Published in final edited form as:

Nature. 2019 April ; 568(7750): 49–54. doi:10.1038/s41586-019-1035-4.

Visualizing DNA folding and RNA in embryos at single-cell resolution

Leslie J. Mateo[1], **Sedona E. Murphy**[1,2], **Antonina Hafner**[1], **Isaac S. Cinquini**[1,3], **Carly A. Walker**[1], and **Alistair N. Boettiger**[1,*]

[1]Department of Developmental Biology, Stanford University

[2]Department of Genetics, Stanford University

[3]Department of Computer Science, Stanford University

Abstract

Establishment of cell types during development requires precise interactions between genes and distal regulatory sequences. Our understanding of how these interactions look in three dimensions, vary across cell types in complex tissue, and relate to transcription remains limited. Here we describe optical reconstruction of chromatin architecture (ORCA), a method to trace the DNA path in single cells with nanoscale accuracy and genomic resolution reaching 2 kilobases. We applied ORCA to a Hox gene cluster in cryosectioned *Drosophila* embryos and labelled ~30 RNA species in parallel. We identified cell-type-specific physical borders between active and Polycomb-repressed DNA, and unexpected Polycomb-independent borders. Deletion of Polycomb-independent borders led to ectopic enhancer-promoter contacts, aberrant gene expression, and developmental defects. Together, these results illustrate an approach for high-resolution, single-cell DNA domain analysis *in vivo*, reveal domain structures that change with cell identity, and show that border elements contribute to formation of physical domains in *Drosophila*.

In embryogenesis, developmental patterning is controlled by a relatively small number of genes, each regulated by multiple enhancers distributed tens to hundreds of kilobases (kb) away. The ability of enhancers to selectively activate target genes and avoid non-target genes is attributed in part to the three-dimensional (3D) folding of the genome. Consequently, examining the sub-megabase-scale folding of developmental control loci and the extent to which it varies between cell types is essential for understanding embryogenesis[1-3].

*Correspondence and requests for materials should be addressed to A.N.B. (boettiger@stanford.edu).

Author Contributions

L.J.M. and A.N.B. designed the experiments. L.J.M. collected the data with assistance from S.E.M., I.S.C., C.A.W. and A.N.B. A.H. collected and analysed the mouse data. L.J.M. S.E.M. A.H. and A.N.B. analysed the data and wrote the manuscript with assistance from I.S.C. and C.A.W.

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

The authors declare no competing financial interest.

Chromatin conformation capture (3C) approaches have mapped *cis*-interactions *de novo* and revealed partitioning of the genome into topological associated domains (TADs), where intra-domain contacts are enriched over inter-domain contacts. TADs frequently span developmental genes and their *cis*-control regions[4,5]. However, these bulk approaches lack the ability to distinguish sub-populations within a heterogeneous sample. Single-cell Hi-C approaches have revealed variability in chromatin organization, but have lower resolution compared to bulk approaches, and do not retain spatial organization[5]. Microscopy approaches can detect sub-populations within tissues *de novo* and resolve large numbers of individual. Accordingly, multicolour fluorescent *in situ* hybridization (FISH)[6-12], oligo-stochastic optical reconstruction microscopy (oligo-STORM)[9,13-17], and sequential FISH[17-19] have revealed cell-type-specific chromatin packaging, compartmentalization, and long-range *cis*-interactions. However, these methods rely on prior data to identify enhancers and have generally not resolved *cis*-interactions within tens of kb, typical of many *cis*-regulatory controls.

We sought a method that combines the *de novo* mapping of TADs and resolution of enhancer-promoter interactions achieved by recent Hi-C with the single-cell resolution and tissue organization provided by FISH. Like Hi-C, the approach should be able to detect regions of the genome which preferentially interact with promoters *de novo.* Like microscopy methods, the approach should be able to detect sub-populations of cells with common properties without requiring cell sorting or dissection. To distinguish cell types within a tissue and relate enhancer-promoter contacts to gene expression, the method should be compatible with simultaneous measurement of mRNAs and nascent transcription in each cell. Finally, to provide a sufficiently sampled view of the tissue, the method should process thousands of individual cells per run. Here, we describe optical reconstruction of chromatin architecture (ORCA), an approach that simultaneously achieves these goals, and apply it to test several predictions about chromatin structure and *cis*-regulation in the Bithorax Complex (BX-C) of *Drosophila* embryos.

Principle of the method

ORCA builds on recent innovations in RNA and DNA FISH, taking advantage of array-derived oligonucleotide (oligo) probes (Oligopaints)[9,13,18,20,21]. ORCA reconstructs the trajectory of a genomic region of interest (100–700 kb), by tiling the region in short sections (2–10 kb) with primary probes that have unique barcodes[20] (Fig. 1a, Extended Data Fig. 1a-c, Supplementary Data Tables 1-5). These barcodes are labelled with a fluorophore and imaged. The signal is then removed via strand displacement (Supplementary Data Table 6). The process repeats for each barcode. This is conceptually similar to recent[18] and concurrent work[17,19], though with improved genomic resolution (Fig. 1b). With high-precision fiducial registration (Extended Data Fig. 1a-c), sequential imaging allows barcoded sections within a diffraction-limited volume to be resolved, as in STORM, while adding sequence resolution across the domain (Fig. 1b). We represent the measured 3D positions of the barcodes as spheres, pseudo-coloured per barcode and linked with a smooth polymer (Fig. 1c), and as distance maps (Fig. 1d).

We applied ORCA to visualize the nanoscale DNA path of the BX-C at 10-kb and 2-kb resolution in 6-μm cryosections of *Drosophila* embryos 10–12 hours post-fertilization (hpf). The 10-kb step size allowed a 700-kb region, including flanking domains of the BX-C, to be traced with 70 barcodes (Fig 1d, e). The 2-kb step size enhanced the resolution over a 130-kb regulatory region spanning *abd-A* to *Abd-B* (Fig.1a, g, h). Missed detection events (Fig. 1d, g, grey lines) were largely stochastic – showing no significant variation between embryos, cell types, or probesets, and limited variation between barcodes (Extended Data Fig. 2). Replicate experiments yielded reproducible measurements (Pearson's $r > 0.95$) (Extended Data Fig. 1d). Comparisons to published Hi-C[22] across a range of contact thresholds revealed qualitatively similar features and quantitatively similar contact distributions (Fig. 1e, f, h, i, Extended Data Fig.1 e, f). In *Drosophila* chromosomes, which are predominantly paired in interphase, our ORCA images reveal that paired homologs follow a common trajectory to within ~50 nm. (Extended Data Fig. 3). Additional ORCA experiments in mouse embryonic stem cells tracing the region containing *Sox2* at 5-kb resolution (Extended Data Fig. 4) exhibited strong correspondence with published Hi-C data[23] (Pearson's $r = 0.96$), illustrating versatility across cell types.

Cell-type-specific chromatin structure

By 10 hpf the *Drosophila* embryo consists of distinct cell types characterized by diverse transcriptional behaviours that are not distinguishable by Hi-C. To characterize these cell types we labelled 29 relevant RNA transcripts, both mRNA and nascent RNA, with unique barcodes (Fig. 2a, b, Extended Data Fig. 5a). Following RNA imaging, we denatured DNA and hybridized ORCA probes, targeting the complete 330-kb BX-C with a resolution of 3 kb per barcode (Fig. 2c). Automated image segmentation of this dataset identified over 20,000 individual cells (Extended Data Fig. 5b). In each cell, we quantified the axial position, the associated counts of mRNAs, the intensity of nascent RNAs, and the 3D DNA structure of the BX-C (Fig. 2).

To conduct an unbiased search for cell-type-specific differences in BX-C spatial organization, we used unsupervised clustering of the RNA expression data to partition single cells into 18 groups (Fig. 2d, Extended Data Fig. 6). Strikingly, several clusters exhibited clear differences from one another in the average spatial organization of the BX-C. As in recent Hi-C analyses[23], we normalized these data for genomic-distance effects, which further accentuated cell-type differences (Fig. 2e-g, Extended Data Fig. 6d).

In anterior cells (e.g. 'brain') we found the BX-C adopted an organization in which long-range interactions were in closer physical proximity than expected and short-range interactions were farther apart than expected (Fig. 2E, Extended Data Fig. 6C, d). The closer-than-expected long-range interactions are consistent with previous reports of Polycomb (Pc) repressed DNA being compact[7,13,14] and the known Pc-repressed status of the BX-C in the anterior cells[24-26]. The farther-than-expected short-range interactions were more surprising, though in agreement with recent conclusions from STORM imaging of Pc-repressed DNA[13]. These data suggest that the 3D organization of Pc-repressed DNA is not only compact, but also organized like a random coil in which the average 3D separation of any two points is weakly dependent on their linear distance. We observed distinct structures

in posterior cells, where one or more BX-C genes were transcribed (Fig. 2f, g, Extended Data Fig. 6c, d). Overall, our combined approach identified *de novo* cell populations that contained substantive differences in BX-C 3D structure.

Cis contacts predict transcription

We asked if any of the promoter contacts with the surrounding sequence were predictive of either the presence (ON) or absence (OFF) of nascent RNA by comparing measured nascent transcription and 3D contacts in each cell (Fig. 3a, b). We hypothesized that cells should exhibit a higher likelihood (quantified by the Odds Ratio) of nascent transcription from a promoter if in close proximity to a genomic region with enhancer function. (Fig. 3a). Indeed, promoter proximity to known enhancers was predictive of nascent transcription (Fig. 3c, coloured circles). However, the correlation was weak. In particular, (1) many active promoters lacked proximity with known[24-27] enhancers, (2) many inactive promoters were in proximity with known enhancers, and (3) known enhancers and regions not previously identified to have enhancer activity were similarly predictive of transcription (Fig. 3c, Extended Data Fig. 7a). These observations in the BX-C are more consistent with a gene activation model where regulatory sequences scan a domain[28,29], rather than a model in which regulatory sequences form stable loops with the promoter[30,31]. Encouragingly, the borders of the domains, which we found to be predictive of transcription, match the regulatory domains of each gene identified by prior enhancer trap assays[25,32] (Fig. 3c, grey boxes).

Interestingly, enhancers of one promoter appeared to silence promoters of other genes (Fig. 3c, red hollow circles). As inter-domain contacts happened more frequently in Pc-compacted domains (Fig. 2f), the correlation of long-range contacts with a silent state is likely an indirect effect of Pc regulation.

Segment-specific TAD boundaries

We next asked how the cell-type-specific structural differences of the BX-C relate to its cell-type-specific epigenetic state. Previous studies have shown that Pc and its associated H3K27me3 epigenetic modification coat the BX-C in a body-segment-specific manner[24,26,33]. We hypothesized that the segment-specific H3K27me3 borders should manifest as segment-specific TAD boundaries, as TAD boundaries often coincide with epigenetic domain borders between active and inactive chromatin[34-36].

To test for segment-specific folding of the BX-C, we manually re-sorted the single cells from Fig. 2 using embryo morphology and the RNA data to annotate body segments (Fig. 4a, see Methods). To emphasize segment-specific structures, we subtracted the average map observed in head cells from that observed in each segment. As expected, in cells from body segments T1 and T2, the BX-C formed a single intermixed region, as observed in head cells (Fig. 4b), with no obvious partitioning (Fig. 4c). In T3, the BX-C split into two TADs, centromeric and telomeric (Fig. 4b, d, single-cell examples in Extended Data Fig. 7b). The telomeric TAD corresponded with the H3K27me3-repressed region previously measured in T3[33] and the centromeric TAD corresponded to the de-repressed region around *Ubx* (Fig.

4d). In A1–A2, as the observed H3K27me3 border[33] shifted right (left column, black arrows), the telomeric TAD boundary in the distance maps followed (right column, black arrowheads) (Fig. 4e-g). In A3–A8, the H3K27me3 border was predicted to continue to shift right[24,26]. Consistent with this hypothesis, we observed that the telomeric TAD continued to shrink further in these segments (Extended Data Fig. 8). These patterns were reproducible between independent experiments (Extended Data Fig. 9).

The segment-specific retraction of the telomeric TAD (Extended Data Fig. 8) suggests a structural basis for the classic "open-for-business" model of BX-C regulation derived from genetic experiments[26,37]. In this model, for each segment posterior of T2, the BX-C "opens" (in a regulatory sense) in a centromeric-to-telomeric manner, one "genetic domain" per body segment until the whole complex is "open" in A9[26]. Each of these 10 genetic domains (labelled in the bottom of Fig. 4c-h) contains enhancers critical to specifying the fate of the segment in which they first become "open"[24-26]. It was recently shown that H3K27me3 covers the "closed" genetic domains and retracts by one genetic domain per segment, at least in the tested T2–A2 segments, providing some epigenetic data to support this model[33]. ORCA data confirm that this epigenetic partitioning is matched by spatial partitioning and extend the observations into more posterior segments than were previously accessible[33] (Fig. 4c-g). The agreement between the 3D boundaries measured by ORCA, the epigenetic domains measured by ChIP-seq, and the genetic domains measured in classic experiments, further validates our approach.

Pc-independent TAD boundaries in BX-C

Contrary to earlier predictions[38], we found that not all TAD boundaries in the BX-C coincided with epigenetic domain borders, and therefore not all can be explained by Pc-activity (Fig. 4c-h, right column, green arrowheads). In A1, the de-repressed portion of the BX-C was split into two TADs (Fig. 4e, dotted black lines, Extended Data Fig. 7b). In A2, an additional TAD boundary formed in the H3K27me3-free region (Fig. 4f, green arrowhead). In A9, where no H3K27me3 boundaries were predicted[24,26], we observed two unexpected structural regions. The *abd-A* and *Ubx* genes and their regulatory regions fused in a centromeric TAD, while the regulatory region of *Abd-B* formed a distinct and decompact structure (Fig. 4h, Extended Data Fig. 7b, Fig. 8). This centromeric TAD in A9 (Fig. 4h) is reminiscent of the Pc-dependent compaction in anterior segments, although Pc is not believed to play a role in BX-C silencing in A9[24-26].

Together, these unexpected TAD boundaries offer an explanation for segment-specific enhancer activity (Fig. 4c-h, left column, green/red boxes) and gene expression (Fig. 4c-h, right column, nascent RNA). For example, in A2, the TAD over genetic domain 2 links the *Ubx* enhancers with the *Ubx* promoter and a separate TAD over genetic domain 3 links the *abd-A* enhancers with the *abd-A* promoter. In A3, the TAD over *abd-A* expanded to include genetic domain 4, coinciding with activation of additional *abd-A* enhancers. Thus, Pc-independent TAD boundaries uncovered by ORCA reflect the promoter specificity of enhancers in several of the BX-C genetic domains.

Boundary elements separate active TADs

Since epigenetic domain borders did not explain all of the observed BX-C spatial boundaries, we explored if genetic elements might account for the physical partitioning. We first examined the TAD border between genetic domains 2 and 3 (Fig. 4f, g). In wild-type A2–A4 cells, both the *Ubx* and *abd-A* regions are in an active epigenetic state, thus epigenetic differences cannot account for the presence of the TAD border (Fig. 5a). We examined embryos homozygous for a 4-kb deletion[39] spanning this border. In these mutants, we observed a fusion of the TADs flanking the deletion, with increased contact frequency extending over 70 kb upstream and downstream of the deletion (Fig. 5a, b). In A2–A4 wild-type cells, *Ubx* and *abd-A* are transcribed in many of the same cells, but at quantitatively different levels due to their different enhancers (Fig. 5c). In mutants, both genes were expressed at similar levels throughout A2–A4 (Fig. 5c), suggesting enhancer cross-talk.

This 4-kb deletion was previously characterized based on its abnormal A1 morphology, and accordingly named *Front-Ultraabdominal (Fub)*[39]. This morphology was attributed to ectopic *abd-A* expression observed in A1. Consistently, we observed aberrant chromatin structure and erroneous enhancer contacts with *abd-A* in A1 (Extended Data Fig. 10). Homozygous mutants died before adulthood, and heterozygous mutants showed multiple developmental defects (Extended Data Fig. 10c-e) in addition to the partial A1 to A2 transformations originally identified[39].

To further test if elements at TAD boundaries identified by ORCA contribute to TAD separation, we examined mutants carrying a ~3-kb deletion[40] spanning another TAD border detected in wild-type between genetic domains 7 and 8 in cells from A7–A9 (Fig. 5d). Notably, both domains are in the same epigenetic state in these cells, thus this border cannot be explained by epigenetic state differences. In mutant embryos, this TAD boundary disappeared and the two TADs fused into one (Fig. 5d, e). This mutation was previously isolated and named *Frontabdominal-7* (*Fab-7*)[40], and while homozygous mutants were fertile, they exhibited a reduction in fecundity and lifespan relative to wild-type (Extended Data Fig. 10c, d).

