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

Leslie J. Mateo1, Sedona E. Murphy1,2, Antonina Hafner1, Isaac S. Cinquini1–3, Carly A. Walker1 & Alistair N. Boettiger1*

The establishment of cell types during development requires precise interactions between genes and distal regulatory sequences. We have a limited understanding of how these interactions look in three dimensions, vary across cell types in complex tissue, and relate to transcription. Here we describe optical reconstruction of chromatin architecture (ORCA), a method that can trace the DNA path in single cells with nanoscale accuracy and genomic resolution reaching two kilobases. We used ORCA to study a Hox gene cluster in cryosectioned Drosophila embryos and labelled around 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, identify domain structures that change with cell identity, and show that border elements contribute to the 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 3D folding of the genome. Consequently, examining the sub-megabase-scale folding of developmental control loci and the extent to which it varies among cell types is essential for understanding embryogenesis1–3.

Chromatin conformation capture (3C) approaches have mapped cis-interactions de novo and revealed partitioning of the genome into topological associated domains (TADs), in which intra-domain contacts are enriched over inter-domain contacts. TADs frequently span developmental genes and their cis-control regions4–5. However, bulk approaches such as 3C lack the ability to distinguish subpopulations 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 organization6. Microscopy approaches can detect sub-populations within tissues de novo and resolve large numbers of individual cells. Accordingly, multicolour fluorescent in situ hybridization (FISH)6–12, oligo-stochastic optical reconstruction microscopy (oligo-STORM)9,13–18, and sequential FISH17–20 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 kilobases, which is an important length scale for cis-regulation.

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 that preferentially interact with promoters de novo. Like microscopy methods, the approach should be able to detect subpopulations of cells that have 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 use it to test several predictions about chromatin structure and cis-regulation in the bithorax complex (BX-C) in 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)9,13,19,21,22. 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 barcodes21 (Fig. 1a, Extended Data Fig. 1a–c). Supplementary Tables 1–5). These barcodes are labelled with a fluorophore and imaged. The signal is then removed via strand displacement (Supplementary Table 6), and the process repeats for each barcode. This is conceptually similar to recent19 and concurrent work17,18,20, although it has improved genomic resolution (Fig. 1b). With high-precision fiducial registration (Extended Data Fig. 1a–c), sequential imaging allows barcoded sections within a diffusion-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 used 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 h 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–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-C23 across a
range of contact thresholds revealed qualitatively similar features and quantitatively similar contact distributions (Fig. 1e–i). In Drosophila chromosomes, which are predominantly paired in interphase, additional ORCA images acquired with STORM revealed that paired homologues follow a common trajectory to within about 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) showed strong correspondence with published Hi-C data (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 more than 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 the spatial organization of the BX-C, 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, we normalized these data for genomic-distance effects, which further accentuated cell-type differences (Fig. 2e–g, Extended Data Fig. 6d).

In anterior cells (for example, ‘brain’), the BX-C adopted an organization in which long-range interactions were 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 and the known Pc-repressed status of the BX-C in the anterior cells. The farther-than-expected short-range interactions were more surprising, though consistent with recent conclusions from STORM imaging of Pc-repressed DNA. 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 DNA structure.

Cis contacts predict transcription

We investigated whether 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 (OR), of nascent transcription from a promoter if it was in close proximity to a genomic position or barcode with which the Ubx promoter made contact. Similar plots are shown for the abd-A and Abd-B promoters (middle, bottom). Known enhancers are marked by coloured dots. The last Abd-B enhancer is labelled with an open red circle in all three plots. Known regulatory domains (grey) are indicated below. Error bars represent 95% confidence intervals.

Segment-specific TAD boundaries

We next investigated how the cell-type-specific structural differences in 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.\(^{25,27,34}\) 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.\(^{35–37}\)

To test for segment-specific folding of the BX-C, we manually re-sorted the single cells from Figs. 2, 3 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 to the H3K27me3-repressed region previously measured in T3\(^{34}\) and the centromeric TAD corresponded to the de-repressed region around Ubx (Fig. 4d).

