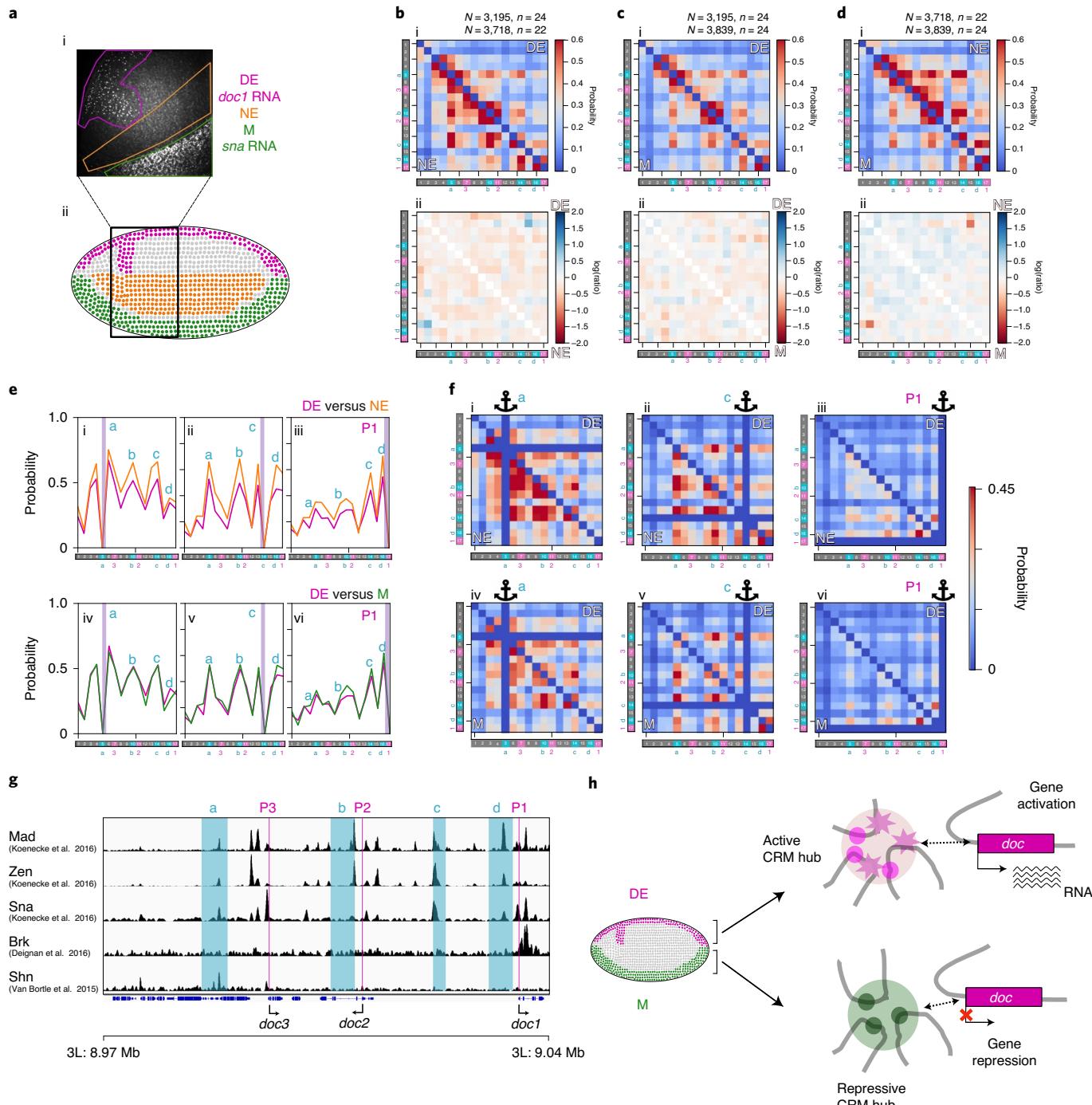


Fig. 2 | CRM-CRM and CRM-P loop frequencies are similar between cell types. **a**, Scheme indicating the three presumptive tissues and their segmentation (ii) based on RNA-FISH labeling (i). **b**, Contact probability maps for DE (upper-right half) and NE (lower-left half) (i). ii, Map of the natural log of the ratio between contact probabilities of DE and NE. Blue indicates a larger contact probability in DE and red in NE. **c**, Similar to **b**, but for DE and M. **d**, Similar to **b**, but for NE and M. **e**, The 4M profiles derived from Hi-M maps for a selected number of anchors. Anchors were placed at CRM_a (i and iv), CRM_c (ii and v) and P1 (iii and vi). **f**, Comparison of three-way contacts for the same tissues and anchors as in **e**. Number of examined nuclei and embryos as in **b-d**, respectively. **g**, ChIP profiles of key transcriptional regulators in the doc-TAD. Accession codes for the datasets associated with the cited papers are listed in Supplementary Table 9. Datasets are from whole embryos at nc14. **h**, Illustration of the double role of CRMs in the doc-TAD.

factors that may be required for the formation of CRM hubs. The pioneer factor Zld has the unusual ability to overcome nucleosome barriers at specific regulatory elements, making them accessible for binding by other classic TFs before activation, as early as nc8-nc11 (refs. ^{54–56,59,68,69}). The doc-TAD is enriched in Zld binding, particularly CRM_{a-d} (Fig. 4f), and Zld was required to ensure proper

expression of *doc* genes, as well as ensuring Pol 2 binding and chromatin accessibility at the *doc* locus (Supplementary Note 6 and Extended Data Fig. 7c–e). To explore whether Zld depletion led to changes in the doc-TAD structure, we performed Hi-M experiments on Zld maternally depleted embryos by using RNA interference (RNAi)⁶⁸. Given the widespread developmental defects

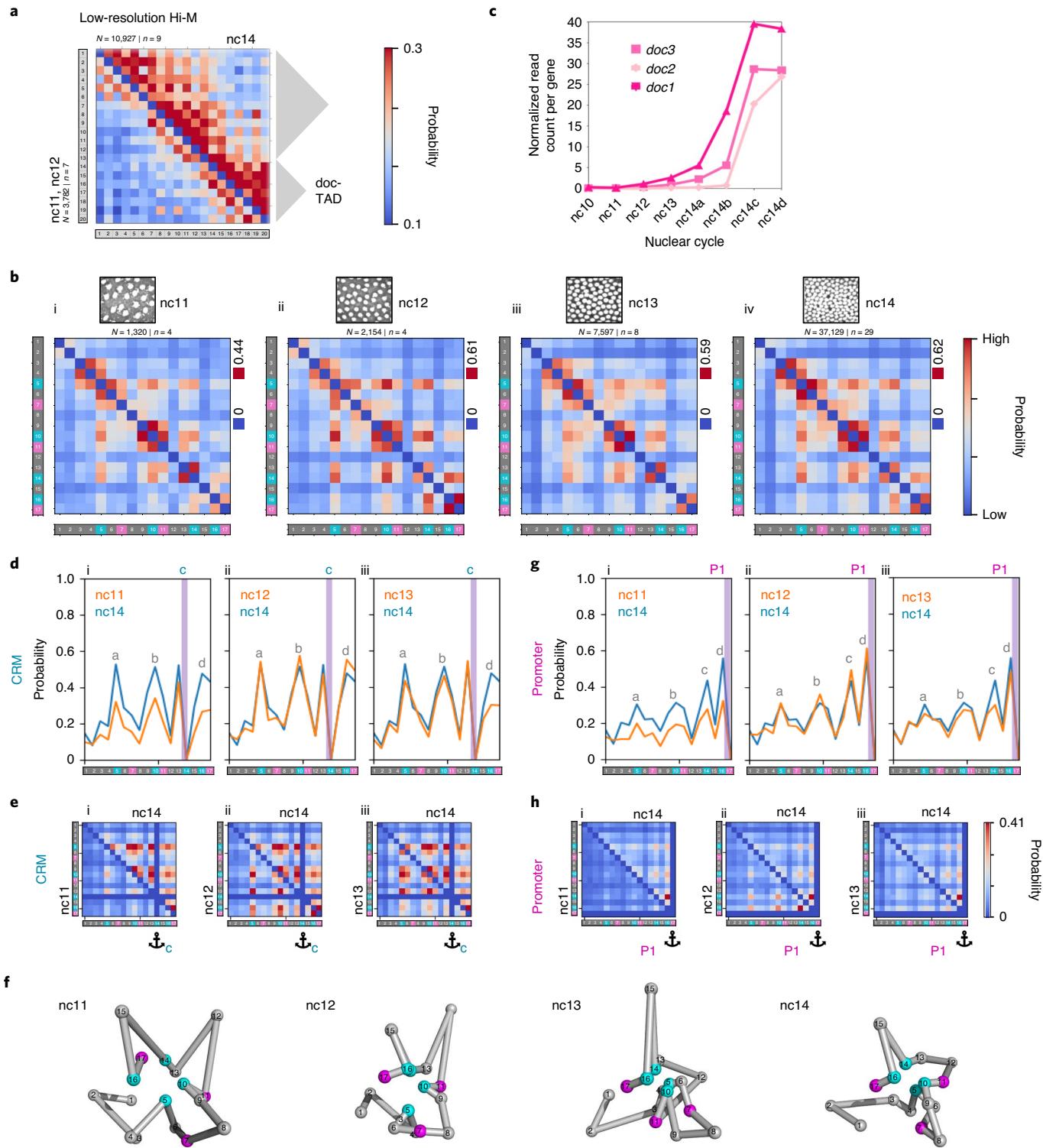
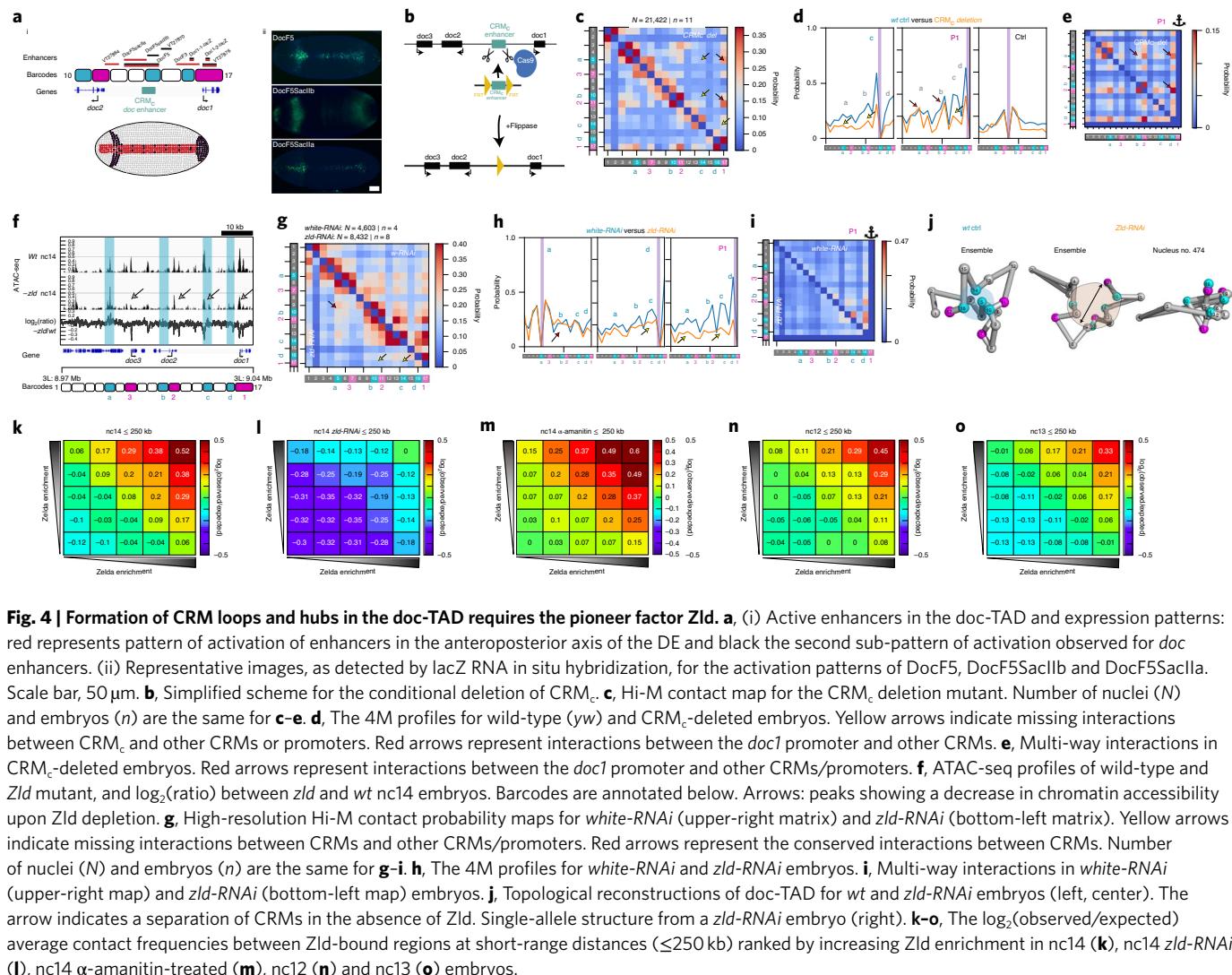