It has recently been proposed that genome organization is driven by interactions among homotypic chromatin states and that vertebrates have evolved an “additional layer” of organization in which boundary elements and cohesin interact, whether through “loop extrusion” or other mechanisms, to further partition the genome[5,38]. We observed that: (1) TAD boundaries formed between regions of a common epigenetic state, (2) these boundaries were marked by cohesin and CTCF (among other proteins, such as CP190, which may contribute to boundary element function[41,42]), and (3) deletion of these boundary elements led to a fusion of the neighbouring TADs. Even if these observations prove to be unique to the BX-C, they challenge the conclusion that the “additional layer” of genome organization is solely a vertebrate innovation[5,38].

ORCA, together with two recent reports[16,17], highlights the promise of microscopy as a relatively inexpensive approach for analysing chromatin nano-structure in large numbers of single cells. Our approach emphasizes the ability of *in situ* imaging to compile multiple

types of information: chromatin structure, mRNA, nascent RNA, cell position in the embryo and the identity of a cell's neighbours. Due to the high resolution of ORCA, we resolved enhancer-promoter contacts on the kb-scale typical of many developmentally-important interactions. Combined with multiplex RNA labelling, ORCA enabled us to address fundamental questions relating to enhancer specificity and boundary element function. While previous results have suggested that genome structure is largely static during development and across cell types[35,43,44], our results support an emerging view that for developmental control loci, cell differentiation is accompanied by extensive 3D remodeling of chromatin structure[23,45-49]. We anticipate the unique properties of ORCA will enable exploration of uncharted genomic regulatory landscapes.

Methods

Microscope setup

All imaging experiments were performed on one of two custom-built microscope platforms optimized for HILO illumination[51] and mounted on a high-performance optical table (Newport RS4000). The systems were constructed similarly, except where noted. The sample was mounted on a 3D stage (Ludl 96S107-N3-LE2) with a piezo *z*-positioner (Ludl 96A602) on a Nikon Ti-U or Ti2 body. Samples were imaged using a 60× 1.4 NA objective (MRD01605) and a Hamamatsu sCMOS camera factory-calibrated for single-molecule imaging (C11440–22CUPLUS-KIT). The system with the Ti2 body had a larger aperture size and used a 1536 × 1536 (pixels) field of view (see ORCA imaging). The Ti-U system used a 1024 × 1024 field of view. In both cases the pixel size was calibrated to be 154 nm. Illumination was provided by a 561-nm solid-state laser (MPB 2RU-VFL-P-2000–560-B1R) and a 647-nm solid-state laser (MPB 2RU-VFL-P-2000–647-B1R). The beams were combined using dichroics (Semrock, Di03-R561-t3–25D, Di03-R635-t3–25D), power modulated using an acousto-optic tunable filter (AOTF) (Gooch and Housego 97–03309-01), and directed to the sample using a custom polychroic mirror (Chroma zt405/488/561/647/752rpc). Emitted and scattered light were further filtered with a custom emission filter (Chroma zet405/488/561/647–656/752m) prior to being collected by the camera. Hardware-based timing control of the camera shutter to the laser power controls and AOTF was accomplished using a DAQ card from National Instruments (PCIe-6353).

A custom-built focus lock was used to measure the distance between the sample coverslip and the objective to correct for axial drift and maintain focus during imaging. This system worked by measuring the distance between a pair of infrared laser beams reflected off the coverslip and was conceptually similar to one previously described[13], with the following modifications. In place of a single beam splitter coupled to a pair of alignment mirrors, two adjustable non-polarizing beam splitters were each coupled with an adjustable mirror to split the 940-nm IR beam (or 915-nm for the Ti-U system) and provided enhanced control in aligning the focus spots. A pellicle beam splitter at the microscope backport was coupled to an absorptive IR filter to remove stray reflections, which would otherwise occasionally interfere with the system performance.

Software-controlled microscope components, including the camera, *x, y*-stage, piezo *z*-stage, AOTF, and laser illumination, were run using the open-source "storm-control"

software project developed by Hazen Babcock and collaborators, with modifications for our hardware configuration as appropriate. Links to the open-source, publically accessible branch of this software are included in the Code Availability section.

Automated fluid handling setup

Fluid handling steps between each imaging round (see ORCA imaging) were performed using a robotic fluid handling setup. First, 200–500 μl probe solution (25% ethylene carbonate and 2x SSC containing "readout" oligo (0.1 nM)) was washed over the sample and allowed to hybridize for 15 minutes (min). Then 1 ml of wash buffer (30% formamide in 2x saline-sodium citrate buffer (SSC)) was washed slowly over the sample for 2 min, followed by a slow wash with 1 ml of rinse buffer (2x SSC) for another 2 min. Afterward, the sample chamber was filled with imaging buffer (0.5 mg/ml glucose oxidase, 40 μg/ml catalase and 10% w/v glucose in 2x SSC). Imaging buffer was stored under a 1–2 cm layer of mineral oil during the experiment due to its sensitivity to oxygen.

As a typical experiment involved more than 50 fluid exchange cycles, the fluid handling steps were performed using a homebuilt robotic system. The system consisted of four primary components: (1) a robotic buffer selection device; (2) an electronic 8-way valve controller; (3) a fluidic chamber; (4) a peristaltic pump and (5) an imaging chamber (FCS2, Bioptechs). Components 1, 2, and 4 could be controlled by the computer through USB connections. The buffer selection device consisted of a buffer reservoir connected beneath a needle mounted on a 3-axis stage. The stage allowed the needle to be inserted into any one of the wells of the 96-well plate or either of two large reservoirs containing wash buffer or rinse buffer. This system was constructed using a repurposed 3-axis CNC router engraver (3040 CNC router, available on Amazon for a fraction of the price of commercial 3D pipette robots). ETFE tubing (1/16 in. outer diameter (OD) and 0.016 in. inner diameter, (ID)) connected the needle to the valve controller (Hamilton 36798) with an 8–7 loop flow junction (Hamilton 36782), allowing the needle to be alternately connected to the either the fluid chamber (to add buffer) or shortcut to the waste (to allow needle rinsing, which occurred whenever imaging buffer was sent to the sample). A second needle permanently connected the imaging buffer reservoir to the same valve controller. The controller disconnected this needle from the peristaltic pump during the hybridization sequence (resulting in no flow) and connected it to the sample chamber when imaging buffer was needed. Two fluid lines connected to a peristaltic pump (Gilson F155006) provided the flow needed for adding buffers to the sample or rinsing the robotically-controlled needle. Open-source software for the control of the fluidics system is described in the "Code Availability" section below.

The fluid exchange steps required for ORCA can also be performed manually, removing dependence on the robotic fluid handling system described above. To facilitate manual washes, one end of Tygon tubing (1/16 inch internal diameter) was fixed to the exit valve of the Bioptechs chamber and the other end placed in a waste collection chamber at a lower elevation than the microscope stage. Another short piece of Tygon tubing was attached to the entrance valve of the chamber and fluid was added to this opening using a 200-μl pipette. Manual fluid handling, and any additional perturbations that it may cause, can be corrected

by the automated stage drift correction, as for the robotic fluid handling (see Quantification of DNA structure).

Oligopaints library design

The sequences in the Oligopaint probe pools were designed computationally building on previously-described approaches[20,52]. Briefly, first the region of interest (100–700 kb) was divided into even segments at the desired resolution (2–10 kb), see Supplementary Data Table 1. Each segment was assigned a unique barcode from a list of 199 orthogonal 20-nucleotide (nt) sequences (see Supplementary Data Table 2). Non-overlapping 40-nt sequences for targeting barcodes to genomic DNA were selected with the following criteria: 1) melting temperature above 65°C and G/C range 20–80%, 2) fewer than 12-nt homology to any barcode in our set, and 3) fewer than 14-nt homology to any genomic repeat elements identified by RepBase (https://www.girinst.org/repbase/). A minimum of twenty of the 40-nt targeting sequences were selected to uniformly paint each barcoded 2–10 kb segment. To minimize crosstalk between barcodes, we required that no more than 17 nt of the 40-nt targeting sequence could match sequences from any other barcoded segment. For probes targeting RNA, the strand of the probe was selected to be anti-sense to the target RNA. For probes targeting DNA regions, the probes were selected to be sense to any transcribed portions of the target genome, in order to avoid cross-talk with RNA. Finally, probesets were also assigned a 20-nt fiducial-probe-binding sequence (catcaacgccacgatcagct), and a unique forward and reverse index primer to allow the probeset to be isolated from an oligopool (see Supplementary Data Table 3). The resulting sequences (22,908 120-nt sequences, see Supplementary Data Table 4), were ordered in an oligopool from CustomArray.

In contrast to many prior DNA FISH probes (30–500 kb), our probes covered shorter stretches (2–10 kb) of DNA, which improved the genomic resolution. It also improved the precision of the distance measurement between labelled regions, as the precision of the centroid distance between two regions cannot exceed the absolute size of the regions themselves, which is ~200 nm for typical 30-kb domains[13,17].

Probe synthesis

Probe synthesis was performed as previously described[13]. Briefly, Oligopaint probes from the oligopool (CustomArray) were amplified from the array-synthesized library via a limited-cycle PCR. All primers used for this amplification were acquired from Integrated DNA Technology (IDT). The PCR product was column purified (Zymo DNA Clean and Concentrator; D4003) and further amplified through an *in vitro* transcription reaction (New England Biolabs, E2040S, Promega RNasin, N2611). Next, the RNA product was converted back to single-stranded DNA (ssDNA) via a reverse-transcription reaction (Thermo Scientific, EP0751, EP075). The ssDNA was column purified (Zymo Research, D4006) and the Oligopaint probes were eluted in 50 μl of ultrapure water. Afterwards the probe was concentrated using a DNA SpeedVac™ (Thermo) and resuspended in 25 μl ultrapure water. Probes were stored at –20°C for future hybridization usage.

Embryo collection

Flies were placed in a collection cage with an apple juice agar plate containing yeast paste at the bottom of the cage. All flies used were Canton-S wild-type strain unless otherwise indicated. Two hours (h) after egg lay in the collection cage, the agar plate was removed and left at room temperature until embryos reached the desired developmental stage. Embryos were dechorionated by treating them with bleach for 1 min while swirling the plate to loosen the embryos from the agar. Embryos were then transferred into collection baskets and rinsed with deionized water for 1 min. Dechorionated embryos were transferred to a scintillation vial containing 5 ml of heptane and 5 ml of aqueous buffer consisting of 8% paraformaldehyde (PFA), 50 mM EGTA in 1x phosphate-buffered saline (PBS) and placed on a shaker table for 25 min for fixation. The aqueous PFA layer was then removed and replaced with 5 ml 100% methanol. The vial was vortexed for 3 min to devitellinize the embryos. Devitellinized embryos sank to the bottom of vial, in methanol layer. The heptane layer was then removed, and the embryos were transferred to 1.5-ml microcentrifuge tubes, underwent three short washes in 100% methanol, and stored in methanol at –20°C or immediately embedded for cryosectioning.

Embryo preparation for labelling with primary probes

The methanol was removed from the embryo storage tubes and the embryos were washed in 1:1 methanol : 1x PBS for 5 min while nutating. Three short washes with 1x PBS were used to remove all remaining methanol and to rehydrate the tissue. Samples were incubated in 30% sucrose in 1x PBS for 2 h at 4°C. The 30% sucrose prevents ice crystals from forming in the tissue during freezing. Afterwards, embryos were embedded in a 15-mm x 15-mm x 5-mm Tissue-Tek Cryomold (VWR). A mixed population of embryos containing males and females were selected at random and pipetted into the cryomold. Any remaining sucrose solution was removed. Using a paintbrush, embryos were aligned along the anterior-posterior (A-P) axis under a dissection scope. Afterwards, Optimal Cutting Temperature (OCT) compound was added, and the cryomold was placed onto a block of dry ice for 15 min. Frozen, embedded embryos were then stored at –80°C until needed for cryosectioning.

Prior to cryosectioning, 40-mm glass coverslips (Bioptechs Inc.) were thoroughly cleaned with 100% methanol and coated with a chromium-gelatin solution for 30 min. The chromium-gelatin solution was prepared by dissolving 1.5 g gelatin, and 0.15 g of chromium potassium sulfate dodecahydrate in 300 ml of deionized water. Embedded embryos were cryosectioned at a thickness of 6 μm and collected onto coated coverslips. The whole embryo was sectioned, and the middle section of the sample was chosen for hybridization due to the clearly visible morphology of the segments along the A-P axis. Coverslips were either stored at –80°C or were directly used for primary probe hybridization.

Mouse embryonic stem cell culture and preparation

Mouse embryonic stem cells, line R1, were cultured in standard conditions as previously described[53]. In preparation for imaging, cells were plated on glass coverslips and fixed in 4% PFA for 10 min. All other steps were performed as described in *Primary probe hybridization*. The mES R1 cells were a generous gift from the Wysocka lab. The cell line was authenticated by colleagues in the Wysocka lab by karyotyping, RNA-seq, ChIP-seq,

and in vitro differentiation all of which gave results consistent with the pluripotent stem cell identity. Additional chimera assays were not performed. The cells tested negative for mycoplasma contamination prior to our experiments.

Primary probe hybridization

Hybridization procedures were carried out at room temperature except where noted. For experiments labelling both RNA and DNA on the same sample, the samples were first hybridized with primary probes targeting RNA.

For RNA labelling, fixed cryosectioned embryos (initially fixed during embryo collection) were again fixed for 10 min with 4% PFA in 1x PBS and then washed three times in 1x PBS. Embryos were then permeabilized for 10 min with 0.5% Triton-X in 1x PBS and washed three times with 1x PBS. Embryos were then incubated for 35 min in hybridization solution (2x SSC + 50% vol/vol formamide and 0.1% Tween). The hybridization solution was aspirated, and 30 μl of probe hybridization buffer (50% formamide, 2x SSC, 0.1% Tween-20, 10% dextran sulfate) containing 750 ng of primary probe was pipetted directly onto the embryos. An 18 × 18-mm coverslip was placed to spread the probe evenly throughout the tissue. The sample was placed onto a heat block at 90°C for 10 min, as we found this improves labelling efficiency and does not result in significant nuclear DNA labelling. The sample was then incubated in a humid chamber overnight at 42°C to hybridize. The following day, the sample was washed two times with 2x SSC at 42°C for ten minutes, and the coverslip was removed after the first wash. Samples were stored at 4°C for up to two weeks or directly mounted into a Bioptechs flow chamber for imaging (see ORCA imaging).

Subsequent barcode hybridization and imaging steps were conducted on the microscope. The 20-nt barcode sequences in the RNA primary probes were labelled through sequential hybridization of complementary readout oligos with a Cy5-fluorophore and imaged. Following RNA image acquisition, the embryos were removed from the microscope for a second round of labelling with DNA primary probes.

For DNA labelling, the hybridization procedure was as follows. The fixed cryosectioned embryos were permeabilized for 10 min with 0.5% Triton-X in 1x PBS and washed three times with 1x PBS. The sample was incubated with 0.1 M of HCl for 5 min, to denature the DNA, and then washed three times in 1x PBS. If primary probes contained any regions which were antisense to RNA transcripts in the labelled region, the sample was incubated for 0.5–1 h at 37°C in RNaseA solution, 10 μg/ml, (Fisher, EN0531) to prevent cross-reactivity with RNA. Embryos were then incubated for 35 min in hybridization solution. The hybridization solution was removed, and 30 μl of hybridization buffer containing 750 ng of primary probe was pipetted directly onto the embryos. An 18 × 18-mm coverslip was placed to spread the probe evenly throughout the tissue. The sample was placed onto a heat block at 90°C for 10 min. The sample was then hybridized in a humid chamber overnight at 42°C. The following day, the sample was washed two times with 2x SSC at 42°C for 10 min, and the coverslip was removed after the first wash. Embryos were postfixed in 8% PFA + 2% glutaraldehyde (GA) in 1x PBS for 1 h and then washed three times in 2x SSC. Samples were stored at 4°C for up to two weeks or directly mounted into a Bioptechs flow chamber

for imaging (see ORCA imaging). Barcode labelling and imaging were then done on the microscope.

ORCA imaging

To label 3D positions of all barcodes, a pipetting robot facilitated robust and continuous rounds of readout oligo hybridizations. The 40-mm coverslip with the hybridized embryos was mounted in a flow chamber (Bioptechs) and connected to a homebuilt, robotic fluidics system (see Automated fluid handling setup). In place of a robotic fluid handling system, probes and wash buffers can also be manually pipetted into the Bioptechs flow chamber, an approach we often found useful when troubleshooting the homebuilt system (see Automated fluid handling setup).

