Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions

Diego I. Cattoni, Andrés M. Cardozo Gizzi, Mariya Georgieva, Marco Di Stefano, Alessandro Valeri, Delphine Chamousset, Christophe Houbon, Stephanie Déjardin, Jean-Bernard Fiche, Inma González, Jia-Ming Chang, Thomas Sexton, Marc A. Marti-Renom, Frédéric Bantignies, Giacomo Cavalli & Marcelo Nollmann

At the kilo- to megabase pair scales, eukaryotic genomes are partitioned into self-interacting modules or topologically associated domains (TADs) that associate to form nuclear compartments. Here, we combine high-content super-resolution microscopies with state-of-the-art DNA-labeling methods to reveal the variability in the multiscale organization of the Drosophila genome. We find that association frequencies within TADs and between TAD borders are below ~10%, independently of TAD size, epigenetic state, or cell type. Critically, despite this large heterogeneity, we are able to visualize nanometer-sized epigenetic domains at the single-cell level. In addition, absolute contact frequencies within and between TADs are to a large extent defined by genomic distance, higher-order chromosome architecture, and epigenetic identity. We propose that TADs and compartments are organized by multiple, small-frequency, yet specific interactions that are regulated by epigenetics and transcriptional state.
The multiscale organization of eukaryotic genomes defines and regulates cellular identity and tissue-specific functions. At the kilo-megabase scales, genomes are partitioned into self-interacting modules or topologically associated domains (TADs)\(^1\). TAD formation seems to require specific looping interactions between TAD borders\(^7,8\), while the association of TADs can lead to the formation of active/repressed compartments\(^8\). These structural levels were often seen as highly stable over time; however, recent single-cell Hi-C studies have reported different degrees of heterogeneity\(^10,11\). Other studies have shown that genomes also display stochasticity in their association with the nuclear lamina\(^12\), in the formation of chromosome territory neighborhoods\(^13,14\), and in gene kissing\(^15\). However, access to single-cell absolute probability contact measurements between loci and efficient detection of low-frequency, long-range interactions are essential to quantify the stochastic behavior of chromatin at different scales.

Here, we combined high-content super-resolution microscopy with state-of-the-art DNA-labeling methods to reveal the variability in the multiscale organization of chromosomes in different cell types and developmental stages in *Drosophila*. Remarkably, we found that stochasticity is present at all levels of chromosome architecture, but is locally modulated by sequence and epigenetic state. Contacts between consecutive TAD borders were infrequent, independently of TAD size, epigenetic state, or cell type. Moreover, long-range contact probabilities between non-consecutive borders, the overall folding of chromosomes, and the clustering of epigenetic domains into active/repressed compartments displayed different degrees of stochasticity that globally depended on cell type. Overall, our results show that contacts

---

**Fig. 1** TAD organization arises from modulation of stochasticity. a, Top, region of Hi-C contact matrix of chromosome 2L. The black-dotted line demarcates a TAD and pink and cyan boxes represent the Oligopaint- labeled TAD borders (TB). Chromatin epigenetic state is indicated at the bottom using the color code of panel b. Bottom, representative three-color 3D-SIM image in two orientations. DAPI, TB2, and TB3 are shown in gray, pink, and cyan, respectively. Scale bar = 1 µm for the main image. The inset displays 5× amplification of the selected region. **b**, Oligopaint libraries in chromosomes 2L and 3R employed in this study (TB1-16 at TAD borders and IT17-19 within TADs). Colored boxes display the chromatin type of TADs as defined in Supplementary Fig. 1a, b. Red: active, blue: repressed, and black: inactive. Dotted colored lines indicate the combinations of libraries measured. **c**, 3D distance distributions between TB2-TB2 and TB2-TB3. The mean colocalization resolution, estimated from two-color labeling of a single border (40 nm, vertical blue dashed line). Blue and black solid lines represent Gaussian fittings. The absolute contact probability between libraries was obtained from the integral of the area of the Gaussian fitting (shaded gray) below 120 nm (Supplementary Fig. 1e). \(N=161\) and 556 for TB2-TB2 and TB2-TB3, respectively, from more than three biological replicates. **d**, Absolute contact probability between consecutive borders vs. genomic distance. Chromatin state of TADs is color coded as defined in panel b. Error bars represent SEM. **e**, Normalized Hi-C counts between consecutive TAD borders (circles) and random loci (solid gray line) as a function of genomic distance for S2 and late embryonic cells. Matrix resolution = 10 kb. Two biological replicates for each cell type were performed. **f**, Schematic representation of contact probability between and within TADs (solid colored lines) for late embryo and S2 cells at the chromosomal region shaded in panel b. Sizes of TADs (gray-shaded triangles) are proportional to genomic length (scale bar on top). Chromatin type is indicated at the bottom. The thickness of the lines and color indicate the absolute contact probability. Dotted lines indicate inter-TAD contacts. Early embryo measurements are depicted in Supplementary Fig. 1k. Numbers of cells for each combination are provided in Supplementary Fig. 11.
between and within TADs are rare, but can be epigenetically modulated to give rise to different levels of higher-order genome organization. We anticipate that our results will guide new statistical models of genome architecture and will be a starting point for more sophisticated studies to understand how a highly variable, multiscale organization can ensure the maintenance of stable transcriptional programs through cell division and during development.