In A1–A2, as the observed H3K27me3 border\(^{34}\) 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\(^{25,27}\). 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.\(^{27,38}\)

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.\(^{27}\) Each of these ten genetic domains (labelled at the bottom of Fig. 4c–h) contains enhancers that are critical for specifying the fate of the segment in which they first become open.\(^{25–27}\) 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.\(^{34}\) ORCA data confirm that this epigenetic partitioning is matched by spatial partitioning and extend the observations into more posterior segments than were previously accessible (Fig. 4c–g). The agreement between the 3D boundaries measured by ORCA, the epigenetic domains measured by chromatin immunoprecipitation with sequencing (ChIP-seq)\(^{34}\), and the genetic domains measured in classic experiments\(^{25–27,38}\), further validates our approach.

Pc-independent TAD boundaries in BX-C

Contrary to earlier predictions,\(^{39}\) we found that not all TAD boundaries in the BX-C coincided with epigenetic domain borders, and therefore not all could be explained by Pc activity (Fig. 4c–h, right). In A1, the de-repressed portion of the BX-C was split into two TADs (Fig. 4e, Extended Data Fig. 7b). In A2, an additional TAD boundary formed in the H3K27me3-free region (Fig. 4f). In A9, where no H3K27me3 boundaries were predicted,\(^{25,27}\) 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 Figs. 7b, 8). This centromeric TAD in A9 (Fig. 4h) is reminiscent of the Pc-dependent compaction in anterior segments, although Pc is not believed to contribute to silencing of BX-C in A9.\(^{25–27}\)

Together, these unexpected TAD boundaries offer an explanation for segment-specific enhancer activity (Fig. 4c–h, left) and gene expression (Fig. 4c–h, right). For example, in A2, the TAD over genetic domain 2 links the Ubx enhancers with the Ubx promoter and a separate TAD over...
Boundary elements separate active TADs

Because epigenetic domain borders did not explain all of the observed BX-C spatial boundaries, we investigated whether 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, so epigenetic differences cannot account for the presence of the TAD border (Fig. 5a). We examined embryos that were homozygous for a 4-kb deletion spanning this border. In these mutants, we observed fusion of the TADs flanking the deletion, with increased contact frequency extending over 70 kb upstream and downstream of the deletion (contact threshold 150 nm). Dotted green lines mark TADs, BX-C genes, CTCF ChIP-seq from wild-type embryos, and RAD21 from a wild-type 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).

Fig. 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. c–h, Predicted (left) and ORCA measured (right) TAD organization for the indicated body segments. Predicted TAD borders are marked by black lines and unexpected TAD borders are indicated by dashed lines. Black arrows (left) indicate measured or predicted H3K27me3 borders and predicted TAD boundaries. Black arrowheads (right) mark 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. 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). 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.

Fig. 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 Fab-7 mutants (right) at 10-kb resolution (contact threshold 150 nm). Dotted green lines mark TADs, BX-C genes, CTCF ChIP-seq from wild-type embryos, and RAD21 from a wild-type embryonic cell line. Comparison of abd-A and Ubx expression from wild-type and Fab-7 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, the dotted lines demarcate the 3D separation between regions upstream and downstream of the Fab-7 locus.
This 4-kb deletion was previously characterized on the basis of its abnormal A1 morphology, and accordingly named *Front-ultraabdominal (Fab)* [46]. This morphology was attributed to ectopic *abd-A* expression 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 [46].

To further test whether elements at TAD boundaries identified by ORCA contribute to TAD separation, we examined mutants carrying a roughly 3-kb deletion [41] spanning another TAD border detected between genetic domains 7 and 8 in wild-type cells from A7–A9 (Fig. 5d). Notably, both domains were in the same epigenetic state in these cells, so 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 *Front-ultraabdominal-7 (Fab-7)* [41], and while homozygous mutants were fertile, they exhibited a reduction in fecundity and lifespan relative to the wild type (Extended Data Fig. 10c, d).