Fig. 3 | CRM loops and hubs precede TAD formation and gene expression. **a**, Low-resolution Hi-M contact probability map of an extended genomic region around the doc-TAD. N , number of nuclei; n , number of embryos. **b**, Representative images of DAPI-stained nuclei for embryos in nc11-nc14 (upper panel). High-resolution Hi-M contact probability maps of the doc-TAD for embryos in nc11-nc14. The minimum and maximum values of the linear color scale are indicated for each matrix using blue and red boxes. **c**, Expression profile of *doc1*, *doc2* and *doc3* during nc10-nc14. The nc14 was divided into four time points according to the extent of cellularization (a, earliest; d, last). Accession codes for the datasets associated with the cited papers are listed in Supplementary Table 9. **d**, Comparison of 4M profiles derived from Hi-M maps at different nuclear cycles. The position of the anchor (CRM_c) is indicated by a vertical purple line. Profiles for nc11, n12 and n13 (orange lines) are compared with nc14 (blue lines) in i-iii, respectively. **e**, Comparison of three-way contacts between nc14 and other nuclear cycles, using CRM_c as anchor. Upper-right half of the matrix always depicts nc14, and the bottom-left half shows nc11 (i), nc12 (ii) and nc13 (iii). Number of examined nuclei and embryos are as indicated for the respective nuclear cycle in **c**. **f**, Topological reconstructions of the doc-TAD for nc11-nc14. CRMs and promoter regions are indicated as cyan and magenta spheres, respectively. **g**, Similar to **c**, but anchor: *doc1* promoter (P1). **h**, Similar to **d**, but anchor: *doc1* promoter (P1).



exhibited by *Zld* RNAi embryos at stage 5 (ref. ⁷⁰), we restricted our analyses to early nc14 *Zld* RNAi embryos.

Depletion of *Zld* did not affect TAD borders at the *doc* locus (Extended Data Fig. 7f) or change the overall compaction of the doc-TAD (Extended Data Figs. 1i and 7g). However, we observed large changes in networks of CRM-CRM and CRM-P interactions (Fig. 4g). Contacts between CRM_a and other CRMs or promoters were very similar in *Zld*-depleted and RNAi *white* control embryos (Fig. 4g,h, red arrows, and Supplementary Note 7). Remarkably, the main differences in contacts occurring in *Zld*-bound genomic regions that lose accessibility upon *Zld* depletion (CRM_c, CRM_d, CRM_b, P1, P2 and P3; Fig. 4g, yellow arrows, and Fig. 4h, middle and right). For example, on *Zld* depletion, CRM_c showed the largest drop in ATAC-seq signal among CRMs (Fig. 4f), and a dramatic drop in its interactions with other CRMs (Fig. 4g,h, yellow arrows). Finally, in *Zld* RNAi embryos, formation of CRM hubs was also considerably impacted (Fig. 4i) and topological reconstructions showed a loss of CRM clustering (Fig. 4j). Altogether, these results suggest a model whereby the pioneering activity of *Zld* plays a key role in the activation of *doc* genes and participates in the formation of CRM-CRM and CRM-P loops during early embryogenesis, possibly through its ability to open chromatin at specific CRMs.

To shed further light on the role of *Zld* in the formation of preferential interaction networks, we selected 5,038 genomic regions displaying *Zld* binding and calculated their pairwise, intra-arm

interaction frequencies⁷¹ using publicly available datasets⁴⁰. In nc14 embryos, *Zld*-bound regions interacted more frequently with each other than with control regions (Fig. 4k). This bias increased with the level of *Zld* binding and was present for short- (< 250 kb) and long-range (> 250 kb) genomic distances (Fig. 4k and Extended Data Fig. 7h). *Zld* depletion led to a considerable decrease in interactions between *Zld*-bound genomic regions (Fig. 4l and Extended Data Fig. 7i), consistent with *Zld* depletion results in the doc-TAD. However, this decrease in interactions was not observed upon transcriptional inhibition (Fig. 4m and Extended Data Fig. 7j), in agreement with other analyses⁴⁰. Importantly, interactions between *Zld*-bound genomic regions were already present in nc12-nc13 embryos (Fig. 4n,o), and were also specific to *Zld*-bound pre-midblastula-transition (MBT) enhancers (Supplementary Note 8 and Extended Data Fig. 7k). Overall, these results are consistent with *Zld* being needed for the formation of a subset of CRM-CRM interactions during early embryogenesis.

Discussion

In the present study, we use a high-resolution, imaging-based, single-cell spatial genomics approach (Hi-M) to link chromosome topology and transcriptional regulation during early *Drosophila* development. This approach has notable advantages, such as the detection of multi-way interactions and transcriptional output with spatial resolution (Supplementary Note 9). We reveal extensive

interaction networks within developmental TADs primarily involving CRMs. Critically, these networks arise thanks to the spatial clustering of multiple enhancers (CRM hubs) and are mostly invariant during cell fate specification and gene activation. Networks of pairwise CRM contacts and CRM hubs arise during early development, before the onset of gene expression and before the emergence of TADs, and require the pioneering activity of the transcription factor Zld.

One of the important results of the present study is that physical proximity between multiple CRMs and promoters is observed with very similar frequencies in cells with distinct fates and appeared during early embryogenesis. These results are consistent with those obtained at later stages of *Drosophila* embryogenesis, showing that enhancers located at considerably larger distances (~100 kb) can also form binary loops that are present in cells from different tissues¹⁷. Similarly, E-P interactions at the mouse *Hoxd* locus were detected in tissues where target genes were not expressed¹⁸. Our results are further supported by a companion paper⁷² that applied Hi-C and micro-C to study tissue-specific *Drosophila* chromosome organization at similar stages of development (Supplementary Note 10). From a developmental perspective, the formation of loops between promoters and distal regulatory elements in cells where genes need to be repressed can be seen as a ‘dangerous liaison’. Indeed, once a loop has been established, transcriptional activation could rapidly occur in cells where that specific promoter should be kept inactive.

This apparent dichotomy can, however, be rationalized in terms of the spatiotemporal patterning of the *cis*-regulatory logic of TFs during embryogenesis. For instance, in M cells, most *doc* CRMs are bound by the spatially localized transcriptional repressor Sna⁴⁴, which acts as a silencer in the M. In this case, communication between promoters and distal CRMs may reinforce transcriptional repression. This interpretation is in agreement with the finding that many enhancers can act as silencers in alternate cell types during *Drosophila* development⁷³; however, other silencing mechanisms may also be at play⁷⁴. Thus, we hypothesize that the optimal mechanism to ensure rapid and efficient activation or repression during development may involve two steps: the rapid priming of key CRMs via ubiquitously maternally deposited pioneer factors (for example, Zld), followed by regulation of transcriptional output by spatially and temporally localized transcriptional activators and repressors. In this model, three-dimensional (3D) chromatin architecture plays a double role because 3D contacts could serve to reinforce both activation and repression at a particular developmental stage while allowing for flexibility at later stages. For example, a repressive CRM loop in a tissue at an early developmental stage may switch to a CRM loop with activation capacities at later stages by changing TF occupancy. Future experiments testing whether CRM loops and hubs display more differences in active and repressed tissues at later stages of development will be important to test these hypotheses.

Previous studies suggested that invariant E-P loops may be pre-established and stable^{14,17,75,76}. In agreement with these results, our data indicate that E-P loops can form early, well before the onset of gene expression. However, in all cases, we measured low frequencies of looping interactions between functional elements. These results are consistent with previous measurements of absolute contact frequencies within TADs and between E-Ps^{49,77–79}. Thus, these results indicate that different sets of multi-way E-E and E-P contacts are present in different cells, and that these contacts may be highly dynamic.

Recent studies reported the existence of enhancer hubs: spatially localized clusters containing multiple enhancers^{29,30,33} that may facilitate transcriptional activation by creating a local microenvironment whereby transcriptional resources are shared, akin to early models of ‘transcription hubs’⁸⁰. Formation of enhancer hubs may require interactions between components of the transcriptional machinery, which could contribute to, or result from, the assembly

of phase-separated condensates^{32,37,38,81–83}. In this model, enhancers need not directly touch their target promoters but merely come into close proximity (~200–300 nm)^{25,84}. Overall, these findings and models are consistent with our observation that multiple endogenous CRMs within a TAD come together in space to form hubs in single, actively transcribing nuclei. We also observed the formation of similar hubs in inactive nuclei, suggesting that repressive elements may also form spatially localized clusters of transcriptional repressors to share resources and reinforce their silencing activities. CRM hubs are formed at early stages of development in pluripotent cells. Thus, we favor a model in which preferential CRM interaction networks are pre-formed at early stages and are subsequently specified (into activation or repression hubs) during nc14 or later.

In *Drosophila*, TADs emerge concomitant with the major wave of zygotic gene activation^{40,46,47}. Previous studies reported the existence of chromatin loops typically at considerably large genomic distances spanning two or more TADs^{17,46} or concerning Polycomb-binding sites^{46,85}. In the present study, we observed that chromatin loops between CRMs within *Drosophila* TADs are widespread, mimicking the common CTCF-mediated chromatin loops present within mammalian TADs^{15,41}. In addition, we found that multiple CRMs can cluster together to form *cis*-regulatory hubs located within TADs, suggesting a mechanism to sequester enhancers in space to reduce the activation of genes in neighboring TADs. Importantly, formation of CRM hubs precedes the emergence of TADs, consistent with the finding in mammalian cells that subsets of E-P contacts arise rapidly after mitosis before TADs are re-formed⁸⁶. Thus, our results suggest that CRM hubs and TADs probably form by different mechanisms. Overall, we hypothesize that CRM hubs represent an additional functional level of genome organization, independent of TADs. This additional layer can also be regulated by priming of enhancers and promoters by paused polymerases^{87–89} or pioneer factors^{54,55}, as well as by chromatin marks^{90,91}. As interactions between Zld CRMs appear before TADs, it is unclear how specificity of CRM interactions may be regulated to favor intra-TAD contacts (Supplementary Note 11).

It is interesting that we observed that interactions between Zld-bound CRMs, as well as interactions between CRMs and cognate promoters, are established very early in pluripotent nuclei, before cell fate commitments. These long-range interactions occur between related CRMs (within doc- and sna-TADs) as well as between unrelated but Zld-bound CRMs (Fig. 4k and Extended Data Fig. 7k), suggesting that a common link could be their regulation by broad factors such as Zld. Critically, preferential contacts involving Zld-bound CRMs were considerably attenuated upon Zld depletion. We and others have recently shown that Zld forms nuclear hubs in early *Drosophila* embryos^{35,36}, and that Zld hubs are re-established by the end of mitosis, before transcriptional activation. Taken together, our results suggest that Zld fosters the formation of CRM hubs by rendering chromatin accessible during early development, as a first step of cell specification to ensure maximum plasticity. Future work involving the detection of a larger number of CRMs will be needed to elucidate the factors and mechanisms involved in spatial clustering of developmental CRMs into nuclear microenvironments.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-021-00816-z>.