For robotic addition of probe sequences, a 96-well plate containing the readout probes was mounted in the fluidics system. Each well held 600 μl of probe solution, consisting of a 50-nt complementary “readout” oligo (0.1 nM), (see Supplementary Data Table 5) bound to a 20-nt Cy5-labelled “imaging” oligo, Cy5–5p-TGGGACGGTTCCAATCGGATC (0.12 nM), and an extended 10-nt toehold sequence (Fig. 1a, Extended Data Fig. 1a) in 25% ethylene carbonate and 2x SSC buffer. The first well also contained a Cy3-labelled “fiducial” oligo complementary to the fiducial tag sequence (0.1 nM), to be used for image registration and drift correction (Extended Data Fig. 1a). The remaining wells also contained a 30-nt strand displacement oligo (0.3 nM) complementary to the previous readout probe (see Supplementary Data Table 6). The fluorescently-labelled oligos, readout oligos and strand-displacement oligos were acquired from IDT. The use of readout oligos that bind a common imaging oligo substantially reduces the cost for sequential hybridization experiments compared to previous approaches[18,20,52,54].

The sample was secured to a motorized stage on a homebuilt microscope (see Microscope setup). Sequential hybridization and imaging were then performed using automated software routines (see Code availability for links to the source-code) as follows (see also Extended Data Fig. 1a): Samples were pre-bleached for 30 seconds (s) with a 647-nm laser to reduce tissue autofluorescence prior to labelling. Readout probe from the first well was flowed onto the sample and hybridized to the barcode for 15 min. The sample was then washed for 140 s with wash buffer (30% formamide in 2x SSC), rinsed with rinse buffer (2x SSC) followed by the addition of imaging buffer. Imaging buffer consisted of 0.5 mg/ml glucose oxidase, 40 μg/ml catalase and 10% w/v glucose in 2x SSC. The imaging buffer was made fresh for every experiment and then stored under a 1–2 cm layer of mineral oil during the experiment due to its sensitivity to oxygen.

All selected fields of view in the sample were imaged sequentially in *z*-stacks of 100-nm steps. At each step, the Cy5 and Cy3 channels were imaged sequentially, allowing for nanoscale drift during acquisition to be corrected. An additional ten frames at the beginning and end of the stack were recorded in the center *z*-position, as a control for stability. Images were collected using a pixel size of 154 nm as an array of 1536 × 1536 pixels or 1024 × 1024 pixels, depending on the microscope platform (see Microscope setup). Overlapping edges were used to ensure reliable mosaic reconstruction of embryo images. Double-counted

cells were identified and removed during the image registration and analysis steps (see Quantification of DNA structure).

Subsequent rounds of hybridization with different readout probes first required addition of the strand-displacement oligo to remove the previous readout oligo, then addition of the new readout oligo to label the next barcode. After each step of labelling, imaging and strand displacement of the barcode, 2x SSC was added to the sample and each position briefly photobleached for 3 s to prevent accumulation of off-target fluorescence. Control experiments omitting the Cy5-labelled oligo and containing only the strand-displacement oligo exhibited efficient removal of signal to undetectable levels within 15 min (Extended Data Fig. 1b). Notably, strand displacement is faster than photobleaching used in previous sequential imaging work[18,20], allowing for analysis of more cells. In contrast to previous work using cleavage of a fluorescent dye[52,54], strand displacement also allows for repeated measurements of the same barcode from the same sample (Extended Data Fig. 1b, c). As discussed below and in *Analysis of barcode detection efficiency*, this is especially useful for error quantification. The hybridization, imaging, and photobleaching cycles repeated for up to 75 rounds. Typical experiments took 2–4 days to complete all rounds of labelling and imaging.

To quantify the combined uncertainty/error introduced by any residual drift, uncertainty in our sub-diffraction measurements, sample movement or degradation, and/or pixel noise in our camera, 4–10 barcodes were selected for repeat labelling and imaging (Extended Data Fig. 1c). The distribution of displacements measured with this approach typically registered to an accuracy of ~30 nm in x-y and ~60 nm in z over the course of the experiment, with displacements measured less than 1 h apart being essentially indistinguishable from those measured at the end of a 96-h experiment (Extended Data Fig. 1c).

STORM imaging

Sequential Stochastic Optical Reconstruction Microscopy (STORM) imaging was performed similar to the ORCA imaging process described except for the following notable changes. A STORM imaging buffer, as previously described, was used in place of the ORCA imaging buffer, and contained 1% 2-mercaptoethanol to promote photo-switching[13]. STORM images of each region, 1024 × 1024 pixels, were imaged for 10,000 frames at 20 Hz using the 647-nm laser set to 1.9 W at the laser head. STORM movies were analysed using 3D-DaoSTORM as previously described[55]. Non-blinking, weakly autofluorescent molecules in the tissue were used for drift correction during image acquisition. Localizations were binned into 15 × 15-nm pixel arrays and the number of peaks with a 150-nm radius centered on each barcode position were recorded. The fraction of localizations containing only a single peak was plotted as a conservative estimate of the number homologs paired more tightly than the resolution of STORM (see Extended Data Fig. 3). To reduce background calls, pixels with fewer than 4 localizations were excluded. This is a conservative estimate of background, as individual Cy5 fluorophores typically exhibit multiple switching cycles before photobleaching and labelled homologs are expected to carry a tile of more than 20 barcodes, each with a Cy5 fluorophore.

Quantification of RNA expression

Sequential images of each field of view were registered and corrected for *x*-*y* stage drift using correlation-based alignment of the Cy3-channel images, which contained pictures of the fiducial label marking all RNAs. Typical drifts of 0–20 pixels were observed between raw images prior to correction. Following correlation-based alignment, the data registered to pixel accuracy or better. All fields of view of the drift-corrected data were then combined into a single mosaic image. Minor stage drift also produced offsets between fields of view (in addition to the drift between repeated imaging rounds of the same field of view, which was already corrected at this step). This drift was also corrected using correlation-based registration of the fiducial image data, using the ~10% overlapping edge pixels between adjacent fields. Individual embryos were cropped from the registered mosaic image by manual selection of a bounding polygon. Drift-corrected RNA data from cropped mosaics was then aligned embryo by embryo to the cropped, drift-corrected mosaics of DNA data through rigid translation and rotation that maximized the cross-correlation between the Cy3-fiducial channels of the DNA and RNA images. The inclusion of highly expressed nascent RNA probes from the same locus targeted by DNA imaging facilitates this alignment. Because minor rotations are introduced in removing and remounting the sample, this autocorrelation corrected for rotational variation, which increased the computational time necessary for alignment. Bright nascent transcripts from different genes in the Cy3-fiducial channel (from the RNA imaging experiments) aligned within 1 μm of their DNA locus of origin labelled in the Cy3-fiducial channel of the DNA data.

To count individual RNA transcripts per cell, bright foci corresponding to single mRNA transcripts in maximum-*z*-projected images were identified using a local-maxima search with manually defined thresholds. Minor differences in the illumination across the field of view were corrected using images measuring field flatness prior to threshold selection. Thresholds were selected for each RNA channel by visual inspection of the quality of agreement of the bright points visible in the data channel with the returned localizations. Embryos were segmented into distinct cells using the ORCA-labelled DNA to mark the nuclei and a watershed algorithm to identify the local nuclear area, as previously described[56]. The total number of RNA molecules identified for each species per cell and density of localizations per cell were recorded. Nascent RNA transcripts were identified using a similar approach and the fluorescent intensity of the detected nascent foci recorded (as a single nascent RNA focus may reflect multiple distinct transcripts).

For the Odds Ratio analysis in Fig. 3, all detected nascent transcripts were used to define the ON state (see Odds Ratio). For the RNA levels per segment shown in Fig. 4, the average nascent transcript brightness per segment in each embryo was computed and then normalized all nascent RNA values relative to the average in the brightest segment.

Clustering of cell states

Following the approach described in previous analyses of single-cell RNA-abundance using sequencing-based methods[57], the expression matrix of all cells from all embryos for all RNAs assayed was projected into a two-dimensional expression space using t-Distributed Stochastic Neighbor Embedding (t-SNE) with a *z*-score normalized Euclidean distance

metric[58] in Matlab (R2017a). The resulting clusters were separated using density-based clustering (DBSCAN)[59] (Matlab implementation, Yarpiz 2015). Only cells assigned to substantial groups (> 300 cells) were considered for further analysis.

Identification of embryo segments

Manual identification of cell types was performed using a combination of embryonic morphology and multiplexed RNA labelling. Segments were identified in germband-retracted (8–11 hpf) embryos by the invaginations of the embryo surface and by the expression of the genes *engrailed* (*en*) and *invected* (*inv*), which mark the posterior compartment of each body segment. *Drosophila* specialists will recall that mutations in regulatory elements in the BX-C affect the parasegments, as the Hox gene expression patterns span the boundaries to the posterior portion of the next anterior segment, rather than align to the segments proper. For a more accurate picture, we therefore partitioned the embryos based on the parasegment boundaries, not the morphological body segment boundaries marked by invaginations of the animal body wall. Following prior convention for simplicity and consistency with earlier work, we refer to these parasegments by their corresponding segment names, not by parasegment numbers. This allows for greater consistency, as the genetic domains controlling each of these parasegments are also named for the corresponding segment (e.g. the genetic domain known as *iab-2* is named segment A2, but controls parasegment 7, which in reality includes the posterior-most part of A1 and the anterior portion of A2). This simplification of using segment names is common in earlier literature[60], and avoids a discussion of details of *Drosophila* Hox genetics that are not essential for understanding of the work presented here.

Identification of *Fub* mutants

Front-Ultraabdominal (*Fub*) mutants were not viable and thus required the use of balancer *Fub-/MKRS*. To examine chromatin organization in mutants with a 4-kb deletion of *Fub,* a known insulator element, embryos from *Fub-/MKRS* were collected (see Embryo collection). Their progeny led to both heterozygous and homozygous *Fub* embryos, which were distinguished by their RNA expression pattern. Previous experiments reported ectopic activation of *abd-A* expression in segment A1 in *Fub-/Fub-* mutants[39]. Through multiplex RNA labelling, homozygous mutants were identified as those with aberrant *abd-A* mRNA and nascent *abd-A* expression compared to heterozygotes, that had wild-type *abd-A* expression pattern (Extended Data Fig. 10). Field of view positions of the mutant embryos were recorded to examine the structural organization of the BX-C at 10-kb resolution (Fig. 5a, b). Note that RNA analysis was not necessary for *Fab-7* experiments (Fig. 5d, e) as these mutants were viable as homozygous.

Quantification of DNA structure

Analysis of ORCA data began with correcting *x-y* stage drift using cross-correlation-based alignment of the Cy3-channel images of the fiducial labels. As described in *Quantification of RNA expression*, all primary probes contained a binding site to hybridize the fiducial Cy3-labelled oligo, thus the fiducial channel recorded an unchanging image of the entire domain to enable drift correction. After correcting for stage drift, the fiducial data were also

used to identify all cells containing the labelled region of interest using a manually-selected threshold to identify these bright foci in the image.

The individual regions were then automatically cropped from the image in a 20 × 20 × 100-voxel box (each voxel corresponding to 154 × 154 × 100-nm for *x, y, z,* respectively) centered on the fiducial label. These single-locus images were then registered to sub-pixel accuracy (typically 20–30 nm) in 3D through nonparametric kernel density estimation, using bicubic interpolation of the fiducial spot and maximum autocorrelation of the resulting kernels from each labelling round to compute the remaining subpixel drift. Each barcode image was corrected by the measured *x, y, z* drift and then fitted in 3D using a 3D-Gaussian spot-fitting algorithm. The position of each barcode in 3D, along with the 95% confidence intervals on the fitted and Gaussian fitted parameters, were saved in a data table (See Data availability). Images with insufficient signal to be fitted by the Gaussian algorithm were recorded as missed detection events, represented with 'nan' symbols in the data table. If multiple Gaussian peaks were detected within the analysed area, only the brightest was reported by the algorithm. The tables constructed for each detected barcode in each cell also recorded a tracking number for all labelled regions in the field of view, their coordinates, and the field of view number. The resulting list of 50–80 *x, y, z* positions for each labelled region was used for reconstructing polymer images of the locus and for computing pairwise distance maps (see Normalization of distance effects).

For the data collected from *Drosophila* embryos between 8–12 hpf, over 90% of cells exhibited paired homologs, and thus a single trajectory was recovered in these cells. STORM experiments (see STORM imaging, Extended Data Fig. 3) demonstrated that individual barcodes were tightly paired, and not resolved even with ~30–50-nm super-resolution imaging. As this distance is on par with the uncertainty measurement in the barcode positions (see Extended Data Fig. 1) and smaller than the typical distance between the centroids of adjacent barcodes, no artefacts were introduced by this averaging of the two polymer paths into a single trajectory (the uncertainty in the trajectory of a single path was already larger than the typical distance between the pair).

Field of view positions were converted to embryonic coordinates by construction of mosaic images from the individual fields of view. For the RNA images of the same embryos, this was completed using the recorded stage positions of each field of view, corrected for drift by registering the overlapping edges of each field of view. Following mosaic stitching, cells from overlapping edge pixels, and thus present in more than one field of view, were identified by detecting the overlapping coordinates and the redundant cells were removed from the analysis. The reconstructed embryo mosaics from the ORCA analysis were aligned to the reconstructed mosaics from the RNA data for integrated analysis. This alignment was accomplished using autocorrelation-based registration (allowing for rotation) of the fiducial channel data corresponding to each mosaic. A multi-scale iterative alignment approach was used to make this registration efficient for mosaic images with hundreds of millions of total pixels.

All image processing computations were carried out on a T630 Dell PowerEdge with 25 cores and 256 GB of RAM.

Analysis of data reproducibility

Biological replicate experiments of the 10-kb and 3-kb resolution datasets were compared by analysing the correlation between the contact frequencies among all barcode pairs and the inter-barcode 3D distance among all barcode pairs (Extended Data Fig. 1d). The replicate experiments at 3-kb resolution represented two slightly different developmental time points (8–10 hpf and 10–12 hpf). However, the 3D structures of the BX-C were highly correlated at these two time points. Segment-by-segment comparisons between these embryos confirmed the qualitative and quantitative similarities between these datasets (Extended Data Fig. 9). Consequently, the two populations were combined for the analyses in Figure 3 and Fig. 4. As the 8–10 hpf dataset contained only 12 RNA species rather than the 29 analysed in the 10–12 hpf dataset, only the latter dataset was used for the unbiased clustering based on gene expression, shown in Figure 2 and Extended Data Fig. 6a.

Analysis of barcode detection efficiency

Detection efficiency per barcode was computed as the frequency of successful localizations of each barcode in a given probeset on an individual polymer. Successful localization required the fluorescent signal to form a sufficiently well defined point-spread function such that the centroid could be reliably fit by our automated detection software (see Quantification of DNA structure). Most barcodes were detected with similar frequencies (~60–70%) within a probeset, though a small number were less efficient (Extended Data Fig. 2). To test for potential bias in detection efficiency per barcode between cells, the distribution of detection efficiency was compared across all barcodes in cells from anterior embryonic regions (where the BX-C is silent) and cells from posterior embryonic regions (where the BX-C contains active genes). These two distinct cell populations were chosen as they are among the most distinct cell populations in the embryo, and as Polycomb silencing of the locus might be expected to affect the efficiency of hybridization in anterior cells. These analyses revealed no significant difference in the detection efficiency per barcode between these cell types for any of our probesets (Extended Data Fig. 2). Similar analyses were performed for the detection efficiency per cell, defined as fraction of barcodes detected in a given cell. This distribution over all cells also did not vary between anterior or posterior cell types (Extended Data Fig. 2). As a further control, the detection efficiency per barcode and detection efficiency per cell between arbitrarily selected pairs of embryos were compared, and again, no bias in either metric between embryos in any of the probesets was detected (Extended Data Fig. 2). As a further control for the impact of missing data, missing data were simulated by randomly removing 50% of detected barcodes. The down-sampled data was highly correlated with the original data for each body-segment analysed (Pearson's $r > 0.92$), indicating that our data are sufficiently over-sampled rather than undersampled (Extended Data Fig. 2f, g).