It has 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 [39]. We observed that: (1) TAD boundaries formed within regions of the same epigenetic state; (2) these boundaries were marked by cohesin and CTCF (among other proteins, such as CP190, which may contribute to boundary element function [42,43]); and (3) deletion of these boundary elements led to 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 [39].

ORCA, together with two recent reports [17,20], 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. Owing to the high resolution of ORCA, we resolved enhancer–promoter contacts on the kilobase-scale that is 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. Whereas previous results have suggested that genome structure is largely static during development and across cell types [36,44,45], our results support an emerging view that for developmental control, cell differentiation is accompanied by extensive 3D remodelling of chromatin structure [24,46–50]. We anticipate that the properties of ORCA will enable exploration of uncharted genomic regulatory landscapes.

**Online content**

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-1035-4.

Received: 1 June 2018; Accepted: 25 February 2019; Published online 18 March 2019.
42. Gambetta, M. C. & Furlong, E. E. M. The insulator protein CTCF is required for correct Hox gene expression, but not for embryonic development in Drosophila. Genetics 210, 129–136 (2018).
43. The modENCODE Consortium et al. Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science 330, 1787–1797 (2010).
44. Ghavi-Helm, Y. et al. Enhancer loops appear stable during development and are associated with paused polymerase. Nature 512, 96–100 (2014).
45. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680 (2014).
46. Kraggesteen, B. K. et al. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. Nat. Genet. 50, 1463–1473 (2018).
47. Andrey, G. et al. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. Science 340, 1234167 (2013).
48. Phanstiel, D. H. et al. Static and dynamic DNA loops form AP-1-bound activation hubs during macrophage development. Mol. Cell 67, 1037–1048 (2017).
49. Narendra, V. et al. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science 347, 1017–1021 (2015).
50. Narendra, V., Bulajić, M., Dekker, J., Mazzoni, E. O. & Reinberg, D. CTCF-mediated topological boundaries during development foster appropriate gene regulation. Genes Dev. 30, 2657–2662 (2016).
51. Van Bortle, K. et al. Insulator function and topological domain border strength scale with architectural protein occupancy. Genome Biol. 15, R82 (2014).

Acknowledgements We thank W. Bender for providing the Fub and Fab-7 mutants, and for discussions; N. Sinnott-Armstrong for help with assembly and programming of the fluid handling robot; D. Kingsley, P. Beachy, M. Fuller, A. Villeneuve, A. Rajpurkar, and T.C. Hung for critical reading of the manuscript; J. Wysocka for the mESCs; and B. 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 (GGM132935A), 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, and I.S.C. was supported by a summer training grant from Stanford Bio-X.

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.

Competing interests The authors declare no competing interests.

Additional information Extended data is available for this paper at https://doi.org/10.1038/s41586-019-1035-4.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1035-4.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to A.N.B.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019
METHODS
No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Microscope setup. All imaging experiments were performed on one of two custom-built microscope platforms optimized for HILO illumination12 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 96A620) on a Nikon Ti-U or Ti2 body. Samples were imaged using a 60 × 1.4 NA objective (MRD16050) and a Hamamatsu sCMOS camera factory-calibrated for single-molecule imaging (ORCA-Flash4.0, C11440-22CUPL-US-KIT). The system with the Ti2 body had a larger aperture size and used a 1,536 × 1,536 (pixels) field of view (see ‘ORCA imaging’). The Ti-U system used a 1,024 × 1,0244 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, D03-R561-25SD, D03-R635-25SD), power modulated using an acousto-optic tunable filter (AOTF) (Gooch and Housego 97-0309-01), and directed to the sample using a custom polyochromatic mirror (Chroma z405/488/561/647/752m). Emitted and scattered light were further filtered with a custom emission filter (Chroma zet405/488/561/674/656/752m) before being captured 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 described13, 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 absorbive IR filter to remove stray reflections, which would otherwise otherwise 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 H. Babcock and collaborators, with modifications for our hardware configuration as appropriate. Links to the open-source, publicly accessible branch of this software are included in the ‘Code availability’ section.