Results

Multiple low-frequency interactions mediate TAD assembly. A major mechanism for TAD formation in mammals involves the stable looping of TAD borders. Stable looping between TAD borders was also recently proposed to be relevant for the maintenance of transcriptional programs during Drosophila development. However, long-lived stable interactions are unlikely to allow for rapid responses in gene regulation. To study this
apparent contradiction, we developed a method to dissect the changes in TADs organization at the single-cell level in three transcriptionally distinct *Drosophila* cell types: early (stage 5) and late (stage 16) embryos; and an immortalized cell line (S2). Pairs of TAD borders were labeled with Oligopaint libraries\(^\text{15}\) and imaged using multicolor three-dimensional structured illumination microscopy (3D-SIM)\(^\text{16, 17}\) (Fig. 1a). TAD chromatin types were defined as active, repressed, or inactive following the distribution of epigenetic marks (Supplementary Fig. 1a). Borders flanking TADs with different chromatin states were imaged in chromosomes 2L and 3R (Fig. 1b and Supplementary Fig. 1b), and appeared in microscopy as well-defined foci (Fig. 1a) whose size increased proportionally with the genomic length of the library (Supplementary Fig. 1c). A large proportion of cells (60–70%) displayed a single focus, consistent with a high degree of homologous pairing independently of the ploidy of each cell type (Supplementary Fig. 1d, e).\(^\text{18, 19}\) Distances between TAD borders were Gaussian distributed for all cell types (Fig. 1c and Supplementary Fig. 1f–h). Remarkably, the width of these distributions was comparable to the mean distance between TAD borders, revealing a high degree of structural variability, independently of TAD size or epigenetic state (Fig. 1c and Supplementary Fig. 1i). Further, the linear relation between dispersion and physical distance (Supplementary Fig. 1i–j) suggests that this variability is regulated by the polymer properties of the chromatin fiber.

Next, we quantified the absolute contact probability between consecutive borders by integrating the probability distribution below 120 nm (99% confidence interval obtained from single-library two-color control experiments, Fig. 1c and Supplementary Fig. 1e). Notably, the contact probability between consecutive TAD borders was below 10%, independently of the cell type or of the epigenetic state of the TAD being flanked (Fig. 1d). Consistently, Hi-C contact frequencies between consecutive TAD borders vs. random genomic loci were indistinguishable (Fig. 1e). These results, combined with the lack of enrichment of CTCF and cohesin at TAD borders in *Drosophila*,\(^\text{20}\) suggest that TAD assembly does not involve stable loops in flies, but rather can be explained by an “insulation–attraction” mechanism.\(^\text{21}\) This model may provide an alternative explanation for the maintenance and formation of more than 50% of metazoan TADs whose boundaries are not formed by looping interactions as defined by Hi-C experiments.\(^\text{8}\)

In agreement with this model, absolute contact probabilities within TADs and between their borders were similar (Fig. 1f and Supplementary Fig. 1k), with inactive/repressed TADs displaying higher contact probabilities than active TADs (7 ± 1% vs. 2.7 ± 1%, mean ± SD). Contact probabilities within TADs were in all cases considerably higher than those with neighboring TADs (Fig. 1f), indicating that stochasticity is locally modulated at the TAD level. Of note, contacts across TAD borders were not uncommon (~3%, Fig. 1f), implying frequent violations of boundary insulation at TAD borders. These results indicate that confinement of chromatin into TADs may require only small differences in absolute contact probabilities (~2-fold). Thus, condensation of chromatin into TADs may arise from a multitude of low-frequency, yet specific, intra-TAD contacts.

**Infrequent long-range contacts modulate chromatin folding.** Recent Hi-C studies suggested that stable clustering between neighboring active TAD borders regulates transcriptional
programs that persist during development\(^7\). We directly tested this hypothesis by measuring the contact probabilities between nonconsecutive TAD borders (Fig. 2a). Hi-C contact frequencies among TAD borders increased nonlinearly with absolute contact probabilities (Fig. 2a and Supplementary Fig. 2a), with both exponential and power-law empirical models fitting the data equally well. Our results highlight the ability of Hi-C to enhance the detection of high-probability contacts and also suggest the need to relate Hi-C data to physical distances with a nonlinear relationship. This would allow a better discrimination of low-frequency contacts (1–3%, Fig. 2a) such as those observed within and between TADs (Fig. 1f) and a more realistic conversion of Hi-C maps into 3D-folded structures.