Normalization of distance effects

To accentuate cell-type specific differences in the data, in Figure 2 the ORCA distance maps were normalized to remove the common polymer nature of the data, whereby regions that are closer in linear space tend to be closer in 3D space. This normalization was computed as follows. From the data tables of polymer coordinates, the pairwise distance was computed

between all elements to generate single-cell distance matrices (in nm). The element-wise average or median of these matrices was taken to provide a statistical, population-based view for groups of single cells. The data from all groupings of cells showed the expected behaviour for a polymer, with a preference for domains nearer in genomic-sequence space to interact more. To adjust for this effect, the average genomic-distance-dependent 3D separation was computed using data from all pairs and all cells and this expected distance map was subtracted from the observed distance map for each cluster. These normalized maps (see Fig. 2e-g and Extended Data Fig. 6d) are referred to as the observed-minus-expected distance maps. Finally, the fraction of cells where pairwise interactions were within 150 nm was computed for each pair and was plotted as the contact frequency map.

Odds Ratio

The odds ratio for transcription, given contact with a given barcoded section of the DNA, as described in Fig. 3 was computed using the standard definition: OR = $(a*d)/(b*c)$. Here a is the number of cells in which contact and transcription was observed, b is the number of cells in which contact was observed and transcription was not observed, c is the number of cells in which no contact was observed but transcription was observed, and d is the number of cells in which neither contact nor transcription was observed. The 95% confidence intervals were computed by bootstrapping. Contact was deemed to be observed between a given barcoded section of the DNA and a given promoter if the region was detected within 150 nm of the promoter.

Fecundity assay

Eight females and twelve males were placed in a fresh vial of dextrose-based fly media for 48 h then removed. The number of adult flies that eclosed (emerged from their pupal cases) over the next three weeks was quantified.

Induced flight assay

Five to twenty flies were transferred to a cylindrical flight chamber, 17-cm tall and 9-cm across, without exposure to carbon dioxide. Carbon dioxide is commonly used to anaesthetize flies for handling and transfer but negatively impacts flight ability for hours after exposure[61], and thus was avoided. Flies were allowed to climb up the sides of the chamber, where they preferentially rest. When all flies were at least 4 cm above the floor, the chamber was perturbed by tapping the sides, causing the flies to take off. Healthy flies took flight and landed back on the sides of the chamber after the perturbation. Uncoordinated or flightless flies fell to the bottom of the chamber. The fraction of flies not found on the bottom of the flight chamber was measured and recorded as the fraction in flight. Fifty-four rounds of induced flight were measured for each genotype.

Sample size choice

Number of embryos and cells were determined by experimental feasibility and not statistical power.

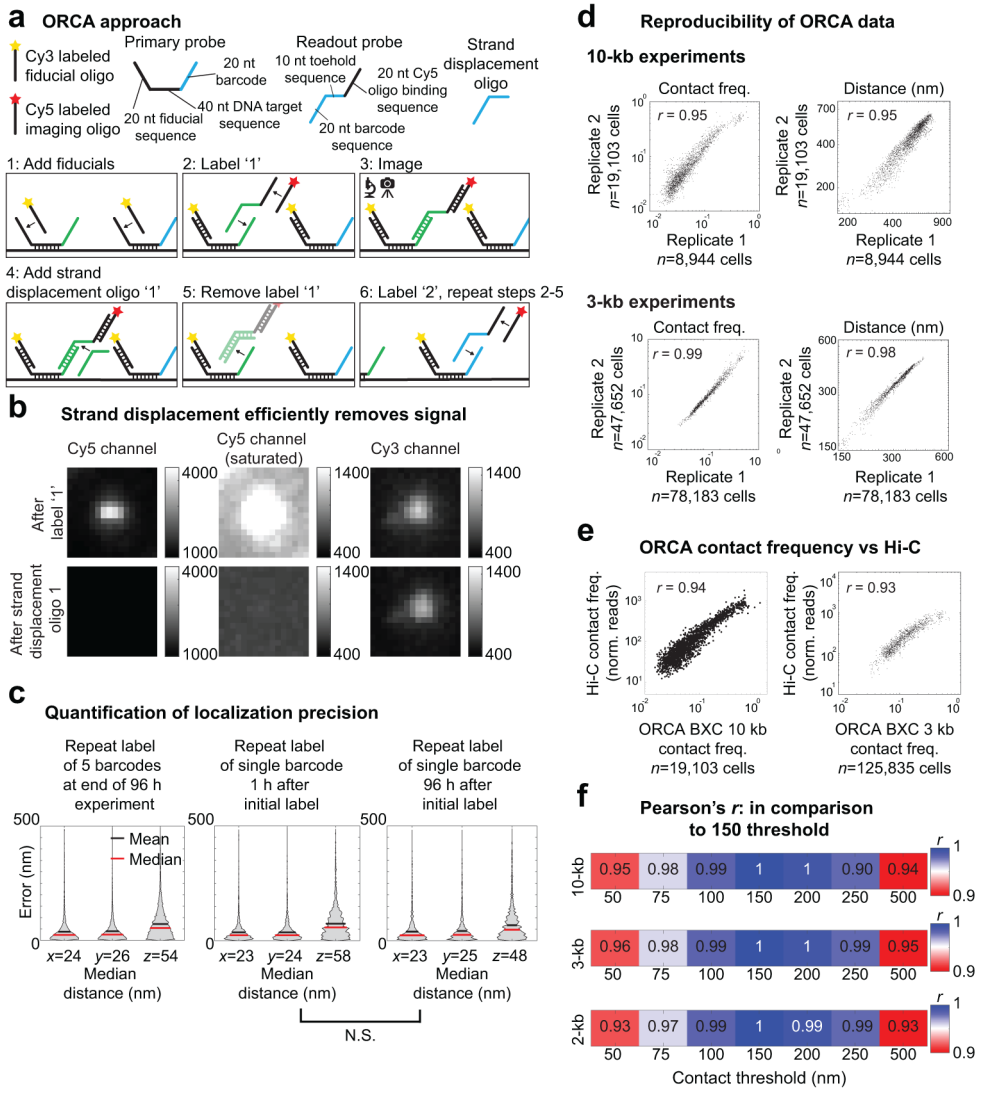
Code availability

All code used in this study is available online at https://github.com/BoettigerLab, except where prohibited by copyright pertaining to derivative works as described below. Software for the control of the automated fluid handling system is available at: https://github.com/BoettigerLab/fluidics-control. Software for the control of the microscope components and for integration of liquid handling and microscope imaging routines is available at: https://github.com/BoettigerLab/storm-control. Code pertaining to the construction of Oligopaint library design used in this study was modified from routines developed in the Zhuang lab at Harvard University[20,52]. The license for those original routines prohibits distribution of derivative works except to other institutions for academic use, precluding the use of publically-accessible code-repositories such as Github. Consequently, these codes will be made available upon request as permitted by the license. Code for the automated analysis of ORCA data is available at: https://github.com/BoettigerLab/ORCA-public.

Data availability

Data tables containing the 3D positions of the barcodes localized by ORCA are available from the online repository: https://github.com/BoettigerLab/BXC-ORCA-data. An annotated map of the domains studied in this work is available as a bed file in the supplemental data and can also be viewed through the UCSC genome browser at this link: https://tinyurl.com/y7wf8t6x. Additional microscopy data is not hosted online due to its size (12.33 Tb), but is available upon reasonable request.

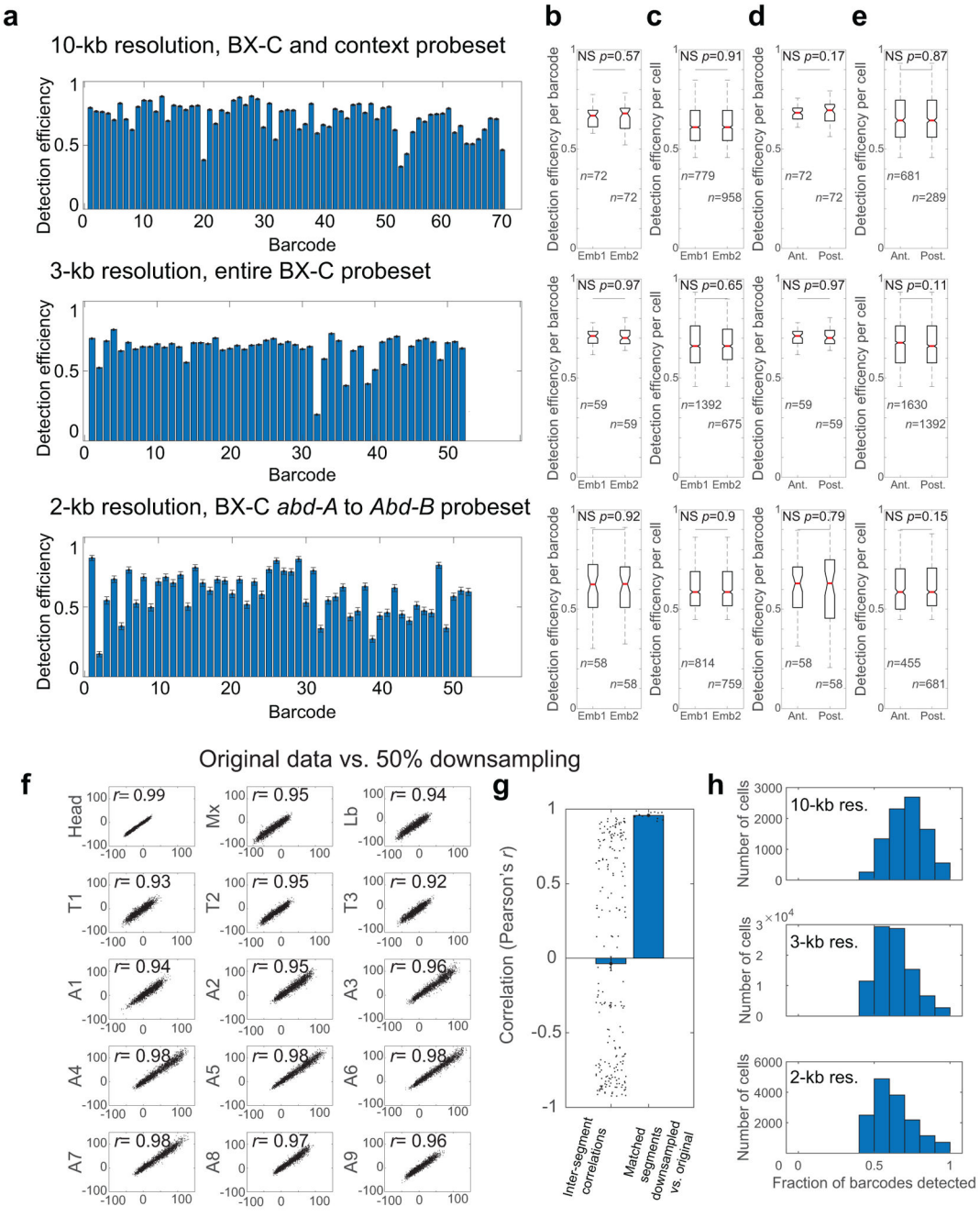
Extended Data



Extended Data Fig. 1 |. ORCA methodology and experimental controls.

a, Schematic of imaging approach. Top panels: oligo species involved in labelling. To minimize costs, only two oligo sequences carry fluorophore labels, a fiducial oligo labelled with a Cy3 dye and an imaging oligo labelled with a Cy5 dye. Sequence-specific binding is achieved by *in situ* hybridization of the primary probes to the genomic DNA. Each primary probe consists of a 40-nt targeting region, a 20-nt arm to bind the fiducial oligo, and a 20-nt barcode to bind the readout oligo. Each barcoded region has at least 20 primary probes, to facilitate detection. Other relevant oligo species include a readout oligo for each barcode (Supplementary Data Table 5) and a strand-displacement oligo for each barcode (Supplementary Data Table 6). Lower panels: data acquisition sequence. The fiducial oligo binds all primary probes and remains bound throughout the experiment. This enables image registration in the downstream image analysis. The readout oligo binds its target barcode '1' and the Cy5 imaging oligo. The now fluorescent barcode '1' is imaged simultaneously with the fiducial Cy3 signal in 3D. The readout oligo '1' is removed by its corresponding strand-displacement oligo, removing the Cy5 signal. This process repeats for the rest of the

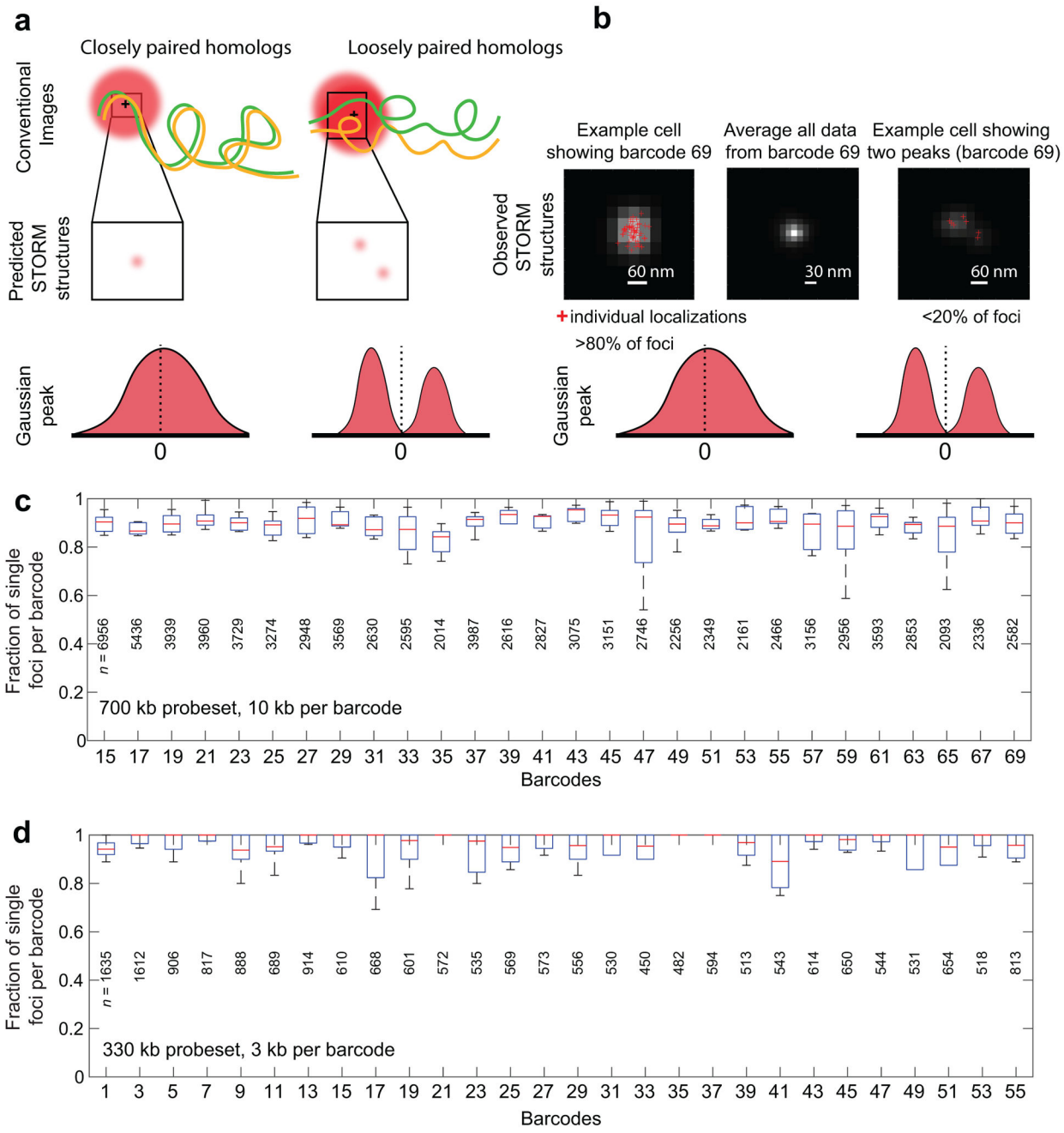
barcodes. **b,** Efficient removal of fluorescent signal by strand displacement: top panels show the fluorescent Cy5 signal from labelled barcode '1' and the corresponding Cy3 signal from the fiducial channel. Bottom panels show the same region after treatment with the strand-displacement oligo. Center panels show the Cy5 channel in increased contrast, showing that all the Cy5 fluorescent signal is removed. Notably the Cy3 image is unchanged. Similar results were obtained in 6,933 cells analysed. **c,** Violin plots of the error in nanometers along each image axis, determined as the difference between the original measurements and measurements repeated at the end of the experiment. Red line marks the median, black marks the mean. Distribution of errors based on 5 repeated barcode-position measurements, n=163,810 repeated barcodes (left). Distribution of errors from measurement of a single barcode's position where the original and repeated measurements are separated by less than one hour (h), n=31,506 repeated barcodes (center). Distribution of errors from measurement of a single barcode's position when the original and repeated measurements were separated by the entire experiment lasting 96 h n=40,032 repeated barcodes (right). N.S. indicates not statistically significant increase in the 3D error (p=0.8, one-sided Wilcoxon test). **d,** Correlation between the replicates of ORCA experiments with both 10-kb barcode resolution (top) and with 3-kb barcode resolution (bottom) when measuring contact frequency (left) and distance (right). Pearson's correlation coefficient, r, is indicated. **e,** Pearson's correlation of contact frequency measured by ORCA and published Hi-C measurements[22] for the 10-kb and 3-kb resolution probesets. **f,** Pearson's correlation coefficient between the 150-nm cutoff (used in the main figures) and alternative cutoff values from 50-to 500-nm. Pearson's r was computed using all unique pairwise combinations of all barcodes measured (70 barcodes in the 10-kb resolution experiments and 52 in both the 3-kb and 2-kb resolution experiments). Relative interaction frequencies and corresponding structural boundaries observed in the contact frequency maps have little dependence on the precise value of the cutoff. Alternate cutoffs remain highly correlated to one-another over a range of values.