Automated fluid handling setup. Fluid handling steps between each imaging round (from fixation to hybridization imaging) were performed using a robotic fluid handling setup. First, 200–500 µl probe solution (25% ethylene carbonate and 2 × saline–sodium citrate buffer (SSC) containing ‘readout’ oligo (0.1 nM)) was washed over the sample with a slow wash with 1 ml of rinse buffer (2 × SSC). Imaging buffer was stored under a 1–2 mH2O fluid pressure.

Probe synthesis. Probe synthesis was performed as previously described13. In brief, Oligopaint probes from the oligopool (CustomArray) were amplified from the array-synthesized library via limited-cycle PCR. All primers used for PCR were generated 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, E20405, 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 then 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 use.

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 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 EDTA in 1 × 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 the vial, in the methanol layer. The heptane layer was then removed, and the embryos were transferred to 1.5-ml microcentrifuge tubes, subjected to three short washes in 100% methanol, and stored in methanol at –20 °C or immediately embedded for cryosectioning.

The probe used in the solution for labeling with primary probes. The methanol was removed from the embryo storage tubes and the embryos were washed in 1:1 methanol:1 × PBS for 5 min while nutating. Three short washes with 1 × PBS were used to remove all remaining methanol and to rehydrate the tissue. Samples were
incubated in 30% sucrose in 1× 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 × 15-mm × 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 chromium potassium sulphate dodecahydrate in 300 ml deionized water. Embodied embryos were cryosectioned at a thickness of 6 μm and collected onto coverslip covers. The whole embryo was sectioned, and the middle section of the sample was chosen for hybridization owing to the clearly visible morphology of the segments along the A–P axis. Coverslips were either stored at −80°C or were used directly for primary probe hybridization.

Mouse embryonic stem cell culture and preparation. Mouse embryonic stem cells, line R1, were cultured in standard conditions as previously described28. 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 laboratory. The cell line was authenticated by colleagues in the Wysocka laboratory by karyotyping, RNA sequencing, 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 before our experiments.

Primary probe hybridization. Hybridization procedures were carried out at room temperature except where noted. For experiments in which both RNA and DNA were labelled 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 croslinked to the coverslip for 10 min with 4% PFA in 1× PBS and then washed three times in 1× PBS. Embryos were then permeabilized for 10 min with 0.5% Triton-X in 1× PBS and washed three times with 1× PBS. Embryos were then incubated for 35 min in hybridization solution (2× SSC + 50% vol/vol formamide and 0.1% Tween). The hybridization solution was aspirated, and 30 μl of probe hybridization buffer (50% formamide, 2× 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 radiation, 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 less than 1 h apart being essentially indistinguishable from those measured at the end of a 96-h experiment (Extended Data Fig. 1c).

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, 2× 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 work29,30, allowing the analysis of more cells. In contrast to previous work using cleavage of a fluorescent dye31,32, strand displacement also allows 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 were 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 and error introduced by any residual drift, uncertainty in our sub-diffraction measurements, sample movement or degradation, and/or pixel noise in our camera, 6–20 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 in a similar way 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× PBS, 4% PFA, 10% methanol to promote photo-switching (Supplementary Table 6). STORM images of each barcode were collected onto 1,024 × 1,024 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-DaSTORM as previously described26. 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 centred on each barcode position were recorded. The fraction of localizations containing only a single peak was plotted as a conservative estimate of the number of homologues paired more tightly than the resolution of STORM (Extended Data Fig. 3). To reduce background calls, pixels with fewer than four localizations were excluded. This is a conservative estimate of background, as individual Cy5 fluorophores typically exhibit multiple switching cycles before photobleaching and labelled homologues 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 before 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 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 auto-correlation 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 before 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 described37. 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 OR 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 all nascent RNA values were normalized 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 methods46, the expression matrix of all cells from all embryos for all RNAs assayed was projected into a 2D expression space using t-distributed stochastic neighbour embedding (t-SNE) with a z-score normalized Euclidean distance metric47 in Matlab (R2017a). The resulting clusters were separated using density-based clustering (DBSCAN)48 (Matlab implementation49). 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 on the basis of 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, which are named for the corresponding segment names, not by parasegment numbers. This allows for greater consistency, as the genetic domains that control each of these parasegments are also named for the corresponding segment (for example, the genetic domain known as lab-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 literature42, 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. Fub mutants (gift from W. Bender), were maintained over the MKRS balance. Their progeny included heterozygous and homzygous Fub embryos, which were distinguished by their RNA expression patterns and reported ectopic activation of abd-A expression in segment A1 in Fub/’Fub’ mutants50. Through multiplex RNA labelling, homozygous mutants were identified as those with aberrant abd-A mRNA and nascent abd-A expression compared to heterozygotes, which 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 Fub-7 experiments (Fig. 5d, e) as these mutants were homozygous viable.