Received: 6 July 2020; Accepted: 9 February 2021;
Published online: 1 April 2021

References

1. Bickmore, W. A. The spatial organization of the human genome. *Annu. Rev. Genomics Hum. Genet.* **14**, 67–84 (2013).
2. Cavalli, G. & Misteli, T. Functional implications of genome topology. *Nat. Struct. Mol. Biol.* **20**, 290–299 (2013).
3. Bulger, M. & Groudine, M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**, 327–339 (2011).
4. Alberts, B. et al. *Molecular Biology of the Cell* 6th edn (W. W. Norton & Company, 2017).
5. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**, 499–506 (2013).
6. Fukaya, T., Lim, B. & Levine, M. Enhancer control of transcriptional bursting. *Cell* **166**, 358–368 (2016).
7. Bartman, C. R., Hsu, S. C., Hsiung, C. C.-S., Raj, A. & Blobel, G. A. Enhancer regulation of transcriptional bursting parameters revealed by forced chromatin looping. *Mol. Cell* **62**, 237–247 (2016).
8. Schwarzer, W. & Spitz, F. The architecture of gene expression: integrating dispersed *cis*-regulatory modules into coherent regulatory domains. *Curr. Opin. Genet. Dev.* **27**, 74–82 (2014).
9. Symmons, O. et al. Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400 (2014).
10. Lüpianez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene–enhancer interactions. *Cell* **161**, 1012–1025 (2015).
11. Ji, X. et al. 3D chromosome regulatory landscape of human pluripotent cells. *Cell Stem Cell* **18**, 262–275 (2016).
12. Dowen, J. M. et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014).
13. Ron, G., Globerson, Y., Moran, D. & Kaplan, T. Promoter–enhancer interactions identified from Hi-C data using probabilistic models and hierarchical topological domains. *Nat. Commun.* **8**, 2237 (2017).
14. Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosome topology. *Science* **361**, 1341–1345 (2018).
15. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **162**, 687–688 (2015).
16. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).
17. Ghavi-Helm, Y. et al. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100 (2014).
18. Montavon, T. et al. A regulatory archipelago controls *Hox* genes transcription in digits. *Cell* **147**, 1132–1145 (2011).
19. Chen, H. et al. Dynamic interplay between enhancer–promoter topology and gene activity. *Nat. Genet.* **50**, 1296–1303 (2018).
20. Bartman, C. R. et al. Transcriptional burst initiation and polymerase pause release are key control points of transcriptional regulation. *Mol. Cell* **73**, 519–532.e4 (2019).
21. Morgan, S. L. et al. Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nat. Commun.* **8**, 15993 (2017).
22. Deng, W. et al. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**, 1233–1244 (2012).
23. Larsson, A. J. M. et al. Genomic encoding of transcriptional burst kinetics. *Nature* **565**, 251–254 (2019).
24. Alexander, J. M. et al. Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *eLife* **8**, e41769 (2019).
25. Benabdallah, N. S. et al. Decreased enhancer–promoter proximity accompanying enhancer activation. *Mol. Cell* **76**, 473–484.e7 (2019).
26. McCord, R. P., Kaplan, N. & Giorgetti, L. Chromosome conformation capture and beyond: toward an integrative view of chromosome structure and function. *Mol. Cell* **77**, 688–708 (2020).
27. van Steensel, B. & Furlong, E. E. M. The role of transcription in shaping the spatial organization of the genome. *Nat. Rev. Mol. Cell Biol.* **20**, 327–337 (2019).
28. Lim, B. & Levine, M. S. Enhancer–promoter communication: hubs or loops? *Curr. Opin. Genet. Dev.* **67**, 5–9 (2021).
29. Allayyar, A. et al. Enhancer hubs and loop collisions identified from single-allele topologies. *Nat. Genet.* **50**, 1151–1160 (2018).
30. Oudelaar, A. M. et al. Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nat. Genet.* **50**, 1744–1751 (2018).
31. Oudelaar, A. M. et al. A revised model for promoter competition based on multi-way chromatin interactions at the α -globin locus. *Nat. Commun.* **10**, 5412 (2019).
32. Hsieh, T.-H. S. et al. Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol. Cell* **78**, 539–553.e8 (2020).
33. Tsai, A., Alves, M. R. & Crocker, J. Multi-enhancer transcriptional hubs confer phenotypic robustness. *eLife* **8**, e45325 (2019).
34. Baudement, M.-O. et al. High-salt-recovered sequences are associated with the active chromosomal compartment and with large ribonucleoprotein complexes including nuclear bodies. *Genome Res.* **28**, 1733–1746 (2018).
35. Mir, M. et al. Dynamic multifactor hubs interact transiently with sites of active transcription in embryos. *eLife* **7**, e40497 (2018).
36. Dufourt, J. et al. Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. *Nat. Commun.* **9**, 5194 (2018).
37. Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).
38. Boija, A. et al. Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842–1855.e16 (2018).
39. Tsai, A. et al. Nuclear microenvironments modulate transcription from low-affinity enhancers. *eLife* **6**, e28975 (2017).
40. Hug, C. B., Grimaldi, A. G., Kruse, K. & Vaquerizas, J. M. Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell* **169**, 216–228.e19 (2017).
41. Vallot, A. & Tachibana, K. The emergence of genome architecture and zygotic genome activation. *Curr. Opin. Cell Biol.* **64**, 50–57 (2020).
42. Porcher, A. & Dostatni, N. The bicoid morphogen system. *Curr. Biol.* **20**, R249–R254 (2010).
43. Bonn, S. & Furlong, E. E. M. *cis*-Regulatory networks during development: a view of *Drosophila*. *Curr. Opin. Genet. Dev.* **18**, 513–520 (2008).
44. Stathopoulos, A. & Levine, M. Localized repressors delineate the neurogenic ectoderm in the early *Drosophila* embryo. *Dev. Biol.* **280**, 482–493 (2005).
45. Schulz, K. N. & Harrison, M. M. Mechanisms regulating zygotic genome activation. *Nat. Rev. Genet.* **20**, 221–234 (2019).
46. Ogihara, Y., Schuettengruber, B., Papadopoulos, G. L., Chang, J.-M. & Cavalli, G. Polycomb-dependent chromatin looping contributes to gene silencing during *Drosophila* development. *Mol. Cell* **71**, 73–88.e5 (2018).
47. Cardozo-Gizzi, A. M. et al. Microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms. *Mol. Cell* <https://doi.org/10.1016/j.molcel.2019.01.011> (2019).
48. Cardozo-Gizzi, A. M. et al. Direct and simultaneous observation of transcription and chromosome architecture in single cells with Hi-M. *Nat. Protoc.* **15**, 840–876 (2020).
49. Mateo, L. J. et al. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54 (2019).
50. Beliveau, B. J. et al. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. USA* **109**, 21301–21306 (2012).
51. Reim, I. & Frasch, M. The Dorsocross T-box genes are key components of the regulatory network controlling early cardiogenesis in *Drosophila*. *Development* **132**, 4911–4925 (2005).
52. Hamm, D. C. et al. A conserved maternal-specific repressive domain in Zelda revealed by Cas9-mediated mutagenesis in *Drosophila melanogaster*. *PLoS Genet.* **13**, e1007120 (2017).
53. Koenecke, N., Johnston, J., Gaertner, B., Natarajan, M. & Zeitlinger, J. Genome-wide identification of *Drosophila* dorso-ventral enhancers by differential histone acetylation analysis. *Genome Biol.* **17**, 196 (2016).
54. Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R. & Eisen, M. B. Zelda binding in the early *Drosophila melanogaster* embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet.* **7**, e1002266 (2011).
55. Nien, C.-Y. et al. Temporal coordination of gene networks by Zelda in the early *Drosophila* embryo. *PLoS Genet.* **7**, e1002339 (2011).
56. Hannon, C. E., Blythe, S. A. & Wieschaus, E. F. Concentration dependent chromatin states induced by the bicoid morphogen gradient. *eLife* **6**, e28275 (2017).
57. Li, X.-Y., Harrison, M. M., Villalta, J. E., Kaplan, T. & Eisen, M. B. Establishment of regions of genomic activity during the *Drosophila* maternal to zygotic transition. *eLife* **3**, 127–144.e23 (2014).
58. Rivera, J., Keränen, S. V. E., Gallo, S. M. & Halfon, M. S. REDfly: the transcriptional regulatory element database for *Drosophila*. *Nucleic Acids Res.* **47**, D828–D834 (2019).
59. Blythe, S. A. & Wieschaus, E. F. Establishment and maintenance of heritable chromatin structure during early embryogenesis. *eLife* **5**, e20148 (2016).
60. Senecal, A. et al. Transcription factors modulate c-Fos transcriptional bursts. *Cell Rep.* **8**, 75–83 (2014).
61. Hao, N., Shearwin, K. E. & Dodd, I. B. Positive and negative control of enhancer–promoter interactions by other DNA loops generates specificity and tunability. *Cell Rep.* **26**, 2419–2433.e3 (2019).
62. Lesne, A., Riposo, J., Roger, P., Cournac, A. & Mozziconacci, J. 3D genome reconstruction from chromosomal contacts. *Nat. Methods* **11**, 1141–1143 (2014).
63. Stein, D. S. & Stevens, L. M. Maternal control of the *Drosophila* dorsal–ventral body axis. *Wiley Interdiscip. Rev. Dev. Biol.* **3**, 301–330 (2014).
64. Deignan, L. et al. Regulation of the BMP signaling-responsive transcriptional network in the *Drosophila* embryo. *PLoS Genet.* **12**, e1006164 (2016).
65. Van Bortle, K., Peterson, A. J., Takenaka, N., O'Connor, M. B. & Corces, V. G. CTCF-dependent co-localization of canonical Smad signaling factors at architectural protein binding sites in *D. melanogaster*. *Cell Cycle* **14**, 2677–2687 (2015).

66. Kwon, E. Z. et al. Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo. *Nature* **512**, 91–95 (2014).
67. Weiss, A. et al. A conserved activation element in BMP signaling during *Drosophila* development. *Nat. Struct. Mol. Biol.* **17**, 69–76 (2010).
68. Sun, Y. et al. Zelda overcomes the high intrinsic nucleosome barrier at enhancers during *Drosophila* zygotic genome activation. *Genome Res.* **25**, 1703–1714 (2015).
69. Schulz, K. N. et al. Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early *Drosophila* embryo. *Genome Res.* **25**, 1715–1726 (2015).
70. Liang, H.-L. et al. The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. *Nature* **456**, 400–403 (2008).
71. Loubiere, V., Papadopoulos, G. L., Szabo, Q., Martinez, A.-M. & Cavalli, G. Widespread activation of developmental gene expression characterized by PRC1-dependent chromatin looping. *Sci. Adv.* **6**, eaax4001 (2020).
72. Ing-Simmons, E. et al. Independence of chromatin conformation and gene regulation during *Drosophila* dorsoventral patterning. *Nat. Genet.* <https://doi.org/10.1038/s41588-021-00799-x> (2021).
73. Gisselbrecht, S. S. et al. Transcriptional silencers in *Drosophila* serve a dual role as transcriptional enhancers in alternate cellular contexts. *Mol. Cell* **77**, 324–337.e8 (2020).
74. Chopra, V. S., Kong, N. & Levine, M. Transcriptional repression via antilooping in the *Drosophila* embryo. *Proc. Natl Acad. Sci. USA* **109**, 9460–9464 (2012).
75. Rubin, A. J. et al. Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat. Genet.* **49**, 1522–1528 (2017).
76. Paliou, C. et al. Preformed chromatin topology assists transcriptional robustness of during limb development. *Proc. Natl Acad. Sci. USA* **116**, 12390–12399 (2019).
77. Cattoni, D. I. et al. Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. *Nat. Commun.* **8**, 1753 (2017).
78. Finn, E. H. et al. Extensive heterogeneity and intrinsic variation in spatial genome organization. *Cell* **176**, 1502–1515.e10 (2019).
79. Gu, B. et al. Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science* **359**, 1050–1055 (2018).
80. Jackson, D. A., Iborra, F. J., Manders, E. M. & Cook, P. R. Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol. Biol. Cell* **9**, 1523–1536 (1998).
81. Cho, W.-K. et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).
82. Chong, S. et al. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **361**, eaar2555 (2018).
83. Lesne, A., Baudem, M.-O., Rebouissou, C. & Forné, T. Exploring mammalian genome within phase-separated nuclear bodies: experimental methods and implications for gene expression. *Genes* **10**, 1049 (2019).
84. Lim, B., Heist, T., Levine, M. & Fukaya, T. Visualization of transvection in living *Drosophila* embryos. *Mol. Cell* **70**, 287–296.e6 (2018).
85. Eagen, K. P., Aiden, E. L. & Kornberg, R. D. Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *Proc. Natl Acad. Sci. USA* <https://doi.org/10.1073/pnas.1701291114> (2017).
86. Zhang, H. et al. Chromatin structure dynamics during the mitosis-to-G1 phase transition. *Nature* **576**, 158–162 (2019).
87. Ghavi-Helm, Y. et al. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* **51**, 1272–1282 (2019).
88. Lagha, M. et al. Paused Pol II coordinates tissue morphogenesis in the *Drosophila* embryo. *Cell* **153**, 976–987 (2013).
89. Saunders, A., Core, L. J., Sutcliffe, C., Lis, J. T. & Ashe, H. L. Extensive polymerase pausing during *Drosophila* axis patterning enables high-level and pliable transcription. *Genes Dev.* **27**, 1146–1158 (2013).
90. Negre, N. et al. A cis-regulatory map of the *Drosophila* genome. *Nature* **471**, 527–531 (2011).
91. Sexton, T. et al. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

Drosophila stocks and embryo collection. Fly stocks were maintained at room temperature (RT) with a natural light–dark cycle and raised in standard cornmeal yeast medium. The *yw* or *UASp-shRNA-w* (BL no. 35573) stocks were used as a control. Flippase stock (BL no. 26902) and CRE stock (BL no. 851) were used for the generation of the CRM_c deletion strain. Zld-depleted embryos were obtained from females from the cross between *nos-Gal4:VP16* (BL no. 4937) and *UASp-shRNA-zld*⁴⁸, a method with a Zld depletion efficiency of ~90% (ref. ³⁶). After a pre-laying step, flies were allowed to lay eggs for 1.5 h on new yeasted 0.4% acetic acid plates. Embryos were then incubated at 25 °C until they reached the desired developmental stage. Embryos were collected and fixed as previously described⁴⁸. Briefly, embryos were dechorionated with 2.6% bleach, rinsed and fixed with a 1:1 mixture of 4% methanol-free formaldehyde in phosphate-buffered saline (PBS) and heptane. Embryos were stored in methanol at -20 °C until further use.