Extended Data Fig. 2 | Barcode detection efficiency.

a, The mean detection efficiency of each barcode for all BX-C probesets. Error bars represent standard error of the mean. n=8,801 cells, n=94,300 cells, n=15,230 cells for the 10-kb, 3-kb, and 2-kb probesets respectively. **b,** Box-and-whisker plots comparing the distribution of detection efficiencies across barcodes between two different embryos. N.S. indicates not statistically significant (two-sided Wilcoxon test, p-values indicated). For the boxplots in (b-e), the red line marks the median, the upper and lower limits of the box mark the interquartile range, and the whiskers extend to the furthest datum within 1.5 times the interquartile range. Notches denote the 95% confidence interval around the median. Sample

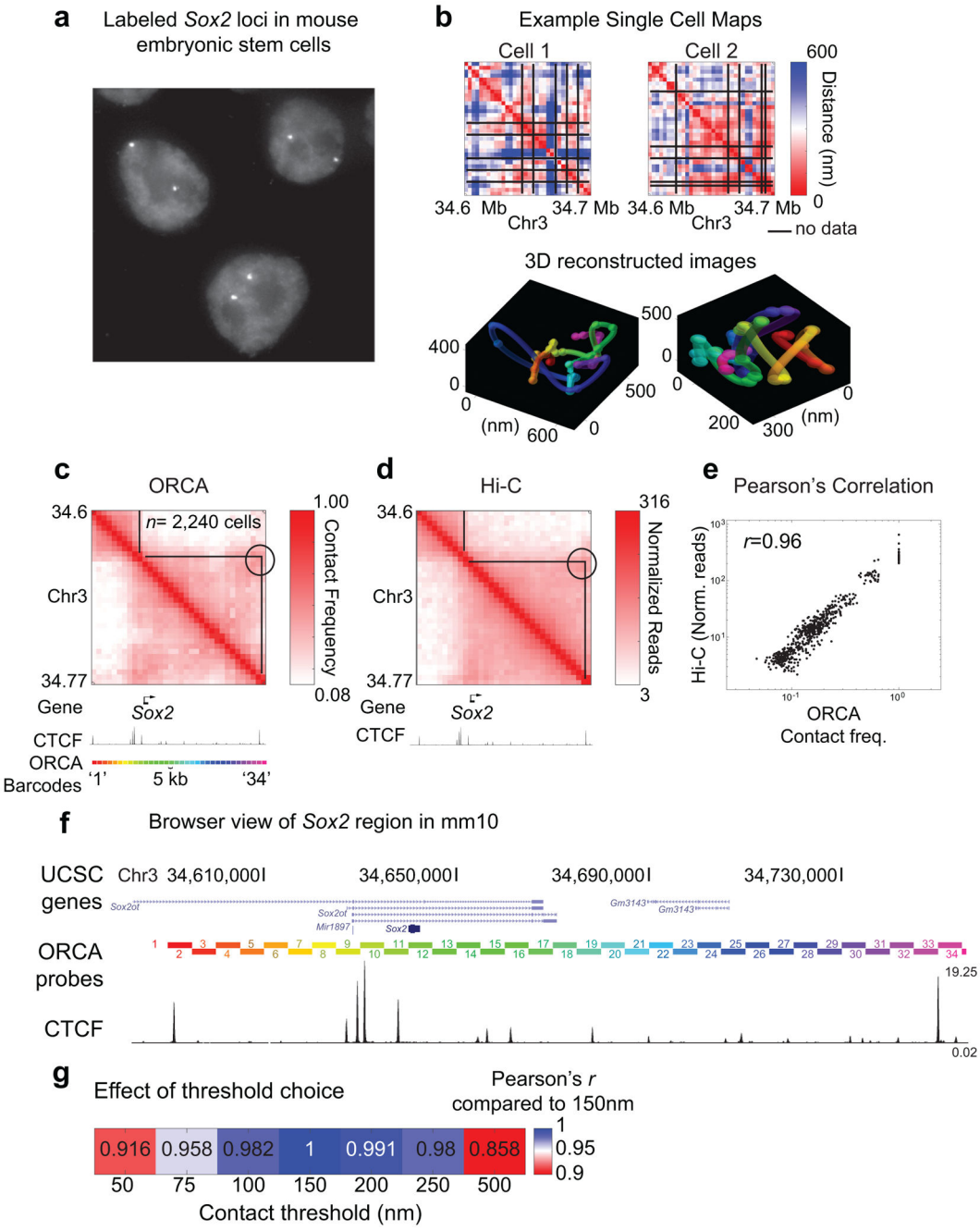
size (*n*) is indicated. **c,** Comparison of the distribution of detection efficiencies between cells of different embryos. **d,** Comparison of the detection efficiency across barcodes between anterior and posterior cells. **e,** Comparisons of the detection efficiency across all cells between anterior and posterior cells. Results are shown for the 10-kb, 3-kb and 2-kb probesets as indicated. **f,** Correlation between normalized distance data for each body segment as shown in Fig. 4 and a modified dataset in which 50% of detected barcodes were removed at random to simulate a high missed detection rate. The correlation coefficients are indicated (Pearson), computed using all unique combinations of 52 barcodes. **g,** Bar graph comparing the mean correlation between matched segments in the original and downsampled data shown in (f) compared to the mean inter-segment correlation. The correlation coefficients are indicated (Pearson), computed using all unique combinations of 52 barcodes. All 15 matched segment pairs shown in f are plotted compared to all 210 inter-segment pairs. This illustrates that the difference between the substantially downsampled data and the original measurements are much more similar than the typical differences between segments, demonstrating the robustness of the segment-specific findings to missing data. Error bars denote standard error of the mean. **h,** Histogram showing the detection efficiency per cell for each probeset.



Extended Data Fig. 3 | STORM analysis of paired *Drosophila* homologs.

a, Schematic of expected outcomes from STORM if chromosomes are tightly paired and follow a similar trajectory, or are loosely paired and follow distinct trajectories that are within a diffraction limited distance of one another. **b,** Representative images from STORM of paired homologs labelled with ORCA barcode 69. Similar images were observed for all barcodes shown in (c-d). Centers from 3D-Gaussian fitting of the point spread function are represented as "+" symbol on spots. Occasional double peaks occur at a frequency similar to that of background non-specific localizations and may represent stray probes, noise in peak detection of weakly-labelled loci, or split homologs. **c,** Quantification of the frequency of

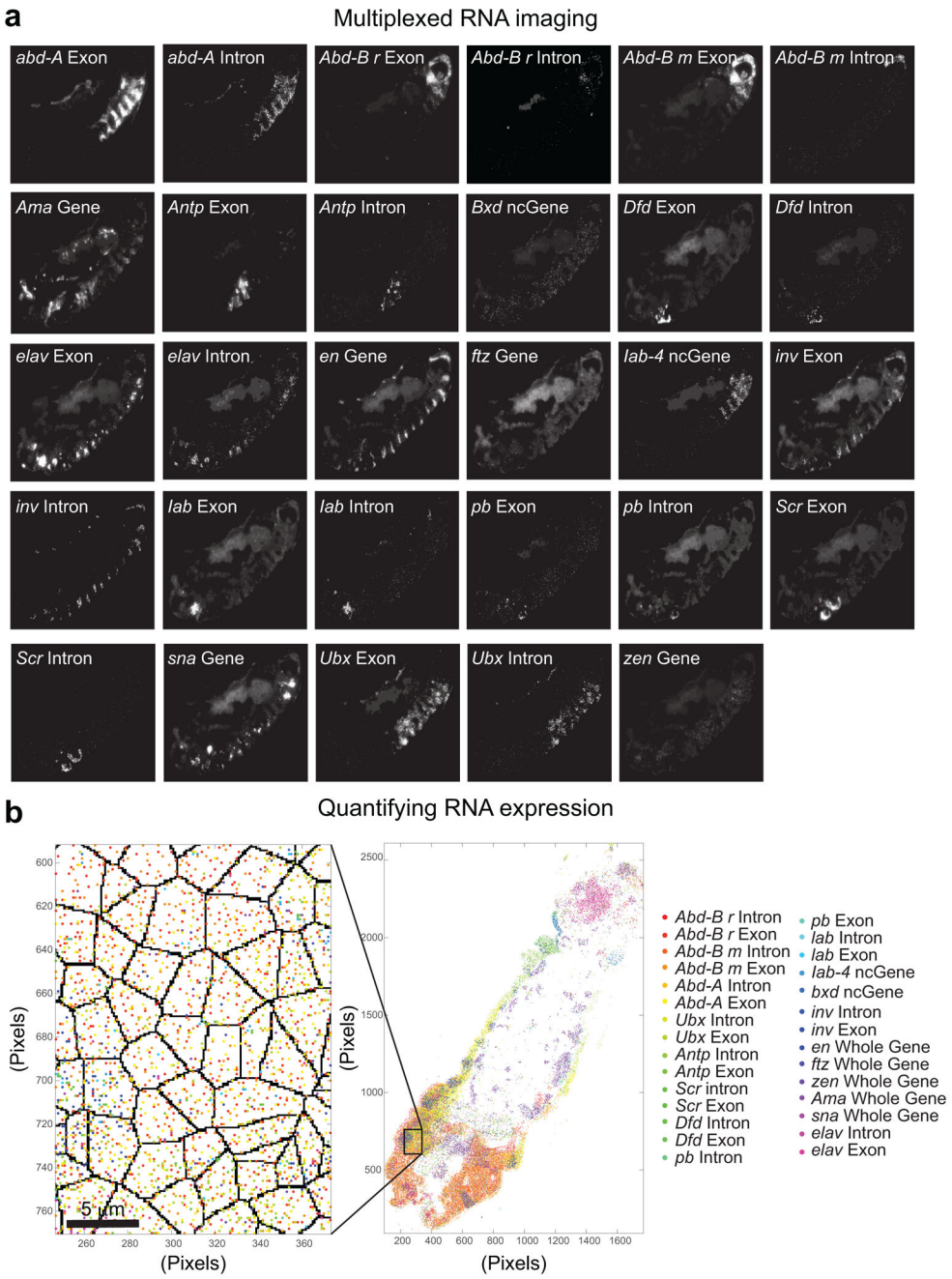
single peak detection for 10 kb per barcode and **d,** 3 kb per barcode. The high level of single peak detection is indicative of the tightly-paired homologs for the BX-C. For the boxplots in (c-d), the red line marks the median, the upper and lower limits of the box mark the interquartile range, and the whiskers extend to the furthest datum within 1.5 times the interquartile range. Number of foci (*n*) is indicated below each box plot.



Extended Data Fig. 4 | ORCA in mouse embryonic stem cells.

a, Raw image of mouse embryonic stem cells with the *Sox2* region labelled, showing two distinct homologs per cell. Similar images were observed for all 2,240 cells. **b,** Distance maps from two single cells, and the corresponding ORCA images. **c,** Average contact frequency as measured by ORCA (150 nm threshold). The location of the CTCF ChIP-seq[23] peaks, the *Sox2* gene, and the ORCA barcodes are indicated. **d,** Average contact frequency as measured by high resolution Hi-C[23]. The TAD containing *Sox2* is outlined with black lines, the weak corner point (CTCF loop) is marked with a circle. **e,** Pearson correlation between the data in (c) and (d). Pearson's correlation and Pearson's *r* in (e, g) were

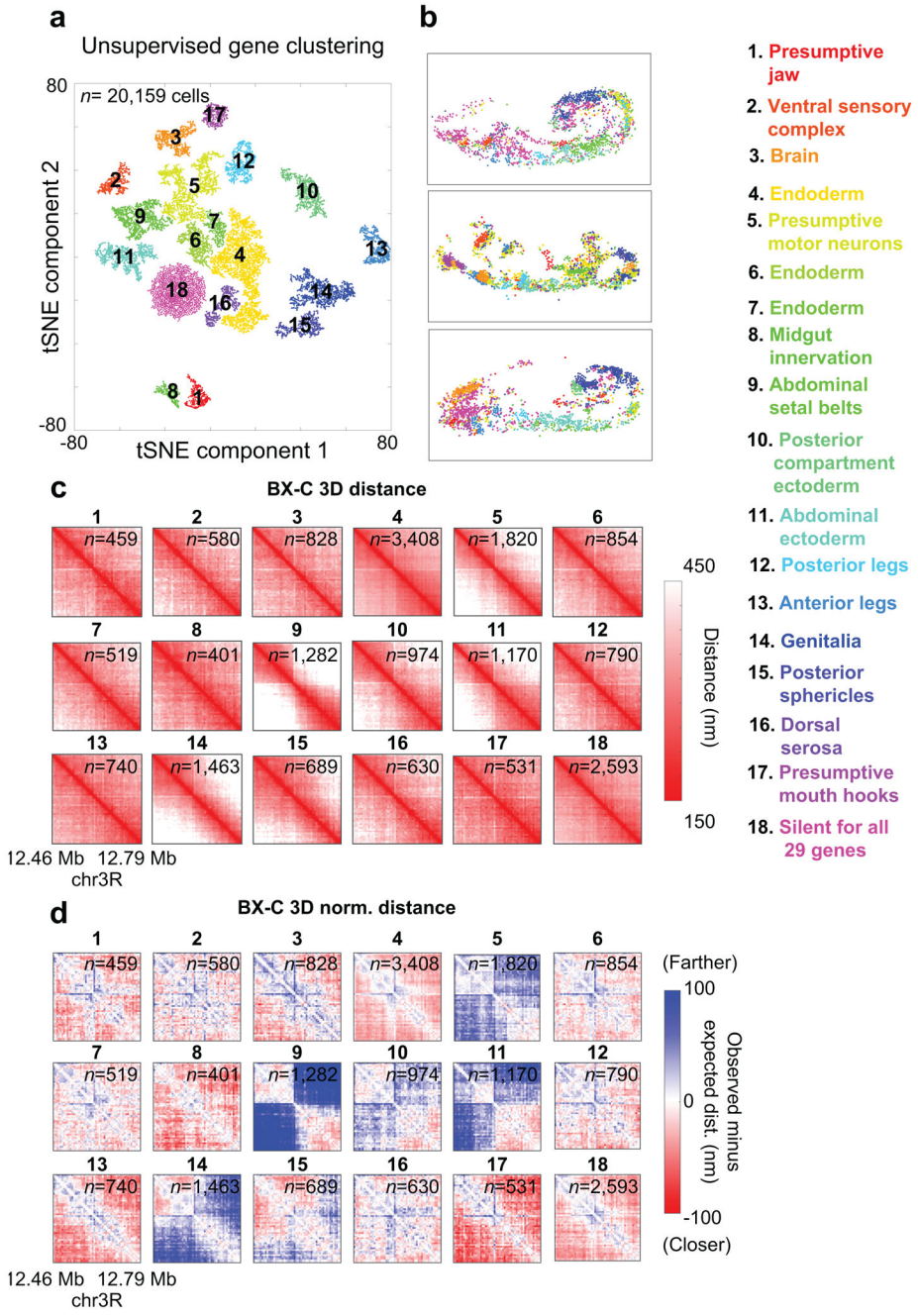
computed using all unique pairwise combinations of all 33 barcodes measured. **f,** Genome browser view of the *Sox2* region, showing the position of the 5-kb barcodes relative to the gene, its regulatory region, and local CTCF peaks. **g,** Pearson's correlation coefficients from comparing different choices of the threshold distance for contact frequency as indicated.



Extended Data Fig. 5 | Multiplex smFISH and quantification.