Quantification of DNA structure. Analysis of ORCA data began with correction of 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 embryo 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, and z, respectively) labelled 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 and 12 hpf, over 90% of cells exhibited paired homologues, 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 (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 pairs from the same region (Extended Data Fig. 4). 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 Figs. 3, 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 (Fig. 2, 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 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). To measure the effect of missing data, missing data were simulated by randomly removing 50% of detected barcodes. The down-sampled data were highly correlated with the original data for each body-segment analysed (Pearson’s $r > 0.92$), indicating that our data are sufficiently oversampled rather than undersampled (Extended Data Fig. 2f, g).

**Normalization of distance effects.** To accentuate cell-type specific differences in the data, in Fig. 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 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 detection efficiency per barcode between these cell types for any of our probesets (Extended Data Fig. 2).

**Odds ratio.** The OR for transcription, given contact with a given barcoded section of the DNA, as described in Fig. 3 was computed using the standard definition: $OR = (ad)/(bc)$. Here $a$ is the number of cells in which contact and transcription were 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 medium 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$^{30}$, 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.** The number of embryos and cells used was determined by experimental feasibility and not statistical power.

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

**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 the Oligopaint library design used in this study was modified from routines developed in the Zheang laboratory at Harvard University$^{25,26}$. 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 Supplementary Data and can also be viewed through the UCSC genome browser at https://tinyurl.com/y7wfl8tx. Additional microscopy data are not hosted online owing to their size (12.3 Tb), but are available upon reasonable request.

52. Tokunaga, M., Imarmoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. Nat. Methods 5, 159–161 (2008).

53. Moffitt, J. R. et al. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. Proc. Natl Acad. Sci. USA 113, 11046–11051 (2016).

54. Gu, B. et al. Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. Science 359, 1050–1055 (2018).

55. Moffitt, J. R. et al. High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. Proc. Natl Acad. Sci. USA 113, 14456–14461 (2016).

56. Babcock, H., Sigal, Y. M. & Zhuang, X. A high-density 3D localization algorithm for stochastic optical reconstruction microscopy. Opt. Nanoscopy 1, 6 (2012).

57. Perry, M. W., Boettiger, A. N., Bothma, J. P. & Levine, M. Shadow enhancers foster robustness of Drosophila gastrulation. Curr. Biol. 20, 1562–1567 (2010).

58. Briggs, J. A. et al. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. Science 360, eaar5780 (2018).

59. Van Der Maaten, L. J. P. & Hinton, G. E. Visualizing high-dimensional data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605 (2008).

60. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters. KDD-96 Proc. 226–231 (1996).

61. Heris, S. DBSCAN Clustering Algorithm https://www.mathworks.com/}

62. Moffitt, J. R. et al. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. Proc. Natl Acad. Sci. USA 113, 11046–11051 (2016).

63. Perry, M. W., Boettiger, A. N., Bothma, J. P. & Levine, M. Shadow enhancers foster robustness of Drosophila gastrulation. Curr. Biol. 20, 1562–1567 (2010).

64. Briggs, J. A. et al. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. Science 360, eaar5780 (2018).

65. Van Der Maaten, L. J. P. & Hinton, G. E. Visualizing high-dimensional data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605 (2008).

66. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters. KDD-96 Proc. 226–231 (1996).

67. Heris, S. DBSCAN Clustering Algorithm https://www.mathworks.com/}

68. Moffitt, J. R. et al. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. Proc. Natl Acad. Sci. USA 113, 11046–11051 (2016).

69. Van Der Maaten, L. J. P. & Hinton, G. E. Visualizing high-dimensional data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605 (2008).

70. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters. KDD-96 Proc. 226–231 (1996).

71. Heris, S. DBSCAN Clustering Algorithm https://www.mathworks.com/}

72. Moffitt, J. R. et al. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. Proc. Natl Acad. Sci. USA 113, 11046–11051 (2016).

73. Van Der Maaten, L. J. P. & Hinton, G. E. Visualizing high-dimensional data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605 (2008).

74. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters. KDD-96 Proc. 226–231 (1996).

75. Heris, S. DBSCAN Clustering Algorithm https://www.mathworks.com/}
Extended Data Fig. 1 | ORCA methodology and experimental controls. a, Schematic of imaging approach. Top, 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 Table 5) and a strand-displacement oligo for each barcode (Supplementary Table 6). Bottom, 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, the fluorescent Cy5 signal from labelled barcode ‘1’ and the corresponding Cy3 signal from the fiducial channel. Bottom, the same region after treatment with the strand-displacement oligo. Centre 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 nanometres 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 five 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 (centre). 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 no 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 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 measurements33 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 | See next page for caption.
Extended Data Fig. 2 | Barcode detection efficiency. a, The mean detection efficiency of each barcode for all BX-C probesets. Error bars represent s.e.m. \( n = 8,801 \) cells, \( n = 94,300 \) cells, and \( n = 15,230 \) 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 boxes mark the interquartile range, and the whiskers extend to the farthest data points 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 smaller than the typical differences between segments, demonstrating the robustness of the segment-specific findings to missing data. Error bars denote s.e.m. h, Histogram showing the detection efficiency per cell for each probeset.
Extended Data Fig. 3 | STORM analysis of paired Drosophila homologues. 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 homologues labelled with ORCA barcode 69. Similar images were observed for all barcodes shown in c, d. Centres from 3D Gaussian fitting of the point spread function are represented as plus symbols on spots. Occasional double peaks occur at a frequency similar to that of background non-specific localizations; these may represent stray probes, noise in peak detection of weakly labelled loci, or split homologues. c, d, Quantification of the frequency of single peak detection for 10 kb (c) and 3 kb (d) per barcode. The high level of single peak detection is indicative of the tightly paired homologues for the BX-C. For the boxplots in c, d, the red line marks the median, the upper and lower limits of the boxes mark the interquartile range, and the whiskers extend to the farthest data points 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 homologues 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 locations of the CTCF ChIP-seq\textsuperscript{24} peaks, the Sox2 gene, and the ORCA barcodes are indicated. d, Average contact frequency as measured by high resolution Hi-C\textsuperscript{24}. The TAD containing Sox2 is outlined with black lines and 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 embryo about 11 hpf. b, A dorsolateral cryosection of another embryo, about 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 (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 different promoters. Similar results to those shown 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.
a Enhancer-promoter distance in active vs. inactive cells

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 or absent. The average distances are marked by plus symbols. 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 of cells, $n$, in each distribution is indicated. The enhancers from each genetic domain (two from each of the largest domains) are labelled by their classical names. The relative positions of enhancers and promoters are shown below. Note that most cells from the silent state come from anterior regions where the BX-C is repressed and compacted by Polycomb. 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 computation of the insulation score (plotted below each graph) and application of 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 that after normalization, organization in the head and anterior thorax is apparently random, and thus uncorrelated between datasets. The systematic features that appear 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 were used to compute the correlation.
Extended Data Fig. 10 | Properties of Fub mutant embryos.