Generation of the CRM_c deletion and reporter lines. A CRISPR (clustered regularly interspaced short palindromic repeats)–Cas9-based strategy was employed to specifically delete CRM_c in a conditional manner. Two flippase recognition target (FRT) sites were inserted flanking an ~860-bp region (3L:9021947–9022805) surrounding the most prominent Zld peaks in CRM_c (impacting barcodes 13–14). Recombination template was cloned into pHD-DsRed plasmid (Addgene, catalog no. 1434). For this, a 5'-homology arm (PCR amplified from genomic DNA) was inserted into a vector previously digested with XmaI/NheI. Then, the PCR-amplified 3'-homology arm was inserted after digestion with SpeI/Ascl. Finally, the PCR-amplified CRM_c flanked by FRT sequences was cloned after digestion by NotI. Guide RNAs were cloned into pCFD3-dU6:3gRNA (Addgene, catalog no. 49410) and digested by BbsI using annealed oligonucleotides. Recombination template and guide RNAs were injected by BestGene Inc. After obtaining the CRISPR-edited stocks, males were crossed with *CRE/CRE; D*/TM3, Sb* virgin females to remove the dsRed marker by the action of a Cre recombinase. Then, *dsRED*^{-/-} males were crossed with *Dp/TM3, Sb* virgin females. Males from this cross were then crossed with *hs-FLP/hs-FLP; Dr/TM3, Sb* virgin females. Larvae from these crosses were heat shocked at 37 °C for 30 min in a water bath for the flippase to be expressed and CRM_c to be deleted. Adult males were then PCR genotyped. Oligonucleotide sequences used for cloning and genotyping are listed in Supplementary Table 2. The deletion removes 287 bp of barcode 13 (5 oligonucleotides out of 50) and 562 bp (10 oligonucleotides out of 50) of barcode 14.

DocFX-lacZ reporter lines were generated by promoter transgenesis of pH-Pelican vectors with CRM fragments in between KpnI and NotI sites analogous to the procedure described in Kahn et al.⁹². Subfragments F5SacIIa and F5SacIIb were generated by removing KpnI-SacII or SacII-NotI fragments, respectively, with subsequent blunt end religation.

Hi-M libraries. Oligopaint libraries, consisting of unique ~35/41-mer sequences with genome homology, were obtained from the Oligopaint public database (<http://genetics.med.harvard.edu/oligopaints>). We selected 20 barcodes in the *doc* locus (3L:8882600..9039000 *Drosophila* release 5 reference genome in all cases) for the low-resolution Hi-M library, 17 barcodes encompassing the doc-TAD (3L:8974562..9038920) for the high-resolution Hi-M library and 65 barcodes (2L:15244500..15630000) for the high-resolution *sna* locus library. For each barcode, we used 45–50 probes, covering ~3 kb. An additional fiducial barcode located at least ~1 Mb away was used for drift correction (see below). The coordinates of the targeted genomic regions are listed in Supplementary Table 3.

Each oligonucleotide in the pool consisted of five regions: (1) a 21-mer forward priming region; (2) a 32-mer (low-res library) or two 20-mers separated by an AT sequence (high-res libraries) readout region unique for each barcode; (3) a 35/41-mer genome homology region; (4) a 32-mer (low-res library) or 20-mer (high-res libraries) readout region; and (5) a 21-mer reverse priming region. The designed template oligonucleotide pools were ordered from CustomArray. The procedure to amplify oligonucleotide pools to obtain the primary libraries was as previously described⁴⁸. It involved a five-step procedure consisting of (1) limited-cycle PCR, (2) amplification via T7 in vitro transcription, (3) reverse transcription, (4) alkaline hydrolysis and (5) purification. The sequences of the primers used for amplification of the libraries are listed in Supplementary Table 4.

For the low-resolution library, we employed 20 unique Alexa Fluor-647-labeled sequence oligonucleotides (imaging oligonucleotides), complementary to the readout region present in the primary oligonucleotide. The fluorophore was attached via a disulfide linkage cleavable by the mild reducing agent tris(2-carboxyethyl)phosphine (TCEP), using a previously described strategy⁴⁸. Alternatively, for the high-resolution libraries, we used ‘adapter’ oligonucleotides, consisting of a 20-mer region complementary to the readout sequence that can recognize the barcode being targeted, a 10-mer spacer sequence and a 32-mer region able to bind to a unique Alexa Fluor-647-labeled oligonucleotide (containing a disulfide linkage). In this approach, a single fluorescent oligonucleotide is required⁴⁹. For fiducial barcodes, a noncleavable, rhodamine-labeled oligonucleotide was used. The sequences of the imaging and adapter oligonucleotides are listed in Supplementary Table 5. PCR and reverse transcription primers used in

probe synthesis, as well as adapter oligonucleotides and fluorescently labeled oligonucleotides, were purchased from Integrated DNA Technology. The whole set of Oligopaints used can be found in Supplementary Table 6.

RNA–FISH probes. RNA probes were obtained by in vitro transcription from a vector containing the sequences targeting *sna* (previously described in ref. ⁴⁷), *doc1*, *doc2* or *doc3* genes in the presence of digoxigenin (DIG) or biotin (BIO) haptene. Vector was linearized before the in vitro transcription with a specific restriction enzyme. RNA probes produced in this manner were then treated with carbonate buffer at 65 °C for 5 min (*sna* probe) or 2 min (*doc1*, *doc2* and *doc3* probes). The information on each probe, including the primers used to clone the target sequences by amplification of genomic DNA, is listed in Supplementary Table 7.

RNA–FISH. In situ hybridization was as described previously⁴⁸, with modifications to allow for the detection of two different species of RNA. The reader is invited to read our detailed protocol in the aforementioned reference. Briefly, fixed embryos were passed through a 1:1 mixture of methanol:ethanol and then pure ethanol. Embryos were then post-fixed with 5% formaldehyde in PBS-T (PBS-T = 0.1% Tween-20 + PBS) for 25 min. Then, embryos were incubated four times with PBS-T for 15 min and permeabilized for 1 h with 0.3% Triton X-100 in PBS. Embryos were rinsed with PBS-T and incubated for 2 h with RHS at 55 °C (where RHS = 50% formamide, 2× saline sodium citrate (SSC), 0.1% Tween-20, 0.05 mg ml⁻¹ of heparin and 0.1 mg ml⁻¹ of salmon sperm). In the meantime, RNA probes were heated at 85 °C for 2 min, transferred to ice for 2 min and then incubated with the embryos in RHS for 16–20 h at 55 °C for RNA hybridization. The next day, embryos were washed four times with RHS at 55 °C and three times with PBS-T at RT. Then, a saturation step was performed with blocking solution (blocking reagent (Sigma-Aldrich, catalog no. 11096176001): 100 mM maleic acid and 150 mM NaCl, pH 7.5) for 45 min.

Then the protocol depends on whether embryos were used for Hi-M (*sna/doc1* double labeling) or to reveal *doc1*, *doc2* and *doc3* expression patterns (Extended Data Fig. 1j). To reveal the expression patterns of *doc* genes, the combination *doc1*-DIG/*doc2*-BIO or *doc2*-DIG/*doc3*-BIO was used. After the saturation step, embryos were incubated with primary antibodies at 1:375 dilution (sheep anti-DIG (Roche, catalog no. 11330389001) and mouse anti-BIO (Life Technologies, catalog no. 03-3700)) overnight at 4 °C. The next day embryos were washed six times in PBS-T for 10 min. Embryos were incubated for 1 h in blocking solution, then for 2 h with secondary antibodies at 1:500 dilution (anti-mouse Alexa Fluor-488 conjugated (Life Technologies, catalog no. A21202)) and anti-sheep Alexa Fluor-555 conjugated (Life Technologies, catalog no. A21436) and washed six times in PBS-T. Finally, embryos were incubated for 10 min with 0.5 mg ml⁻¹ of DAPI solution, washed with PBS-T and mounted in ProLong Diamond Antifade.

For Hi-M, both *sna* and *doc1* probes were DIG labeled. By taking advantage of the differential spatial expression pattern, we labeled both RNAs simultaneously by the combination of both probes during incubation, and the use of a single anti-DIG antibody and a tyramide signal amplification reaction. After RNA hybridization and the saturation step, the activity of endogenous peroxidases was eliminated by incubating with 1% H₂O₂ in PBS-T for 30 min. After rinsing with PBS-T, embryos were incubated overnight at 4 °C with sheep anti-DIG conjugated with POD (Sigma-Aldrich, catalog no. 11207733910) with 1:500 working dilution in PBS-T. The next day, embryos were washed with PBS-T and incubated for 30 min with tyramide-coupled Alexa Fluor-488. Next, H₂O₂ was added to a final concentration of 0.012% over another 30 min. Embryos were washed with PBS-T and stored at 4 °C until further use.

Hybridization of Hi-M primary library. Hybridization followed a previously described protocol⁴⁸. Briefly, embryos were RNase treated for 2 h, permeabilized for 1 h with 0.5% Triton X-100 in PBS and rinsed with sequential dilutions of Triton X-100/pH buffer to 100% pH (2× SSC, 0.1 M NaH₂PO₄, pH 7, 0.1% Tween-20 and 50% formamide (v:v)). Embryos in pH buffer were preheated at 80 °C, the supernatant was aspirated and 30 µl of FHB (50% formamide, 10% dextran sulfate, 2× SSC and 0.5 mg ml⁻¹ of salmon sperm DNA) containing 225 pmol of the primary library was pipetted directly onto the embryos. Mineral oil was added on top and the tube was incubated overnight at 37 °C. The next day, oil was carefully removed and embryos were washed twice for 20 min at 37 °C with 50% formamide, 2× SSC and 0.3% 3-(3-cholamidopropyl)dimethylammonio-1-propanesulfonate (CHAPS). Next, embryos were sequentially washed for 20 min at 37 °C with serial dilutions of formamide/PBS-T to 100% PBS-T. Embryos were rinsed with PBS-T and stored in PBS-T at 4 °C until the imaging step.

Imaging system. All experiments were performed on a homemade, wide-field epifluorescence microscope built on a RAMM modular microscope system (Applied Scientific Instrumentation) coupled to a microfluidic device, as described previously^{47,48}. Samples were imaged using an ×60 Plan-Achromat water-immersion objective (numerical aperture = 1.2; Nikon). The objective lens was mounted on a closed-loop piezoelectric stage (Nano-F100, Mad City Labs Inc.). Illumination was provided by four lasers (OBIS-405/488/640 nm and

Sapphire-LP-561 nm, Coherent). Images were acquired using an sCMOS camera (ORCA Flash 4.0V3), with a final pixel size calibrated to 106 nm. A homemade autofocus system was used to correct for axial drift in real time and to maintain the sample in focus as previously described⁴⁷.

A fluidic system was used for automated sequential hybridizations, by computer-controlling a combination of three eight-way valves (HVXM 8-5, Hamilton) and a negative pressure pump (MFCS-EZ, Fluigent) to deliver buffers and secondary readout probes on to a FCS2 flow chamber (Biotechs). Software-controlled microscope components, including camera, stages, lasers, pump and valves, were run using a homemade software package developed in LabView 2015 (National Instruments).