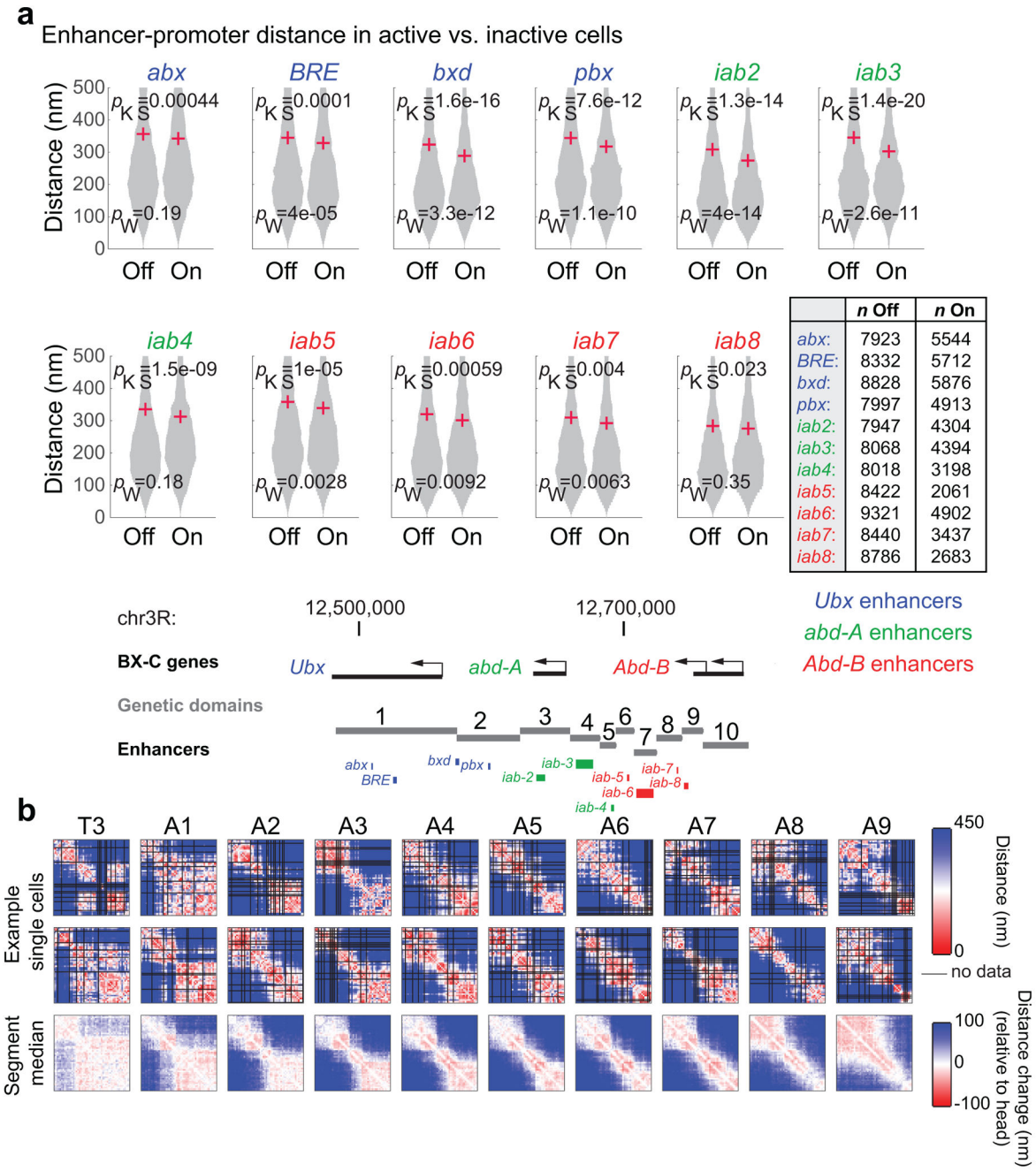
a, Sequentially acquired images from multiplex smFISH and intron-FISH for 18 cytosolic mRNAs and 11 nascent RNAs, respectively, in a laterally cryo-sectioned *Drosophila* embryos ~11 hpf. **b,** A dorsal-lateral cryosection of another embryo, ~11 hpf, showing the positions of all RNAs quantified. The zoomed-in view shows the segmented cells and a higher-resolution depiction of the RNA positions. Bright foci for each mRNA transcript (see Fig. 2a) were identified using an automated counting algorithm with a manually defined threshold. The embryo images were segmented to identify cell boundaries and each RNA transcript identified is depicted as a single spot, colour-coded by RNA species. Note, *Abd-B*

has two differentially expressed isoforms (known as *m* and *r*) which are transcribed from two different promoters. Similar results to (a-b) were obtained for 15 embryos.



Extended Data Fig. 6 | Distinct spatial organization and chromatin structure per cell type. **a,** t-SNE 2D projection of the RNA expression data, clustered with DBSCAN (see Methods). Annotations identified by manual inspection are indicated by matching colours and numbers (labelled on the right). This number/colour legend is used for all panels. **b,** The relative embryonic spatial positions of cells in each group shown for three embryos. **c,** Distance maps from ORCA for each assigned group. **d,** Normalized distance maps (observed minus expected distance) for each group. The expected distance accounts for the polymer nature of DNA, whereby sequences closer in linear position along the genome are expected to be closer together in 3D space. The expected distance was calculated by fitting a power

law to the distribution of 3D separation distances between barcodes as a function of their linear separation for all data in all cells.

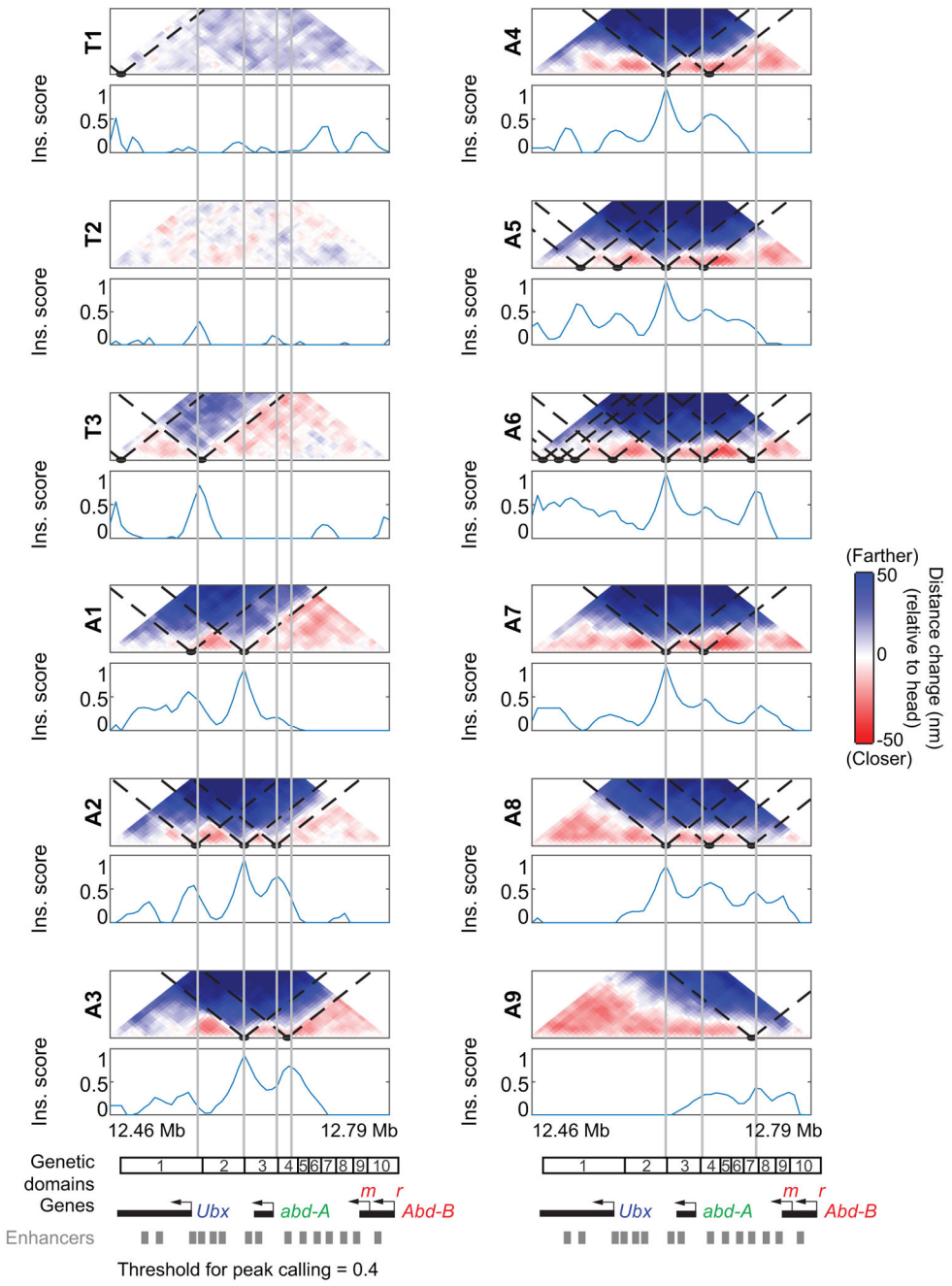


	n Off	n On
abx:	7923	5544
BRE:	8332	5712
bxd:	8828	5876
pbx:	7997	4913
iab2:	7947	4304
iab3:	8068	4394
iab4:	8018	3198
iab5:	8422	2061
iab6:	9321	4902
iab7:	8440	3437
iab8:	8786	2683

Extended Data Fig. 7 | Enhancer-promoter interactions and single cell examples.

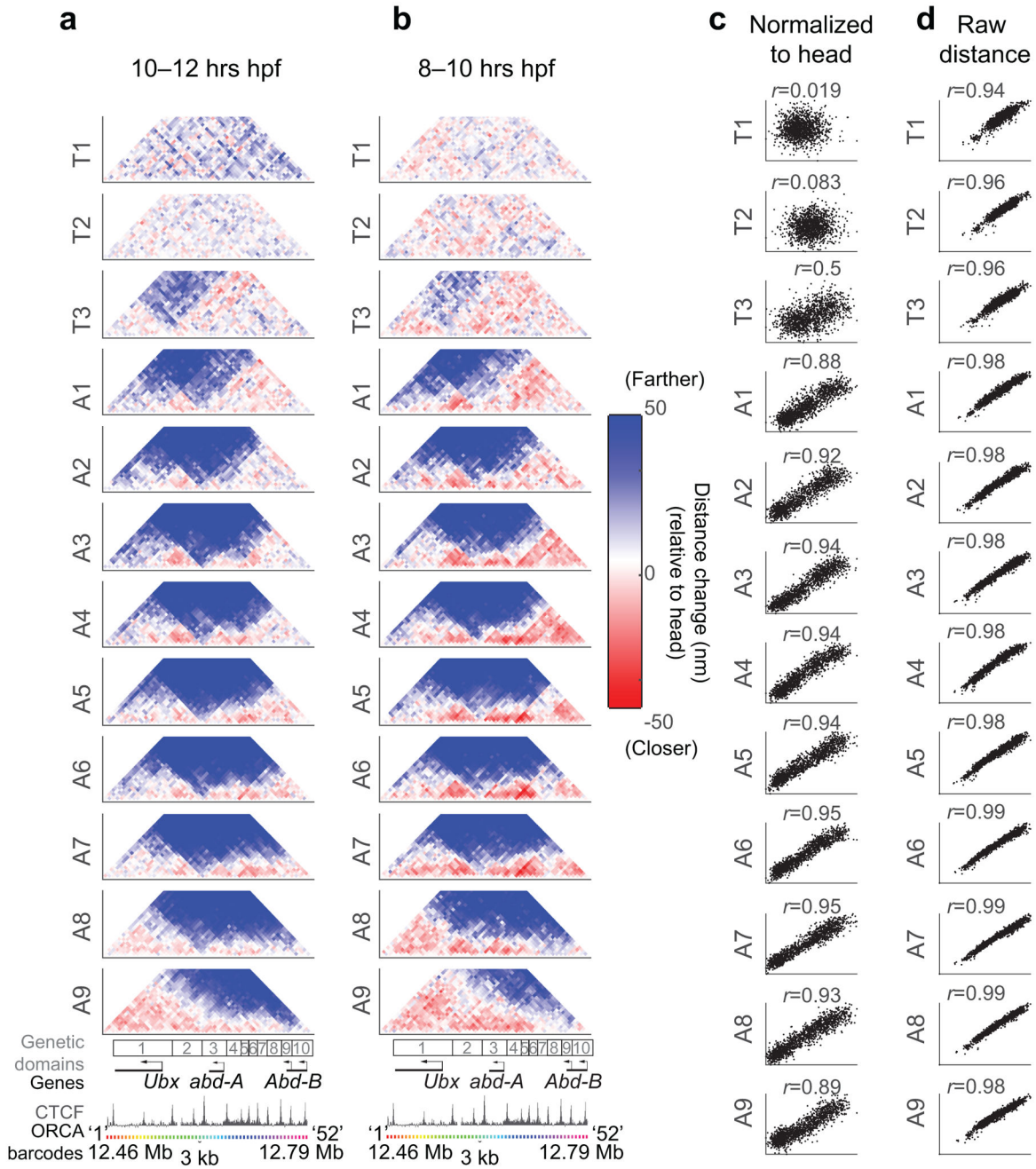
a, Violin plots showing the distance between each indicated enhancer and its cognate promoter in cells in which nascent transcripts were detected vs. absent. The average distances are marked by '+'. The *p*-values from a two-sided KS-test between the distributions (p_{KS}) and from a two-sided Wilcoxon test (p_W) are shown. The total number cells, *n*, in each distribution is indicated. The enhancers from each genetic domain (2 from each of the largest domains) are labelled by their classical names. The relative position of enhancers and promoters are shown below. Note most cells from the silent state come from anterior regions where the BX-C is repressed and compacted by Polycomb. It is interesting

to note that despite this compaction of the repressed state and relative decompaction of the active state, all enhancer-promoter pairs are on average slightly closer when active. A more substantial difference may be expected for domains in which the silent state is not largely attributable to Polycomb activity. **b,** Two example distance maps from single cells for each embryonic segment as indicated. Population-average normalized maps as in Fig. 4 are shown below for comparison.



Extended Data Fig. 8 |. Automated identification of TAD boundaries from insulation score. Relative distance maps for each segment, plotted at 3-kb resolution. TAD boundaries (dotted black lines) were identified by an automated algorithm based on computing the insulation score (plotted below each graph) and applying a threshold. A threshold score of 0.4 was used to call TAD boundaries, which agree with the boundaries detected by manual inspection and annotated as in Fig. 4. The insulation score for each segment, was computed as follows: For each barcode we first measured a score for the upstream region by computing the fraction of red bins (normalized distance <0) in a block of 6×6 bins and a similar score for the downstream block and for the interblock region. To compute the

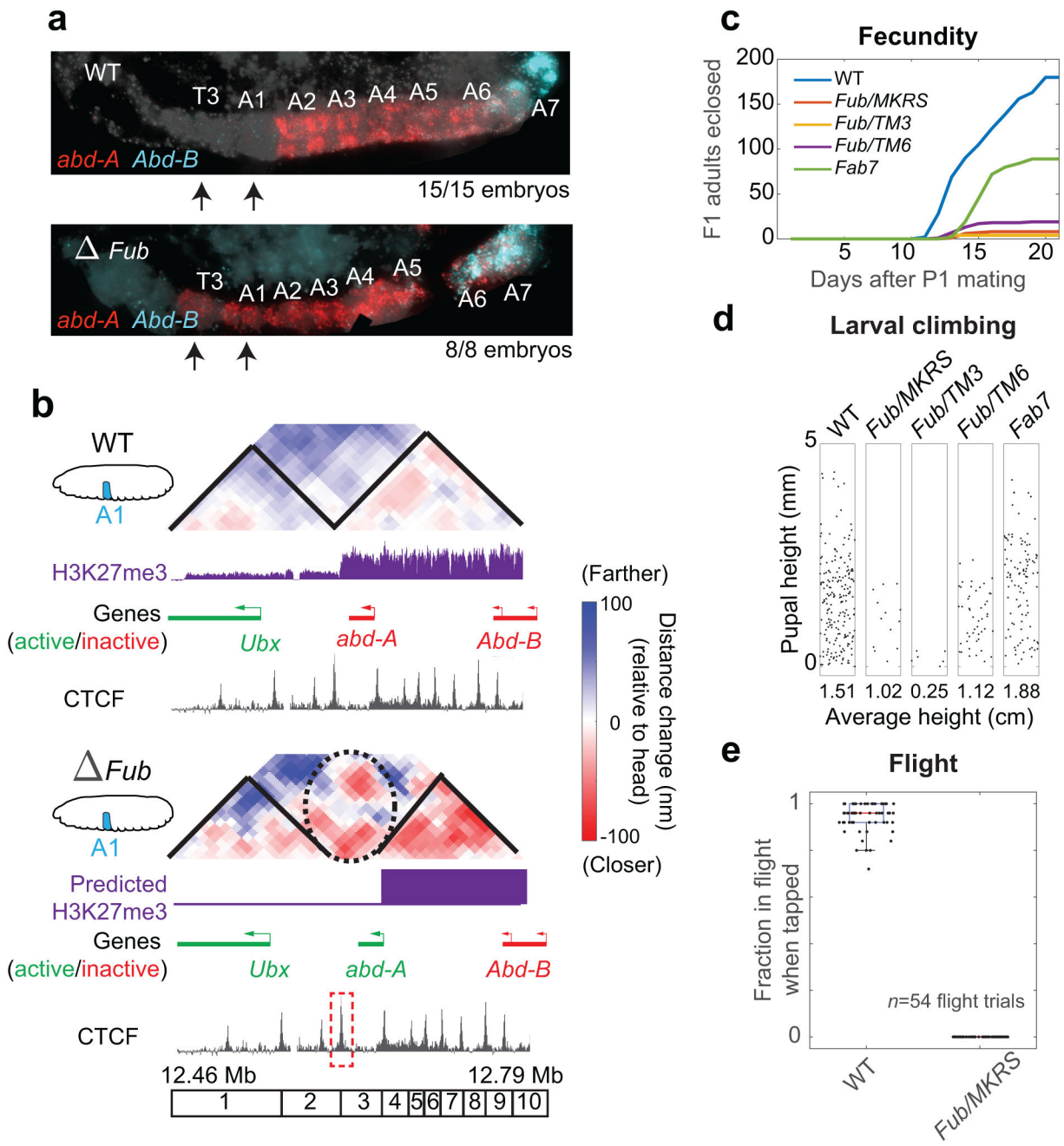
insulation score, we then took the average scores for the two blocks minus the score of the interblock.