a, Expression of abd-A and Abd-B in wild-type (WT) and Fub mutant embryos. Note that abd-A expression expands to T3 in mutant embryos, compared to wild-type 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 in wild-type 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 34 in wild-type embryos and the predicted extent of H3K27me3 in Fub mutants are shown in purple. Predicted H3K27me3 is based on the activation of abd-A observed in the mutant and previous descriptions of phenotypes of genetic border deletions in Drosophila25–27. The region deleted in Fub mutants is marked by red dotted lines. c, Fecundity of wild-type, homozygous Fab-7 and balanced Fub lines, plotted as the number of viable adults produced by eight fertilized females in a 2-day window. d, The height that each Drosophila larva, represented as a dot, climbed before pupation. Average height is indicated on the x-axis. e, Comparison of flight ability between wild-type and Fub/MKRS lines, using the metric developed by Bartholomew et al.63. The metric reports the fraction of adult Drosophila that 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 data points within 1.5 times the interquartile range.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

| n/a | Confirmed |
|-----|-----------|
| ☐   | ☑         |

The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement

☐ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

☐ The statistical test(s) used AND whether they are one- or two-sided

☐ Only common tests should be described solely by name; describe more complex techniques in the Methods section.

☐ A description of all covariates tested

☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

☐ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

☐ For null hypothesis testing, the test statistic (e.g. $F$, $t$, $r$) with confidence intervals, effect sizes, degrees of freedom and P value noted

☐ Give P values as exact values whenever suitable.

☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

☐ Estimates of effect sizes (e.g. Cohen's $d$, Pearson's $r$), indicating how they were calculated

☐ Clearly defined error bars

☐ State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

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 University31,62. 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. Descriptions of all code are provided in the readme files in the indicated repositories.

Data analysis

Additional code used in the analysis of the ORCA data is available at https://github.com/BoettigerLab/ORCA-public. A description of the code is provided in the Readme included.

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

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All DNA paths from all cells are available upon request as indicated in the Supplementary materials (the files are too large to append as Extended Data Tables). Image data is available upon request. No datasets requiring public deposition were produced for this study.

Field-specific reporting

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

- [x] Life sciences
- [ ] Behavioural & social sciences
- [ ] Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

| Sample size | Sample size was determined by experimental feasibility and by any associated statistics used in comparisons. All sample sizes are indicated and meet or exceed standard practice for single cell data. |
|-------------|------------------------------------------------------------------------------------------------------------------|
| Data exclusions | No data were excluded. |
| Replication | All results were reproduced as described and indicated in the corresponding data panels. No data failed to reproduce. |
| Randomization | Embryos were selected at random prior to cryosectioning and labeling. |
| Blinding | No blinding was used in these experiments. |

Reporting for specific materials, systems and methods

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| [x] | Unique biological materials |
| [x] | Antibodies |
| [x] | Eukaryotic cell lines |
| [x] | Palaeontology |
| [x] | Animals and other organisms |
| [x] | Human research participants |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| [x] | ChIP-seq |
| [x] | Flow cytometry |
| [x] | MRI-based neuroimaging |

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s) mES R1 cells were a generous gift from the Wysocka lab.

Authentication

These cells were 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.

Mycoplasma contamination

The cells tested negative for mycoplasma contamination prior to our experiments.

Commonly misidentified lines

(See ICLAC register) N/A
## Animals and other organisms

Policy information about [studies involving animals; ARRIVE guidelines](https:// ARRIVEguidelines.org) recommended for reporting animal research

| Category                | Description                                                                                                                                 |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------------|
| Laboratory animals      | Wild type and mutant flies and embryos (Drosophila melanogaster), both male and female, were used in this study as described in the methods. Wildtype strain was CantonS (Stanford Fly Core Facility), mutants used were Fab-7 and Fub, which were gifted from Welcome Bender. Embryo from age 8 hours post fertilization through 12 hours post fertilization were analyzed as indicated. |
| Wild animals            | No wild animals were used in these studies.                                                                                                                                                           |
| Field-collected samples | No field collected samples were used in these studies.                                                                                                                                               |