Acquisition of Hi-M datasets. Embryos were attached to a poly(L-lysine)-coated coverslip and mounted into the FCS2 flow chamber. Fiducial readout probe (25 nM rhodamine-labeled probe, 2× SSC, 40% v:v formamide) was flowed on to the sample and hybridized for 15 min, washed for 10 min with readout washing buffer (2× SSC, 40% v:v formamide) and for 5 min with 2× SSC before injecting 0.5 mg ml⁻¹ of DAPI in PBS to stain nuclei. The imaging buffer (1x PBS, 5% w:v glucose, 0.5 mg ml⁻¹ of glucose oxidase and 0.05 mg ml⁻¹ of catalase) was injected. Subsequently, 10–15 embryos were selected according to developmental stage and orientation, and segmented into a mosaic of multiple fields of view (200 × 200 μm²). After brightfield image recording, z-stacks were taken with 405-, 488- and 561-nm laser illuminations. The z-stacks had a step size of 250 nm with a total range of 15 μm.

Next, the sample was sequentially hybridized with different secondary readout probes, imaged in the rhodamine and the Alexa Fluor-647 channels, and photobleached. For each round of secondary hybridization, the sample was treated with secondary hybridization buffer (25–50 nM imaging oligonucleotide, 2× SSC, 40% v:v formamide, which also included 50 nM of adapter oligonucleotide in the case of the high-res libraries; see Hi-M libraries) for 15 min, then washed with readout washing buffer and with 2× SSC before injecting imaging buffer. After imaging, the fluorescence of the readout probes was extinguished using a chemical bleaching buffer (2× SCC, 50 mM TCEP hydrochloride) for 10 min, and then the sample was washed with 2× SSC for 5 min before a new hybridization cycle started. All buffers were freshly prepared and filtered for each experiment. The imaging buffer used for a single experiment was stored under a layer of mineral oil and renewed every 12–15 h. Further details can be found on our previously published protocol⁴⁸.

Image processing. Our homemade Hi-M microscope produced z-stacks in DCIMG format, which were converted to TIFF using proprietary software from Hamamatsu. TIFF images were then deconvolved using Huygens Professional v.20.04 (Scientific Volume Imaging, <https://svi.nl>). Further analysis steps were performed using a homemade analysis software that implemented the steps described previously⁴⁸. Briefly, images were first z-projected using either sum (DAPI channel) or maximum intensity projection (barcodes, fiducials) and image-based cross-correlation was used to align the fiducial channels. These corrections were then used to align DAPI and barcode images. Next, the positions of the centers of barcodes were detected with subpixel resolution using local maximum-fitting functions from the ASTROPY package⁴³. Nuclei were segmented from projected DAPI images by adaptive local thresholding and watershed filtering⁴⁸. RNA images were segmented by manually drawing polygons over the nuclei displaying a pattern of active transcription. Barcodes and RNA status were then attributed to each single nucleus by using the XY coordinates of the barcodes, the projected DAPI masks of nuclei and the transcriptional status from manual masking. Finally, pairwise distance matrices were calculated for each single nucleus. From the list of pairwise distances obtained from any two barcodes, we calculated the contact probability as the number of nuclei in which pairwise distances were within 250 nm normalized by the number of nuclei containing both barcodes. This definition was similar to that used in other studies^{77,78}, and avoided biases due to uneven barcode detections. Contact frequencies obtained using this pipeline and those using previous pipelines⁴⁸ produced highly correlated results. Image processing was carried out from Linux terminals connected to a server running Linux PopOS 19.10, with four GeForce GTX 1080Ti GPU cards (SCAN computers). Assessment of Hi-M dataset size was done using a bootstrapping approach (Extended Data Fig. 1h).

4M profiles and multi-way interactions. The 4M profiles were obtained by slicing the corresponding Hi-M contact map across a given anchor. Multi-way interactions were obtained by selecting an anchoring barcode and calculating the single-nucleus pairwise distances to all possible pairs of barcodes. If both barcode-anchor distances for a given barcode pair in a single nucleus are below the contact threshold (250 nm), this nucleus is considered to have a three-way interaction for this anchor and barcode pair. The three-way contact frequency is then obtained by dividing the number of nuclei that show a three-way interaction by the number of nuclei where the three barcodes involved in the three-way interaction have been detected. The calculation of three-way interactions does not require the detection of contiguous barcodes, and therefore the calculation of three-way frequencies is not restricted to nuclei displaying all the barcodes.

ShRec3D structures. The 3D topological representations were obtained from Hi-M pairwise distance maps using our own Python implementation of the approaches described by Lesne et al. and Morlot et al. for ShRec3D^{62,94}. Starting from the single-cell pairwise distance matrix, an ensemble pairwise distance matrix was calculated using the first maximum of the kernel density estimation. These pairwise distances were converted into 3D coordinates for each barcode using nonclassical metric multidimensional scaling. When necessary, structures were mirrored and a ball-and-stick representation was rendered with PyMOL (PyMOL Molecular Graphics System, v.2.3 Schrödinger, LLC.).

Genome-wide analysis of Zld-mediated interactions. First, we extracted lists of Zld peaks genome wide. For the Zld peaks used in Fig. 4k–o and Extended Data Fig. 7i–k: datasets from Harrison et al.⁵⁴ (accession no. GSM763062) were used to extract the autosomal regions bound by Zld at 3 hours post fertilization using their corresponding chromatin immunoprecipitation (ChIP)-sequencing intensity⁵⁴. For the Zld peaks used in Extended Data Fig. 7l: a list of putative enhancers of pre-MBT genes ($N=62$) was obtained by selecting the Zld peaks nearest to the transcription start site of pre-MBT genes⁹⁵ (Supplementary Table 8). BED coordinates were remapped from dm3 (BDGP R5) to dm6 (BDGP R6) using FlyBase's sequence coordinates converter (FB2020_05, released 14 October 2020).

Second, we characterized interactions between Zld-bound regions using the 5-kb resolution Hi-C dataset from Hug et al.⁴⁰. Peaks were sorted and classified into different categories based on protein occupancy. If multiple Zld peaks were contained within the same 5-kb bin, only the one with the highest intensity was considered. After filtering, we analyzed 5,038 bins occupied by Zld in the different autosomal chromosomes. For each biological condition, intra-arm chromosomal contacts were distance normalized by computing the log₂(observed/expected). The average interaction frequencies at long (>250 kb) or short (<250 kb) ranges were then ranked in five groups by increasing the Zld ChIP signal. From low to high Zld peak intensity, each group contains, respectively, 3,124, 706, 480, 248 and 480 peaks.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The Oligopaint public database (<http://genetics.med.harvard.edu>) was used to select Oligopaints. Publicly available datasets used in the present study (accession nos. GSE86966, GSE25180, E-MTAB-4918, GSM763062, GSE58935, GSE16245, GSE68983, GSE68654, E-MTAB-1673, GSE62904 and GSE65441) are detailed in Supplementary Table 9. Data for matrices in Figs. 1–4 and in Extended Data Figs. are publicly available at <https://github.com/NollmannLab/Espinola-Goetz-2021>. Source data are provided with this paper.

Code availability

Code used in this manuscript is available at <https://github.com/NollmannLab/Espinola-Goetz-2021>.

References

92. Kahn, T. G. et al. Interdependence of PRC1 and PRC2 for recruitment to polycomb response elements. *Nucleic Acids Res.* **44**, 10132–10149 (2016).
93. The Astropy Collaboration et al. The Astropy Project: building an open-science project and status of the v2.0 core package. *Astronom. J.* **156**, 123–1412 (2018).
94. Morlot, J.-B., Mozziconacci, J. & Lesne, A. Network concepts for analyzing 3D genome structure from chromosomal contact maps. *EPJ Nonlinear Biomed. Physics* **4**, 2 (2016).
95. Chen, K. et al. A global change in RNA polymerase II pausing during the *Drosophila* midblastula transition. *eLife* **2**, e00861 (2013).
96. Boettiger, A. N. et al. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016).

Acknowledgements

We thank N. Benabdallah, G. Cavalli, T. Forne, T. Robert, J. Bonnet and members of the Lagha and Nollmann laboratories for their critical reading of the manuscript. We thank A. Makrini and D. Cattoni for help with bioinformatic analysis. This project was funded by an ERC Consolidator Grant from the European Union's Horizon 2020 Research and Innovation Program (grant no. 724429 to M.N.). We thank the Bettencourt-Schueler Foundation for their prize Coup d'élán pour la recherche Française, the France-BioImaging infrastructure supported by the French National Research Agency (grant no. ANR-10-INBS-04, Investments for the Future), the Labex EpiGenMed (ANR-10-LABX-12-01) and the *Drosophila* facility (BioCampus Montpellier, CNRS, INSERM, Université de Montpellier, France). M.G. was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation; project no. 431471305). M.L.'s laboratory is supported by an ERC Starting Grant (SyncDev, grant no. 679792) and CNRS. M.B. and O.M. are supported by an FRM PhD fellowship.

Author contributions

A.M.C.G., M.L. and M.N. conceived the study and the design. S.M.E., C.H., M.B. and I.R. acquired the data. M.G., S.M.E., M.B., O.M., I.R. and M.N. analyzed the data. M.G., M.N., J.B.F. and O.M. provided the software. S.M.E., M.G., O.M., I.R., M.B., M.N. and M.L. interpreted the data. M.L. and M.N. wrote the manuscript. J.-B.F., C.H. and I.R. provided the reagents. S.M.E., M.G. and M.N. did the visualization of the study. M.N. and M.L. supervised the study. M.N. and M.L. acquired funds.

Competing interests

The authors declare no competing interests.

Additional information

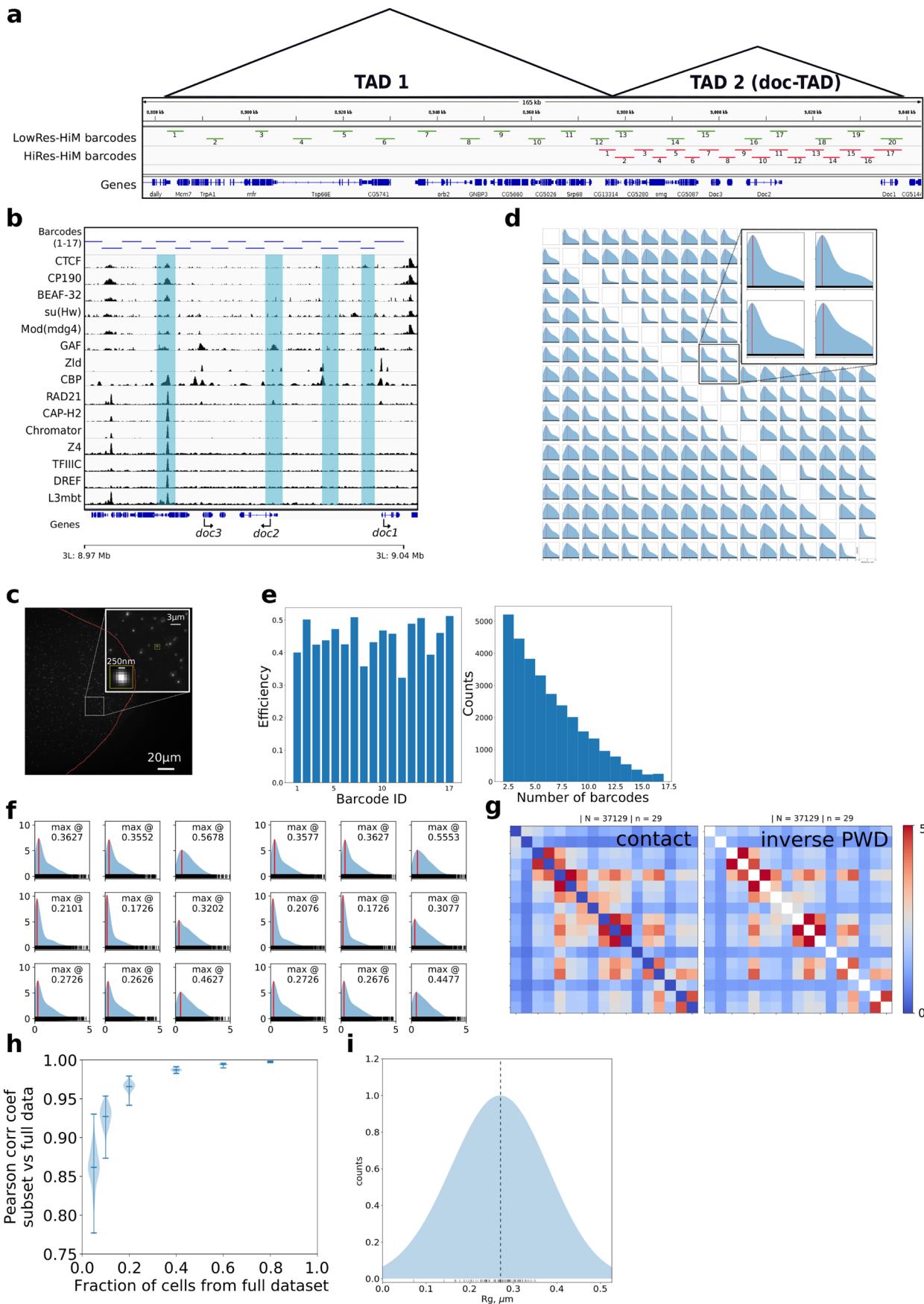
Extended data is available for this paper at <https://doi.org/10.1038/s41588-021-00816-z>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41588-021-00816-z>.