Extended Data Fig. 9 |. Comparison of 3-kb segment-specific datasets.

a, Normalized distance maps from dataset 1, embryos 10–12 hpf. **b,** Normalized distance maps from dataset 2, embryos 8–10 hpf, showing similar patterns to dataset 1. All maps are normalized relative to cells from the head segments, as in Fig. 4. **c,** Pearson correlation between datasets 1 and 2 by segment. Note after normalization, organization in the head and anterior thorax is apparently random, and thus uncorrelated between datasets. The systematic features appearing in posterior segments, in which regulatory sequences of the BX-C are active, are similar between datasets. **d,** Pearson correlation of non-normalized

distance data. In (c) and (d), values from all unique combinations of the 52 barcodes are used in computing the correlation.



Extended Data Fig. 10 |. Properties of *Fub* mutant embryos.

a, *abd-A* and *Abd-B* expression in wild-type (WT) and *Fub* mutant embryos. Note *abd-A* expression expands to T3 in mutant compared to WT embryos **b,** ORCA data at 10-kb resolution for the 330-kb BX-C region, normalized as in Fig. 4, for cells from segment A1 WT and *Fub* mutant embryos. Predicted TADs are marked by black lines. The dotted circle over the *Fub* mutant distance map highlights aberrant interactions. The extent of H3K27me3 measured in A1[33] in WT and the predicted extent of H3K27me3 in *Fub* mutants is shown in purple. Predicted H3K27me3 is based on activation of *abd-A* observed in the mutant and previous descriptions of phenotypes of genetic border deletions in *Drosophila*[24-26]. The

region deleted in *Fub* mutants is marked by red dotted lines. **c,** Fecundity of WT, homozygous *Fab-7* and balanced *Fub* lines, plotted as the number of viable adults produced by 8 fertilized females in a 2-day window. **d,** The height that each *Drosophila* larva, represented as a dot, climbed prior to pupation. Average height is indicated on the *x*-axis. **e,** Comparison of flight ability between WT and *Fub*/*MKRS* lines, using the metric developed by Bartholomew et al.[61]. The metric reports the fraction of adult *Drosophila* which re-land on the sides of the chamber rather than falling to the bottom, after being displaced by light tapping. Each dot represents a separate flight trial. Red line represents the median, boxes mark the upper and lower quartiles, whiskers extend to the farthest datum within 1.5 times the interquartile range.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Welcome Bender for providing the *Fub* and *Fab-7* mutants, and for helpful discussions. We thank Nicholas Sinnott-Armstrong for help with assembly and programming of the fluid handling robot. We thank David Kingsley, Philip Beachy, Margaret Fuller, Anne Villeneuve, Aparna Rajpurkar, and TzuChiao Hung, for a critical reading of the manuscript. We thank Joanna Wysocka for donating the mESCs and Bo Gu for assistance with cell culture. This work was supported by the Searle Scholars Program, Burroughs Wellcome Careers at the Scientific Interface Award, Dale Frey Award, Beckman Young Investigator Program, NIH New Innovator Award DP2 (DGM132935A), and the Packard Fellows Program (A.N.B). L.J.M. is supported by the Stanford School of Medicine Dean's Fund, S.E.M. is supported by the Stanford Genome Training Program, A.H. is supported by a Walter V. and Idun Berry Postdoctoral Fellowship, I.S.C. was supported by a summer training grant from Stanford Bio-X.

References

1. Long HK, Prescott SL & Wysocka J Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell 167, 1170–1187 (2016). [PubMed: 27863239]
2. Spielmann M, Lupiáñez DG & Mundlos S Structural variation in the 3D genome. Nat. Rev. Genet. 7, 85–97 (2018).
3. Furlong EEM & Levine M Developmental enhancers and chromosome topology. Science 361, 1341–1345 (2018). [PubMed: 30262496]
4. Schwartz YB & Cavalli G Three-Dimensional Genome Organization and Function in *Drosophila*. Genetics 205, 5–24 (2017). [PubMed: 28049701]
5. Rowley MJ & Corces VG Organizational principles of 3D genome architecture. Nat. Rev. Genet. 19, 789–800 (2018). [PubMed: 30367165]
6. Chambeyron S & Bickmore WA Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. Genes Dev. 18, 1119–1130 (2004). [PubMed: 15155579]
7. Eskeland R et al. Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. Molecular cell 38, (2010).
8. Bantignies F et al. Polycomb-Dependent Regulatory Contacts between Distant Hox Loci in Drosophila. Cell 144, 214–226 (2011). [PubMed: 21241892]
9. Beliveau BJ et al. In Situ Super-Resolution Imaging of Genomic DNA with OligoSTORM and OligoDNA-PAINT. in (ed. Erfle H) 1663, 231–252 (Springer New York, 2017).
10. Williamson I, Lettice LA, Hill RE & Bickmore WA Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity. Development 143, 2994–3001 (2016). [PubMed: 27402708]

11. Cheutin T & Cavalli G Loss of PRC1 induces higher-order opening of Hox loci independently of transcription during Drosophila embryogenesis. Nat. Commun. 9, 3898 (2018). [PubMed: 30254245]
12. Cattoni DI et al. Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. Nat. Commun. 8, (2017).
13. Boettiger AN et al. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature 529, 418–422 (2016). [PubMed: 26760202]
14. Kundu S et al. Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. Mol. Cell 65, 432–446.e5 (2017). [PubMed: 28157505]
15. Fabre PJ et al. Nanoscale spatial organization of the HoxD gene cluster in distinct transcriptional states. Proc. Natl. Acad. Sci. U. S. A. 112, 13964–13969 (2015). [PubMed: 26504220]
16. Szabo Q et al. TADs are 3D structural units of higher-order chromosome organization in Drosophila. Science advances 4, eaar8082 (2018). [PubMed: 29503869]
17. Bintu B et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science 362, eaau1783 (2018). [PubMed: 30361340]
18. Wang S et al. Spatial organization of chromatin domains and compartments in single chromosomes. Science 353, 598–602 (2016). [PubMed: 27445307]
19. Cardozo Gizzi AM et al. Microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms. bioRxiv 434266 (2018). doi:10.1101/434266
20. Chen KH, Boettiger AN, Moffitt JR, Wang S & Zhuang X Spatially resolved, highly multiplexed RNA profiling in single cells. Science 348, 412 (2015).
21. Beliveau BJ et al. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. Proceedings of the National Academy of Sciences 109, 21301–21306 (2012).
22. Schuettengruber B et al. Cooperativity, Specificity, and Evolutionary Stability of Polycomb Targeting in Drosophila. Cell Rep. 9, 219–233 (2014). [PubMed: 25284790]
23. Bonev B et al. Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell 171, 557–572.e24 (2017). [PubMed: 29053968]
24. Kyrchanova O et al. The boundary paradox in the Bithorax complex. Mech. Dev. 138, 122–132 (2015). [PubMed: 26215349]
25. Maeda RK & Karch F Chapter 1 The Bithorax Complex of Drosophila. An Exceptional Hox Cluster. Curr. Top. Dev. Biol. 88, 1–33 (2009). [PubMed: 19651300]
26. Maeda RK & Karch F The open for business model of the bithorax complex in Drosophila. Chromosoma 124, 293–307 (2015). [PubMed: 26067031]
27. Rivera J, Keränen SVE, Gallo SM & Halfon MS REDfly: the transcriptional regulatory element database for Drosophila. Nucleic Acids Res. 47, D828–D834 (2018).
28. Symmons O et al. Functional and topological characteristics of mammalian regulatory domains. Genome Res. 24, 390–400 (2014). [PubMed: 24398455]
29. Symmons O et al. The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances. Dev. Cell 39, 529–543 (2016). [PubMed: 27867070]
30. Levine M, Cattoglio C & Tjian R Looping back to leap forward: transcription enters a new era. Cell 157, 13–25 (2014). [PubMed: 24679523]
31. Chen H et al. Dynamic interplay between enhancer-promoter topology and gene activity. Nat. Genet. 50, 1296–1303 (2018). [PubMed: 30038397]
32. Bender W & Hudson A P element homing to the Drosophila bithorax complex. Development 127, 3981–3992 (2000). [PubMed: 10952896]
33. Bowman SK et al. H3K27 modifications define segmental regulatory domains in the Drosophila bithorax complex. Elife 3, e02833 (2014). [PubMed: 25082344]
34. Sexton T et al. Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. Cell 148, 458–472 (2012). [PubMed: 22265598]
35. Dixon JR et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380 (2012). [PubMed: 22495300]

36. Nora EP et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485, 381–385 (2012). [PubMed: 22495304]
37. Lewis EB A gene complex controlling segmentation in Drosophila. Nature 276, 565–570 (1978). [PubMed: 103000]
38. Rowley MJ et al. Evolutionarily Conserved Principles Predict 3D Chromatin Organization. Mol. Cell 67, 837–852 (2017). [PubMed: 28826674]
39. Bender W & Lucas M The border between the Ultrabithorax and abdominal-A regulatory domains in the Drosophila bithorax complex. Genetics 193, 1135–1147 (2013). [PubMed: 23288934]
40. Gyurkovics H, Gausz J, Kummer J & Karch F A new homeotic mutation in the Drosophila bithorax complex removes a boundary separating two domains of regulation. EMBO J. 9, 2579–2585 (1990). [PubMed: 1973385]
41. Gambetta MC & Furlong EEM The Insulator Protein CTCF Is Required for Correct Hox Gene Expression, but Not for Embryonic Development in Drosophila. Genetics 210, 129–136 (2018). [PubMed: 30021792]
42. The modENCODE Consortium et al. Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science 330, 1787–1797 (2010). [PubMed: 21177974]
43. Ghavi-Helm Y et al. Enhancer loops appear stable during development and are associated with paused polymerase. Nature 512, 96–100 (2014). [PubMed: 25043061]
44. Rao SSP et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell 159, 1665–1680 (2014). [PubMed: 25497547]
45. Kragesteen BK et al. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. Nat. Genet. 50, 463–473 (2018).
46. Andrey G et al. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. Science 340, 1234167 (2013). [PubMed: 23744951]
47. Phanstiel DH et al. Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during Macrophage Development. Mol. Cell 67, 1037–1048 (2017). [PubMed: 28890333]
48. Narendra V et al. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science 347, 1017–1021 (2015). [PubMed: 25722416]
49. Narendra V, Dekker J, Mazzoni EO & Reinberg D CTCF-mediated topological boundaries during development foster appropriate gene regulation. Genes Dev. 30, 2657–2662 (2016). [PubMed: 28087711]
50. Van Bortle K et al. Insulator function and topological domain border strength scale with architectural protein occupancy. Genome Biol. 15, R82 (2014). [PubMed: 24981874]

Methods References

51. Tokunaga M, Imamoto N & Sakata-Sogawa K Highly inclined thin illumination enables clear single-molecule imaging in cells. Nat. Methods 5, 159–161 (2008). [PubMed: 18176568]
52. Moffitt JR et al. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. Proceedings of the National Academy of Sciences 113, 11046–11051 (2016).
53. Gu B et al. Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. Science 359, 1050–1055 (2018). [PubMed: 29371426]
54. Moffitt JR et al. High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. Proceedings of the National Academy of Sciences 113, 14456–14461 (2016).
55. Babcock H, Sigal YM & Zhuang X A high-density 3D localization algorithm for stochastic optical reconstruction microscopy. Optical Nanoscopy 1, 6 (2012).
56. Perry MW, Boettiger AN, Bothma JP & Levine M Shadow enhancers foster robustness of Drosophila gastrulation. Curr. Biol. 20, 1562–1567 (2010). [PubMed: 20797865]
57. Briggs JA et al. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. Science 360, eaar5780 (2018). [PubMed: 29700227]

58. Van Der Maaten LJP & Hinton GE Visualizing high-dimensional data using t-sne. J. Mach. Learn. Res. 9, 2579–2605 (2008).

59. Ester M, Kriegel H-P, Sander J & Xu X A Density-Based Algorithm for Discovering Clusters. KDD-96 Proceedings 226–231 (1996).

60. Maeda RK & Karch F The ABC of the BX-C: the bithorax complex explained. Development 133, 1413–1422 (2006). [PubMed: 16556913]

61. Bartholomew NR, Burdett JM, Vandenbrooks JM, Quinlan MC & Call GB Impaired climbing and flight behaviour in Drosophila melanogaster following carbon dioxide anaesthesia. Sci. Rep. 5, 15298 (2015). [PubMed: 26477397]

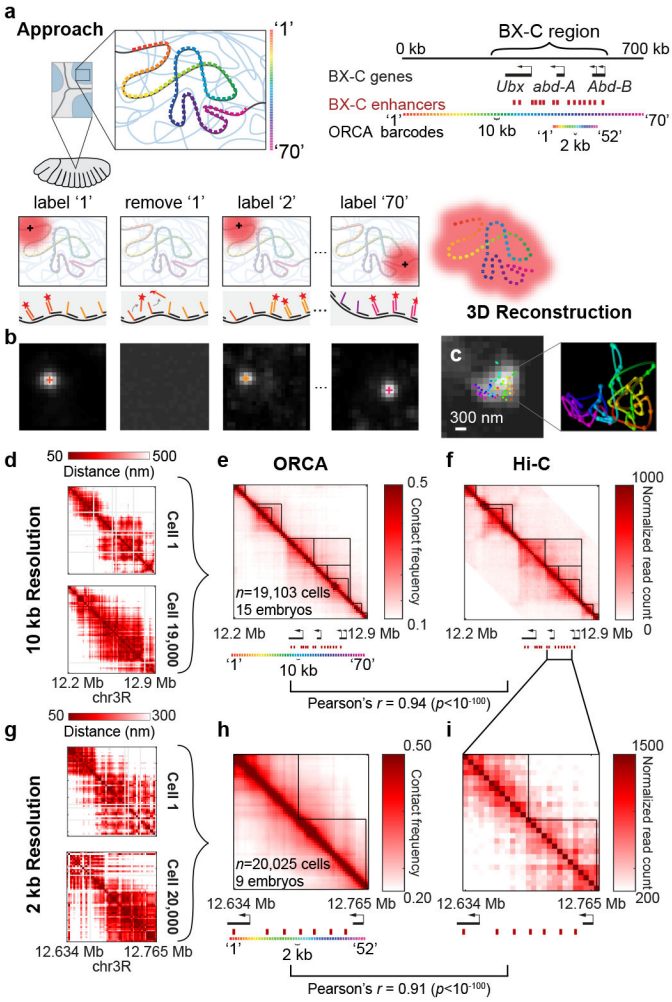


Figure 1 |. Optical reconstruction of chromatin architecture (ORCA).

a, A region of interest is labelled with primary probes, partitioning the region into 70 segments with unique barcodes. Each barcode is imaged by sequentially introducing a complementary readout oligo carrying a fluorophore, which is removed after imaging, and the process repeats. **b,** Example data from imaged barcodes. Centers from 3D-Gaussian fitting of the point spread function are represented as "+" on spots. **c,** Diffraction-limited image of the entire domain, overlaid with coloured spots indicating the positions of each barcode. Zoomed-in view shows the same spots connected in order of genomic position (ORCA image). Similar images were collected for all 19,103 cells analysed in (e). **d,** Maps from two individual cells from wild-type embryos 10-12 hpf, showing pairwise distances between all barcodes that traced a 700-kb region (chr3R:12.20-12.90 Mb (dm3)) at 10-kb resolution (distance maps). **e,** Frequency across all cells in the embryo with which any two barcodes were found in contact (separation <150 nm). TADs are marked with black lines. **f,** Previously published Hi-C[22] from wild-type embryos 0-12 hpf plotted at 10-kb resolution. **g,** Distance maps from two individual cells, constructed from ORCA of the region: chr3R: 12.634–12.765 Mb (dm3) at 2-kb resolution. **h,** Population-level contact frequency map. **i,**

Hi-C data from (f), re-plotted at 5-kb resolution[22]. Pearson's *r* was computed using all unique pairwise combinations for the barcodes shown.