Correspondence and requests for materials should be addressed to M.L. or M.N.

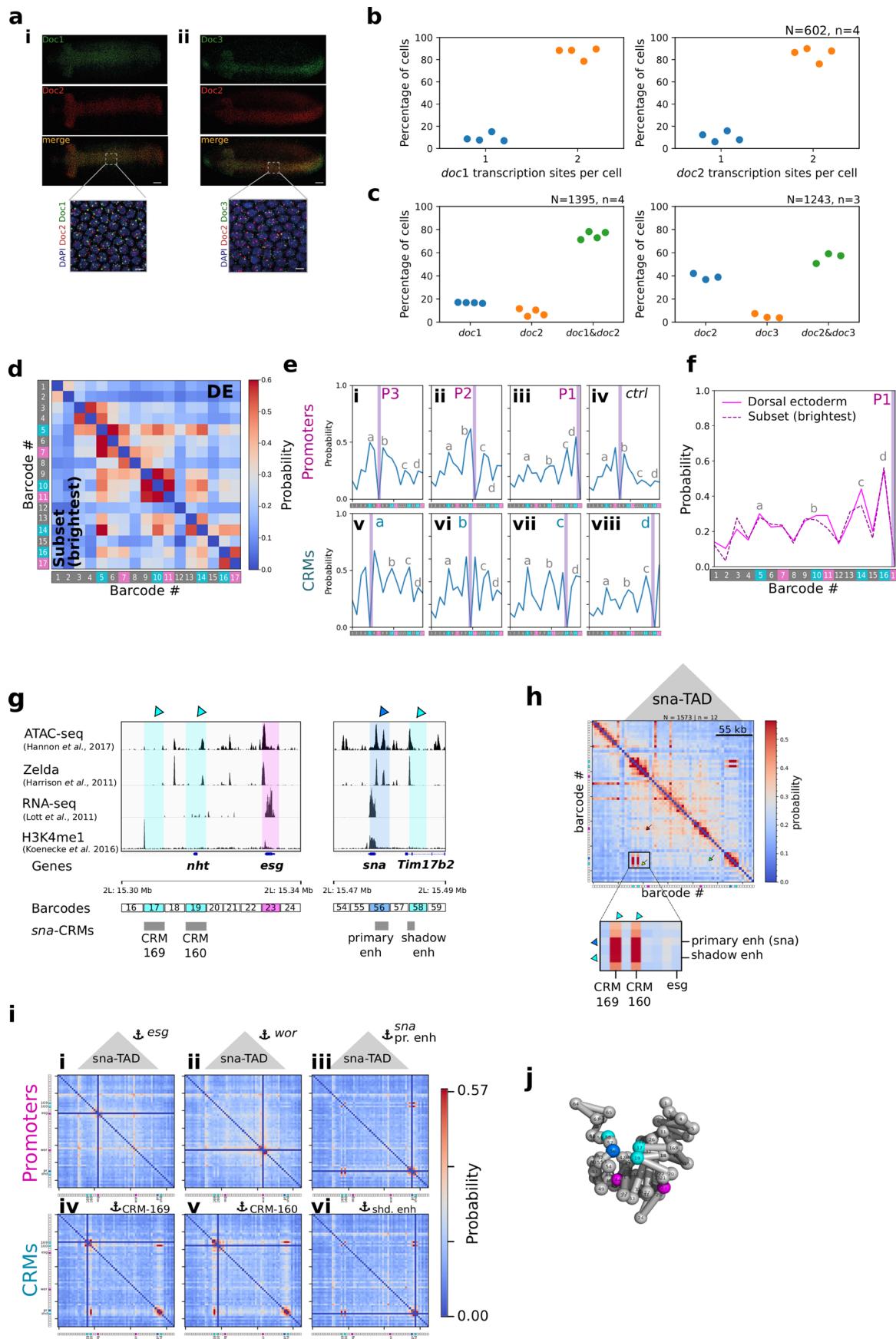
Peer review information *Nature Genetics* thanks Justin Crocker and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.



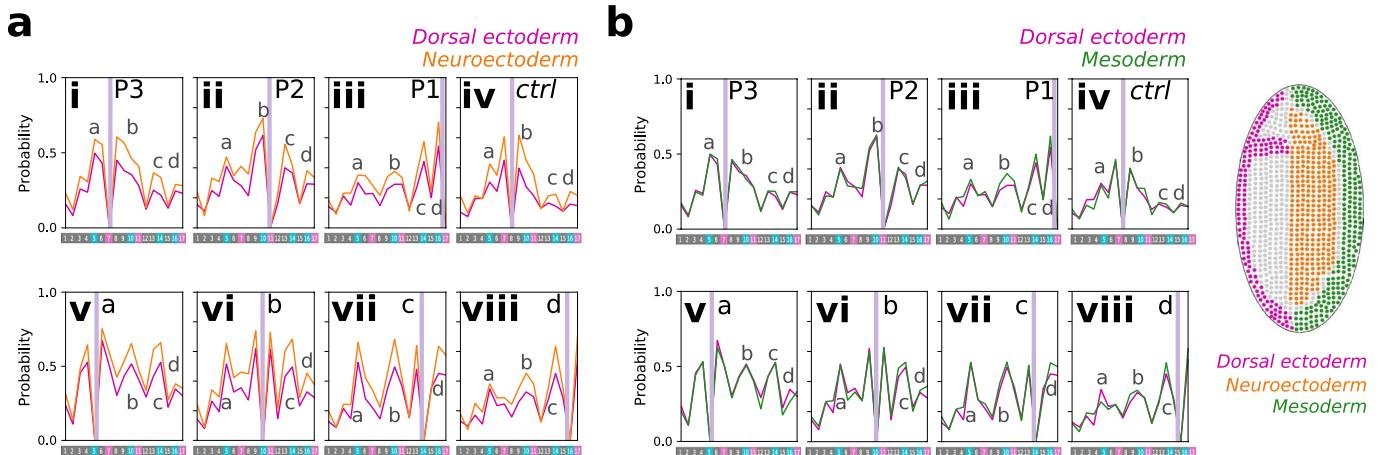
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Hi-M allows high-resolution chromatin tracing in the doc-TAD. **a**, Schematic representation of the genomic positions of barcodes for the low (green) and high (red) resolution doc Hi-M libraries. Triangles demarcate the two TADs registered in this genomic region⁹⁰. **b**, Chip-seq profiles for architectural proteins in the doc-TAD^{54,90}. Cis-regulatory modules (CRM_{a-d}) from Fig. 1c are highlighted by blue bars. **c**, Typical maximum intensity projection displaying the fluorescence emission signal from a single barcode in a section of an embryo (outline in red). Emissions from individual barcodes appear as diffraction-limited spots. **d**, Map of pairwise distance distributions for all barcode combinations. The order of the distributions follow that in the Hi-M matrix (Fig. 1f). Blue shade represents a kernel density estimation with a bandwidth of $0.2\text{ }\mu\text{m}$, red line represents the maximum of the distribution, and black vertical lines on the x-axis represent individual data points. **e**, Efficiency of barcode detection and distribution of number of barcodes detected per cell. **f**, To verify that uneven barcode efficiencies did not affect our results, we plotted the pairwise distance distributions for the full dataset (right) and half the data (left, here cells were randomly chosen). Map of pairwise distributions is centered at the barcode bin (4,13). **g**, Hi-M contact probability map (left) and inverse pairwise distance map (right) for the same experiment (doc-TAD, all cells). $N=37129$, $n=29$. **h**, Pearson correlation coefficient of the contact probability of the full doc-TAD Hi-M dataset (nc14, dorsal cells displaying *doc1* expression) against subsets with a fraction of cells. One hundred random subsets were generated for each tested subset size. The central bar indicates the mean and the error bars indicate the extreme values of the distribution. **i**, Distribution of radii of gyration for the doc-TAD calculated from single cells. Blue shade represents a kernel density estimation with a bandwidth of $0.1\text{ }\mu\text{m}$, black vertical lines on the x-axis represent individual data points. Dashed line represents the maximum of the distribution. The size of the doc-TAD, as estimated from its radius of gyration ($0.27 \pm 0.1\text{ }\mu\text{m}$), was comparable with that of TADs of similar genomic sizes⁹⁶.

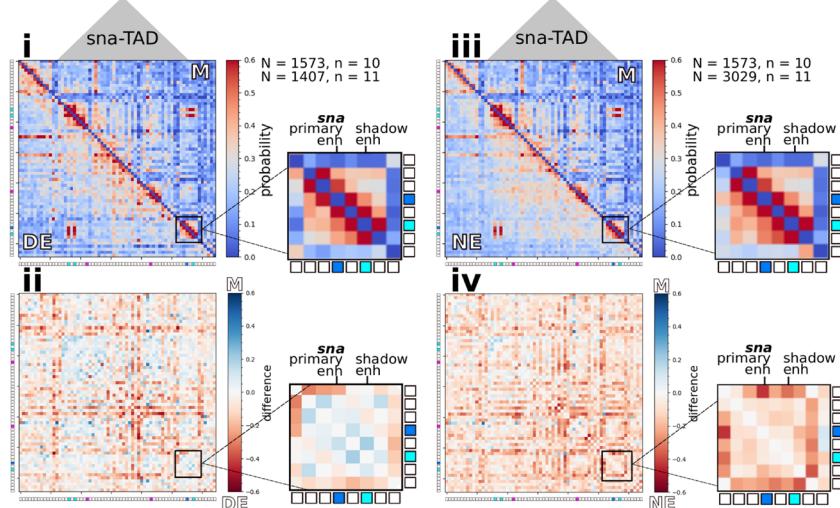
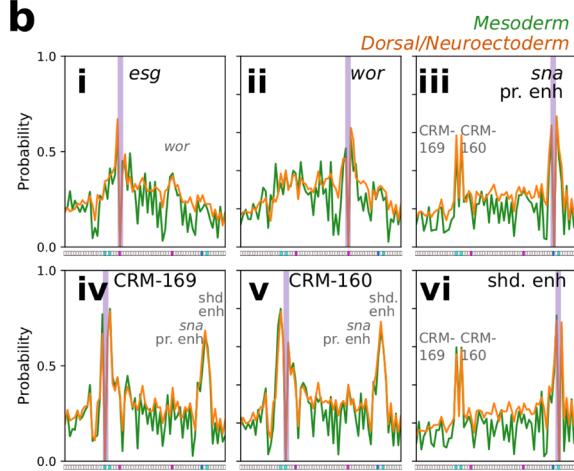
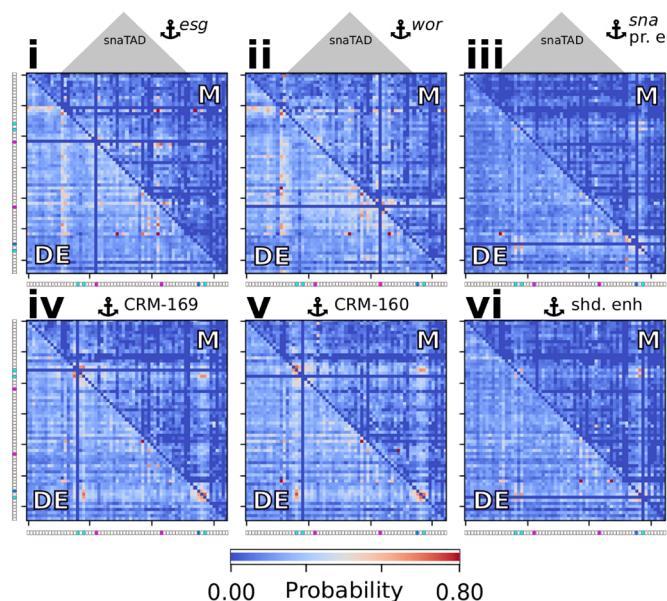
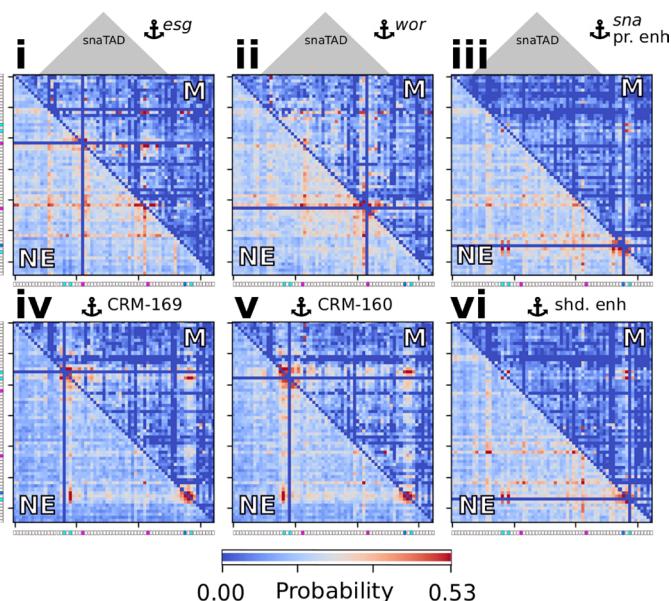


Extended Data Fig. 2 | See next page for caption.

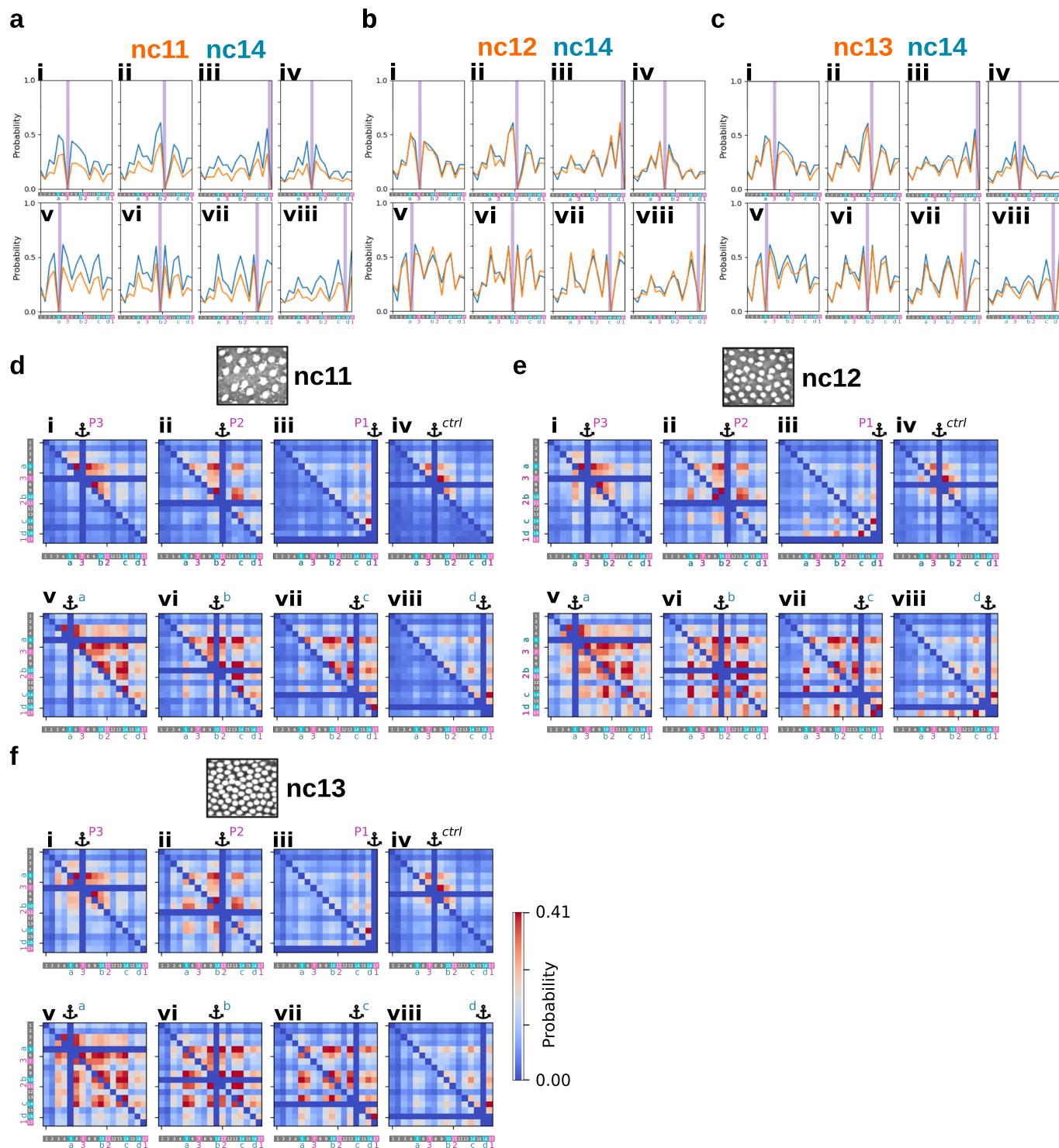
Extended Data Fig. 2 | doc genes are highly co-expressed and doc CRMs spatially cluster, as do CRMs in the sna locus. **a**, RNA-FISH staining for *doc1*, *doc2* and *doc3* in late nc14 embryos. Scale bars: 50 μm / 5 μm (inset). **b**, Percentage of one or two active transcription sites/nuclei for *doc1* and *doc2*. **c**, Percentage of cells displaying active transcription spots from 2-color RNA-FISH imaging of *doc1-doc2* and *doc2-doc3*. Most nuclei (>70%) displayed co-activation of *doc1+doc2* and of *doc2+doc3*. For this latter, a larger percentage of nuclei expressed only *doc2* (~40%), because of the low efficiency of labeling of *doc3* nascent mRNA (small intronic size). **d**, Comparison of contact maps from nuclei displaying at least one active *doc1* RNA-FISH spot (top right matrix) and from a subset (33%) of nuclei displaying the strongest *doc1* RNA-FISH signals. **e**, 4 M profiles derived from Hi-M maps of dorsal ectoderm cells in nc14. **f**, 4 M virtual profile for nuclei displaying at least one active *doc1* RNA-FISH spot (solid pink line) and from a subset of nuclei (33%) with the highest *doc1* signals (dashed dark pink). **g**, Epigenetic profile of selected regions around the *esg* and *sna* genes within the *sna*-TAD. Accessibility (ATAC-seq), pioneer factor binding (Zelda), transcriptional activity (RNA-seq), chromatin marks for active enhancers marks (H3K4me1), and for the transcriptional activators Dorsal, Zen and Mad are shown. A subset of barcodes were annotated as cis-regulatory modules (shown in cyan): CRM₁₆₉ harbors the canonical H3K4me1 active enhancer mark; CRM₁₆₀ and shadow *sna* enhancer were described in the RedFly database. Magenta barcode harbours the *esg* promoter and the blue barcode contains the *sna* promoter and its primary enhancer. See Supplementary Table 1 for more details. **h**, Hi-M contact probability map of the *sna* locus. Yellow arrow shows interactions between CRMs, red arrow between CRMs and promoters, and green arrow between promoters. **i**, Multi-way interactions between promoters (panels i-iii) and CRMs (panels iv-vi). Number of nuclei and embryos examined as in panel r. **j**, 3D topological representation of the *sna*-TAD. Bead colors are as in panel d. Barcode 44 contains the *wor* promoter.



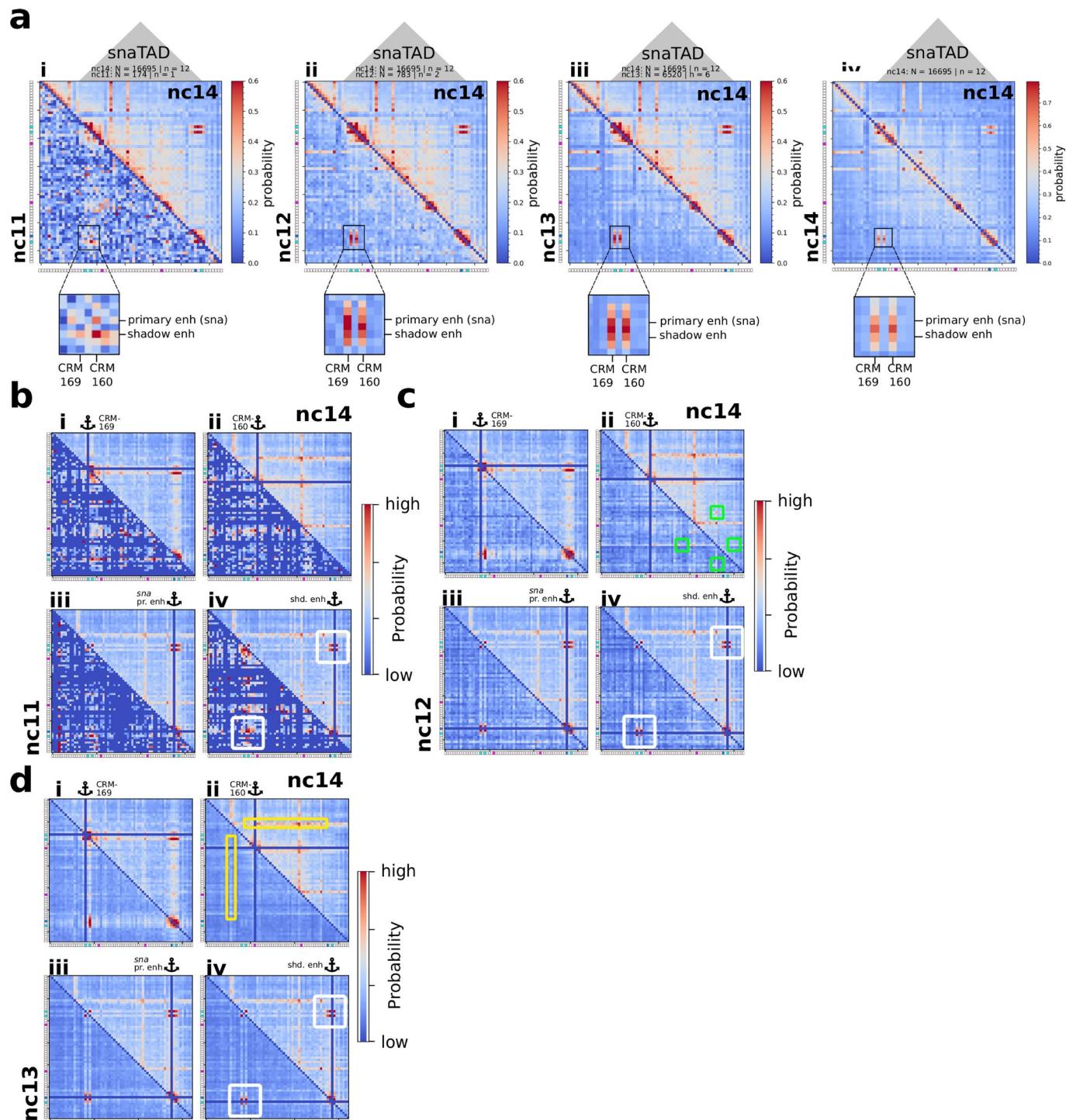
Extended Data Fig. 3 | doc CRM loops are similar between three presumptive tissues. **a**, Comparison of 4 M profiles between DE (magenta) and NE (orange) for different anchors within the doc-TAD (Panels i-iii: promoters. Panel iv: control. Panels v-viii: CRMs). Anchors are indicated by vertical purple lines. Peaks in the profiles are annotated with the corresponding CRMs (a-d) **b**, Comparison of 4 M profiles between DE (magenta) and M (green) for different anchors within the doc-TAD (Panels i-iii: promoters. Panel iv: control. Panels v-viii: CRMs). Anchors are indicated by vertical purple lines. Peaks in the profiles are annotated with the corresponding CRMs (a-d). Right panel: scheme indicating the three presumptive tissues.