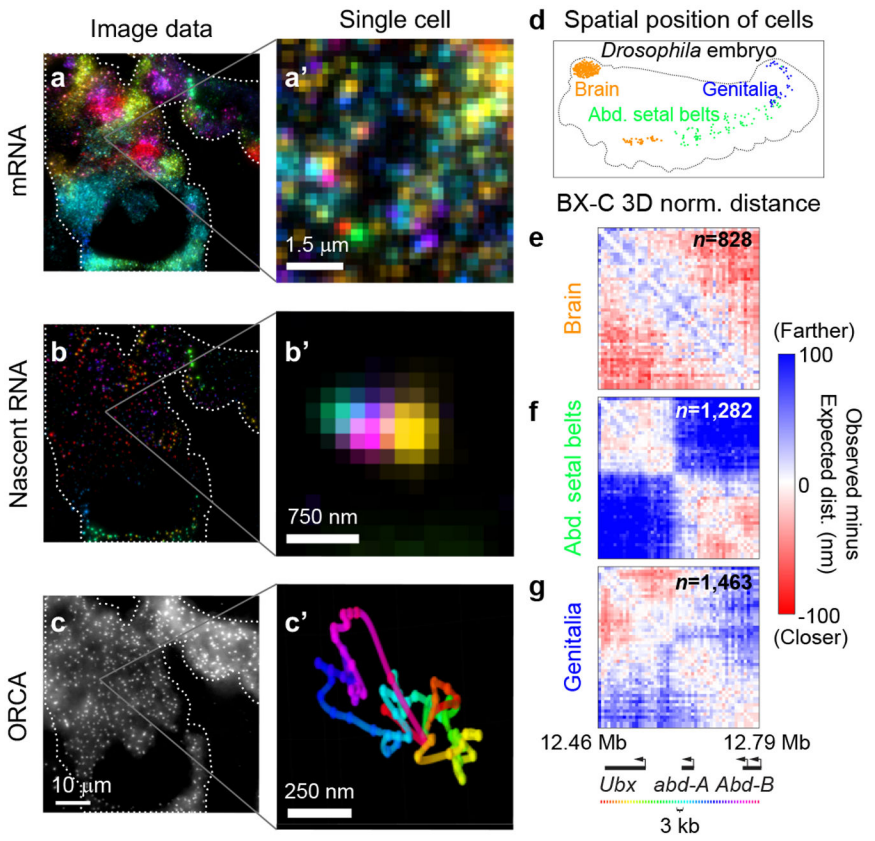


Figure 2 |. Multiplexed smFISH combined with ORCA identifies cell-type-specific chromatin folding.

Cryosectioned embryos 10–12 hpf, labelled with **a,** 18 mRNAs (8 shown), **b,** 11 nascent RNAs (for all RNAs, see Extended Data Fig. 5), and **c,** ORCA for the 330-kb BX-C. **a'-c',** Corresponding zoomed-in views at the scale of a single cell. Similar images were observed for all 20,159 cells analysed in (**d–g**). **d,** Spatial organization of cells assigned to three clusters within one embryo (for all clusters, see Extended Data Fig. 6b). **e–g,** Normalized population-median distance maps for the clusters shown in (d), (for all clusters, see Extended Data Fig. 6d).

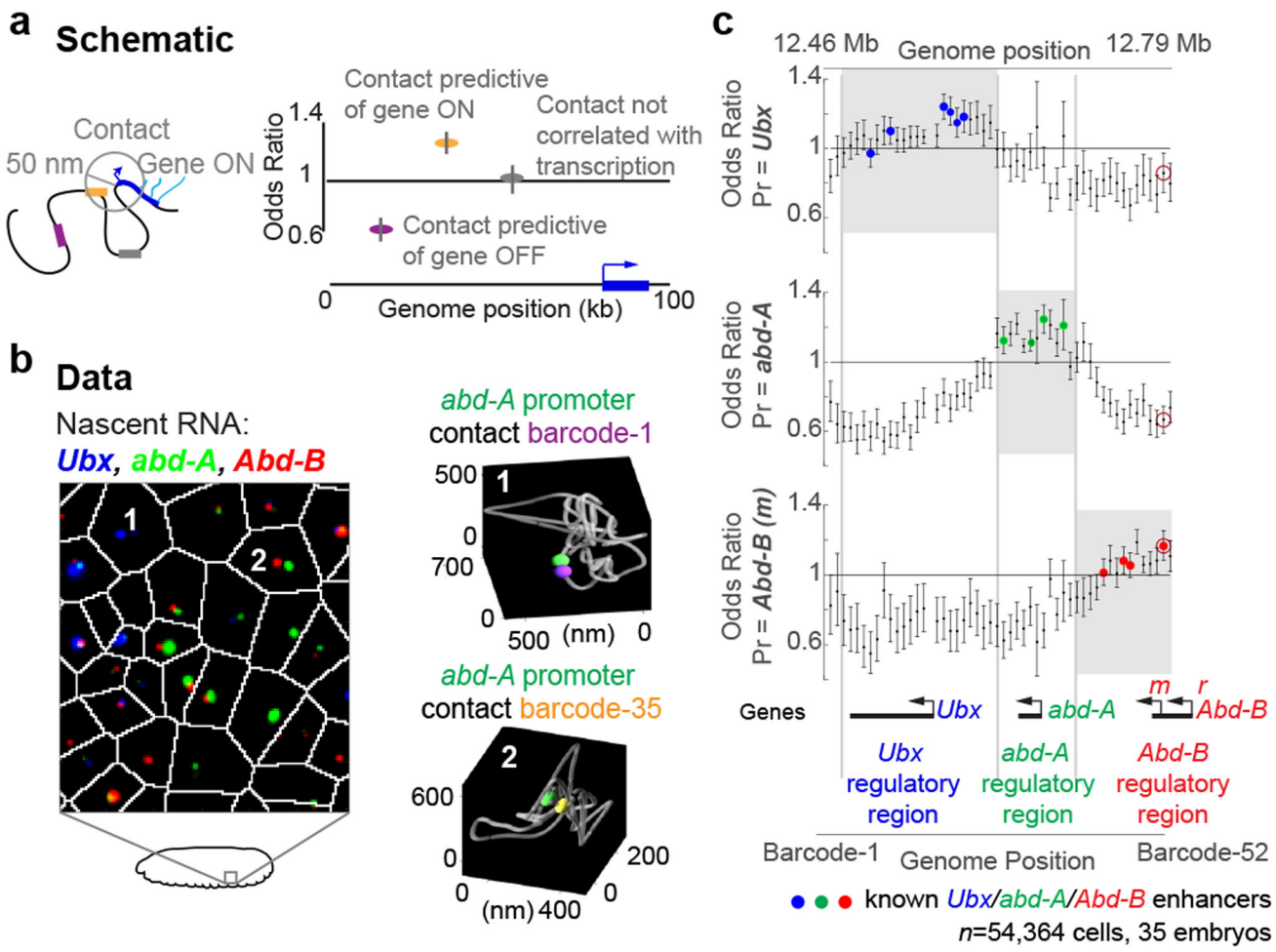


Figure 3 |. Connecting RNA expression and genome structure in single cells.

a, Schematic for inferring *cis*-regulatory activity from pairwise contacts and nascent RNA. DNA sequences located <150 nm from the promoter were considered "in contact". The Odds Ratio (OR) was used to infer which of these contacts are likely to induce transcription (OR > 1, as expected for enhancers, orange), to repress transcription (OR < 1, as expected for silencers, purple) or to have no contribution to transcription (OR statistically indistinct from 1, as expected for non-regulatory sequences, grey). OR is plotted as a function of genomic distance. **b,** Example image from a posterior portion of an embryo showing nascent RNA for all three BX-C genes (left), and ORCA images (right) showing promoter contacts with barcoded sequences. Matching cells are indicated. Similar images were observed for the 54,364 cells in the analysis. **c,** Top plot, the OR (*y*-axis) for *Ubx* nascent transcriptional activity as a function of the genomic position/barcode (*x*-axis) with which the *Ubx* promoter made contact. Similar plots are shown for the *abd-A* and *Abd-B* promoters. Known enhancers are marked by coloured dots. The last *Abd-B* enhancer is labelled with a hollow red circle for all three plots. Known regulatory domains (in grey) are indicated below. Error bars represent 95% confidence intervals.

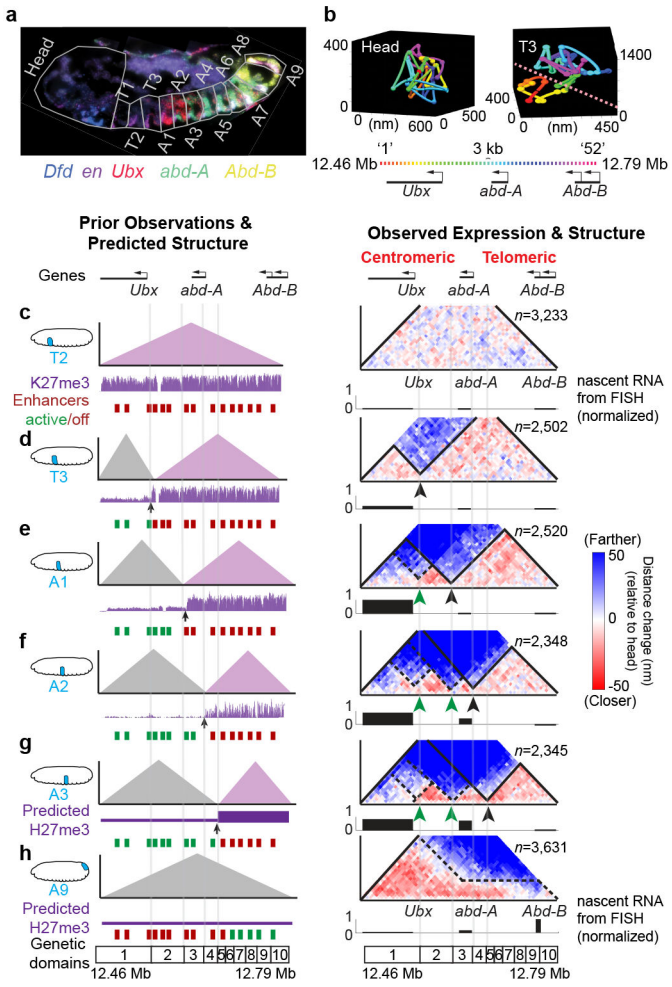


Figure 4 |. ORCA uncovers cell-type-specific domain structures.

a, *Drosophila* embryo with body segments manually annotated (see Methods). mRNAs are shown for *en* (purple), *Dfd* (blue), *Ubx* (red), *abd-A* (green) and *Abd-B* (yellow). **b,** ORCA images of the BX-C from a cell in the head and a cell in T3, colour-coded by barcode as shown below. Both replicate datasets produced similar images. Dashed line marks the H3K27me3 border determined by ChIP-seq[33]. **c–h,** Predicted (left column) and ORCA measured (right column) TAD organization for the indicated body segments. Predicted TAD borders are marked by black lines, unexpected TAD borders are indicated by dashed lines. A black arrow (left column) indicates the measured[33] or predicted[24,26] H3K27me3 border and predicted TAD boundary. A black arrowhead (right column) marks the corresponding boundary position below the ORCA data. Green arrowheads in (e–g) mark unexpected TAD boundaries. Vertical grey lines follow the genetic domain borders previously identified[24-26]. Average nascent RNA intensity for each BX-C gene is indicated by the height of the bar in the plot below the ORCA data (right column). Values are normalized to the highest-expressing segment in the embryo. The position and length of the bar matches the extent of the transcription unit. The number of cells, *n*, for each map is indicated and aggregated from 35 embryos.

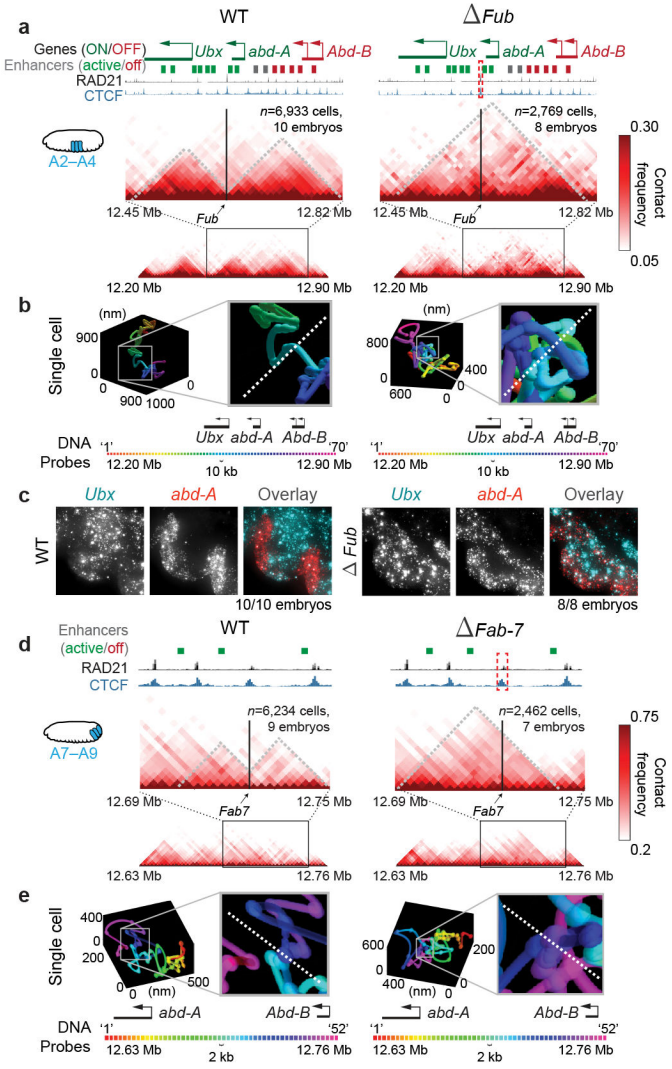


Figure 5 | Boundary deletions fuse TADs and lead to enhancer crosstalk.

a, Pairwise contact frequency computed from ORCA for A2–A4 in wild-type embryos (left) and *Fub* mutants (right) at 10-kb resolution (contact threshold 150 nm). Dotted grey lines mark the TADs. BX-C genes, CTCF ChIP-seq from wild-type embryos[42], and RAD21 from a wild-type embryonic cell line[50], are shown below. Red dotted lines mark the region deleted in *Fub* mutants. **b,** ORCA images of a 700-kb domain spanning the BX-C, from one of the 6,933 wild-type cells from a single experiment and one of the 2,769 *Fub* mutant cells from a single experiment. Barcode position is indicated by colour. In the zoomed-in views, a dotted line demarcates the 3D separation of regions upstream and downstream of the *Fub* locus. **c,** Comparison of *abd-A* and *Ubx* expression from wild-type and *Fub* mutant embryos within individual segments. **d,** Pairwise contact frequency computed from ORCA for A7–A9 comparing wild-type and *Fab-7* mutants for the region around the *Fab-7* locus at 2-kb resolution. **e,** ORCA images from one of 6,234 wild-type cells collected over two experiments and one of the 2,462 mutant cells collected over two experiments. In the

zoomed-in images, a dotted line demarcates the 3D separation between regions upstream and downstream of the *Fab-7* locus.