a**b****c****d****Extended Data Fig. 4 | See next page for caption.**

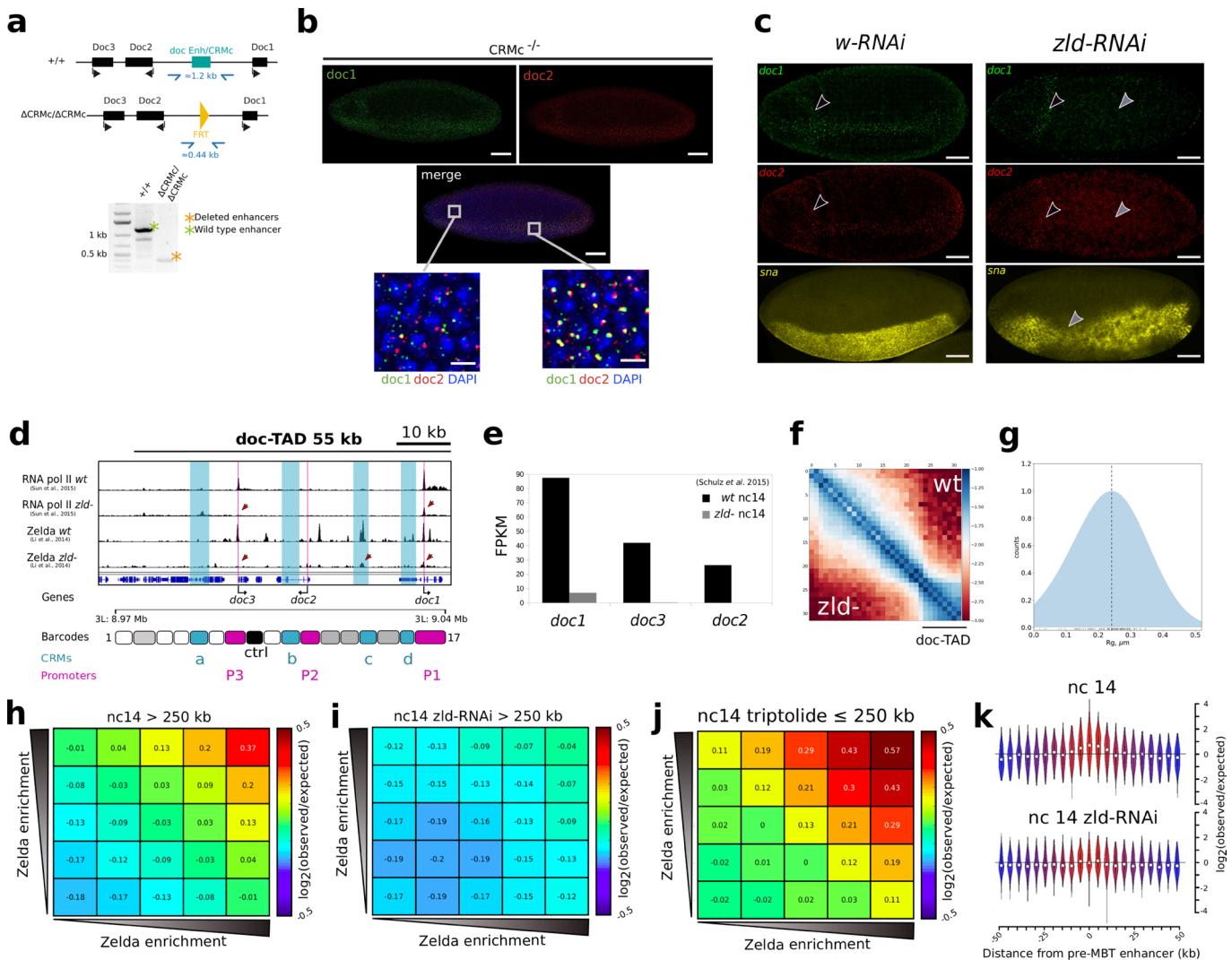
Extended Data Fig. 4 | *sna* CRM loops are similar between three presumptive tissues. **a**, Panel i: Hi-M contact probability map of the *sna* locus for M (upper-right map) versus DE (lower-left map). Inset show a magnification of the region around *sna*. Panel ii: Same but for the difference between M and DE Hi-M maps. Blue indicates larger contact probabilities in M whereas red indicates larger contact probabilities in DE. Panel iii: Similar to panel i, but for M (upper-right map) versus NE (lower-left map). Panel iv: Similar to panel ii, but for M versus NE. N: number of nuclei. n: number of embryos. **b**, Comparison of 4 M profiles between M (green) and DE/NE (orange). Anchors within the *sna*-TAD are indicated in each panel by vertical purple lines. A subset of peaks is annotated using the nomenclature from Fig. 1d. **c**, Comparisons of multi-way maps for M (upper-right map) versus DE (lower-left map) in the *sna* locus using the anchors indicated in each panel by pictograms and dark blue crosses. Maps are color-coded according to the scale bar on the right. Number of embryos and nuclei as in panel c. **d**, Similar to panel e, but for M (upper-right map) versus NE (lower-left map). Number of embryos and nuclei as in panel c.



Extended Data Fig. 5 | doc CRM loops are established early in development. **a**, Comparison of 4M profiles between embryos in nc14 (blue lines) and nc11 (orange) for different anchors within the doc-TAD (Panels i-iii: promoters. Panel iv: control. Panels v-viii: CRMs). The position of the anchor is indicated by a vertical purple line. Peaks in the profiles are annotated with the corresponding CRMs (a-d). **b**, Similar to panel A, but comparing 4M profiles between embryos in nc14 (blue lines) and nc12 (orange). **c**, Similar to panel A, but comparing 4M profiles between embryos in nc14 (blue lines) and nc13 (orange). **d**, Comparison of multi-way interaction matrices of nc14 (upper-right map) and nc11 (lower-left map). Anchors (dark blue crosses) are as follows: *doc3*, *doc3*, *doc1* promoters (panels i-iii), control region (panel iv), CRM_{a-d} (panels v to viii). Representative image of DAPI-stained nuclei for nc11 is shown on top. Barcodes are shown on the left and bottom of multi-way maps. Number of nuclei (nc11): N=1320, number of embryos (nc11); n=4. Number of nuclei (nc14): N=37129, number of embryos (nc14); n=29. **e**, Similar to panel d, but for nc14 (upper-right map) and nc12 (lower-left map). Representative image of DAPI-stained nuclei for nc12 is shown on top. Number of nuclei (nc12): N=2154, number of embryos (nc12); n=4. **f**, Similar to panel d, but for nc14 (upper-right map) and nc13 (lower-left map). Representative image of DAPI-stained nuclei for nc13 is shown on top. Number of nuclei (nc13): N=7597, number of embryos (nc13); n=8.



Extended Data Fig. 6 | sna CRM loops are established early in development. **a**, Comparison of Hi-M contact probability maps in the *sna* locus for nc14 (upper-right map) and nc11 (panel i), nc12 (panel ii), nc13 (panel iii) and 14 (panel iv) (lower-left maps). Maps are color-coded according to the scale bar on the right. Inset on the bottom of each map shows a magnification of the region around *esg* and *sna* CRMs (see Extended Data Fig. 2 and Supplementary Table 1). **b**, Comparison of multi-way contact maps between nc14 (upper-right maps) and nc11 (lower-left maps). Maps are color-coded according to the scale bar on the right. The position of anchors are indicated by dark blue crosses. White boxes indicate contacts already present at nc11 that persist through nc14. Number of nuclei and embryos examined as indicated in panel a. **c**, Similar to panel b, but comparing nc14 to nc12. Green boxes indicate contacts that emerge at nc12 and persist at nc14. Number of nuclei and embryos examined as indicated in panel a. **d**, Similar to panel b, but comparing nc14 to nc13. Yellow boxes indicate interactions that appear at nc13 (at the TAD border). Number of nuclei and embryos examined as indicated in panel a.



Extended Data Fig. 7 | Perturbation of gene expression and CRM loops by enhancer deletion and Zld depletion. **a**, Scheme of the wild type *doc* locus (+/+) and the *doc* locus after CRISPR/Cas9 genome editing ($\Delta\text{CRM}_c/\Delta\text{CRM}_c$). *doc* enhancer/CRM_c, FRT sequence and primers used for genotyping are in teal, yellow and blue, respectively. Genotyping PCR products on agarose gel electrophoresis are shown in the lower panel. Orange and green stars correspond to the bands of the expected sizes after amplification using primers flanking the *doc* enhancer/CRM_c sequence. See Methods for further details. **b**, RNA-FISH imaging of *doc1* and *doc2* in the CRM_c-deletion mutant. Scale bars: 50 μm / 5 μm (inset). **c**, RNA-FISH imaging of *doc1*, *doc2* and *sna* in control (RNAi white) and RNAi *Zld* embryos. Black arrows show the *doc1* and *doc2* expression patterns in the anterior part of the embryo. Grey arrows indicate the absence (*doc1*, *doc2*) or perturbation (*sna*) of gene expression patterns in RNAi-*Zld* embryos. Scale bar: 50 μm . **d**, Tracks for pioneer factor binding (Zelda) and RNA Pol2 binding in the doc-TAD. See Supplementary Table 1 for assignment of CRM_{b-d}. **e**, Transcription levels (RNAseq) of *doc1*, *doc2* and *doc3* in wild-type versus *zld-* embryos. **f**, Hi-C matrix for a genomic region containing doc-TAD in wild-type and Zld-depleted embryos. Data from Hug et al. (2017)⁴⁰. **g**, Distribution of radius of gyration for the doc-TAD in Zld-depleted embryos (see Extended Fig. 1i for wild-type). **h-j**, Log2(observed/expected) average contact frequencies between Zelda bound regions at long-range distances (> 250 kb) ranked by increasing Zelda enrichment in nc14 (panel g), nc14 zld-RNAi (panel h) and at short-range (≤ 250 kb) in nc14 triptolide-treated embryos (panel i). **k**, Violin plot of intragroup Log2(observed/expected) distribution between 62 selected pre-MBT enhancers and neighbouring sequences (± 5 kb) in nc14 (upper panel) and nc14 zld-RNAi (lower panel). The central white marker indicates the median and the vertical black lines indicate the extreme values of the distribution. The coordinates of enhancers and closest pre-MBT genes are listed in Supplementary Table 8.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	custom made software package developed in Labview 2015 (https://github.com/NollmannLab/HiMacquisitionSoft); ZEB 2.3 SP1 (Zeiss LSM880 confocal), Leica Application Suite Advanced Fluorescence 2.3.1 (Leica SP5 II confocal)
Data analysis	image processing and analysis: MATLAB R2019-2020, DCIMG to TIFF conversion software (hamamatsu proprietary MATLAB library), Huygens Professional version 20.04, pyHIM 0.4 (https://github.com/NollmannLab/Espinola-Goetz-2021), ATROPY package v4.0.1 post1, PyMol v2.3, ImageJ v1.53c, Genomic data processing and visualization: FlyBase's sequence coordinates converter version FB2020_05, integrative genomics viewer v2.8.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Oligopaint public database (<http://genetics.med.harvard.edu>) was used to select oligopaints. Publicly available datasets used in this study (GSE86966, GSE25180, E-MTAB-4918, GSM763062, GSE58935, GSE16245, GSE68983, GSE6854, E-MTAB-1673, GSE62904, GSE65441) are detailed in Supplementary Table 9. Data used to build figures 104 in this study are available at <https://github.com/NollmannLab/Espinola-Goetz-2021>. Uncropped version of agarose gel in Extended Fig. 7a is provided in Source Data Extended Data Fig. 7

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No a priori sample size calculation was performed. The nature of the experiment allows for the acquisition of thousands of nuclei across 5-10 embryos per experiment. Pearson correlations were used to verify that even a small proportion of the dataset (~10%) correlated highly with the whole dataset. See Figure legends for more information on sample sizes.
Data exclusions	no data were excluded from the analysis
Replication	We calculated person correlations between datasets, and verified that matrices of different embryos displayed the same patterns.
Randomization	Individual embryos were randomly chosen on coverslips during image acquisition.
Blinding	No blinding was necessary for this study, biological samples and groups were defined before measurements. Individual embryos were chosen randomly for acquisition. Images were collected as a pool for the different conditions and all images were processed. Analysis were performed using a computational pipeline in an automated and identical manner between different samples.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern		

Antibodies

Antibodies used	sheep anti-DIG, Roche cat #11333089001, mouse anti-Biotin, Life technologies cat #03-3700, anti-digoxigenin-POD Fab fragments Sigma-Aldrich cat #11207733910; anti-mouse alexa488-conjugated, life technologies cat #A21202; Anti-sheep alexa555-conjugated, Life technologies cat #A21436
Validation	All antibodies used in this study are commercially available and validated by manufacturers. Furthermore, they were validated before in Cardozo et al Mol. Cell 2019, and Dufourt et al (2018). See citations in manuscript.

Animals and other organisms

Policy information about [studies involving animals; ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Drosophila melanogaster wild type embryos (yw). For zelda depletion experiments, females from the cross between nos-Gal4:VP16 (BL4937) and UASp-shRNA-zld (see Ref. 67). For CRISPR-cas9 line with CRM _c deletion: dsRED-/- males were crossed with Dp/TM3,Sb virgins females. Males from this cross were then crossed with hs-FLP/hs-FLP;Dr/TM3,Sb virgins females. Embryos were from developmental stage nuclear cycle 11 to 14.
Wild animals	No wild animals used
Field-collected samples	No samples collected in the field

Ethics oversight

this study did not require ethical approval

Note that full information on the approval of the study protocol must also be provided in the manuscript.