Contact probabilities between nonconsecutive TAD borders were in all cases low (<9%, Fig. 2b) and decreased monotonically with physical and genomic distance following a power-law behavior (Fig. 2b and Supplementary Fig. 2b–c). Notably, the decay exponents were different between cell types (Fig. 2b and Supplementary Fig. 2b), indicating that levels of stochasticity are globally modulated between cell types, possibly reflecting cell
Impact of long-range contacts in chromosome-wide folding. To quantitatively dissect stochasticity at larger genomic scales, we labeled 69 quasi-equidistant TAD borders encompassing 90% of chromosome 3R (Fig. 3a and Supplementary Fig. 3a, b). Tens of foci were resolved in embryonic and S2 cells by 3D-SIM (Fig. 3a). The paired probability distance distribution $p(r)$ between any two foci exhibited moderate single-cell variations (Fig. 3b) but was considerably different between cell types (Fig. 3b and Supplementary Fig. 3c). The chromosome elongation and mean volume, obtained from the maximum pairwise distance ($D_{\text{max}}$) and the radius of gyration ($R_g$, Fig. 3c), decreased to almost half when comparing S2 and late embryonic cells, while early embryonic cells adopted intermediate values (Fig. 3b).

Reinforcing these findings, changes in Hi-C contact frequency of S2 vs. late embryo for the 69 TAD borders were notable in the sub-Mb scale (200–600 kb), and they extended to genomic distances as high as ~10 Mb (Fig. 3d), suggesting that changes in chromosome compaction between cell types arise from an increased frequency of interactions affecting all genomic scales. All in all, these data indicate that chromosome folding is highly variable, with mild, cell type-specific increases in the probability of long-range contacts being sufficient to produce large changes in the manner in which chromosomes occupy the nuclear space (Fig. 3c).

**Stochastic nanoscale organization of epigenetic marks.** Interchromosomal and intrachromosomal Hi-C maps have revealed that active and repressed TADs may associate to form two types of compartments (namely A and B)\textsuperscript{9, 28}. To study this higher-order level of organization in single cells and at the single-molecule level, we immunolabeled active and repressive epigenetic marks (histones H3K4me3 and H3K27me3, respectively) and performed multicolor direct stochastic optical reconstruction microscopy (dSTORM)\textsuperscript{29–31}, a method that provides a higher spatial resolution than 3D-SIM. DSTORM imaging revealed that active and repressive histone marks distributed non-homogeneously across the cell nucleus, forming discrete compartments of tens to hundreds of nanometers for all cell types (Fig. 4a and Supplementary Fig. 4a). Repressed and active
chromatin marks were strictly segregated at the nanoscale for all cell types, as revealed by coordinate-based colocalization analysis (aCBC43, Fig. 4b). These findings were confirmed by independent colocalization methods and by additional controls using doubly labeled nuclear factor and noncolocalizing epigenetic marks (Supplementary Fig. 4b–d). Interestingly, active marks were often observed at borders of/or demarcating large repressed compartments, mirroring their alternating one-dimensional genomic distributions (Fig. 4c).

To investigate if active and repressed compartments also varied among cell types and development, we resorted to one-color dSTORM using Alexa 647 as the fluorophore of choice (the results were similar when using other fluorophores, Supplementary Fig. 4e). Compartments were detected using a Voronoi diagram-based algorithm (Fig. 4d, e)33. In all cases, active compartments were smaller than repressive compartments in agreement with two-color dSTORM observations (Fig. 4e–e and Supplementary Fig. 5a, b). Interestingly, for both marks, the number of compartments and their sizes showed variations between single cells of the same type (Supplementary Fig. 5c, d). To further evaluate if changes in compartment sizes correlated with changes in local chromatin folding, we quantified the density of single-molecule detections in active and repressed compartments. Notably, the local density of compartments was higher for both types of marks in embryonic cells than for S2 cells (Supplementary Fig. 6a), consistent with our previous findings (Fig. 2e, f) and with compartment contact density from Hi-C counts (Supplementary Fig. 6b).

To study whether the nanoscale organization of repressive and active marks reflected the epigenomic domain organization from ensemble genome-wide methods, we predicted the physical sizes of epigenomic domains (Supplementary Fig. 6c) and compared them with those obtained by direct observation. The predicted size distributions failed to recover the largest compartments observed by microscopy (Fig. 4f, g and Supplementary Fig. 6c). We reasoned that large compartments are likely to arise from clustering of smaller epigenetic domains ("clustered compartments").

To quantify this phenomenon, we calculated the percentage of compartments not accounted for by the distribution of epigenetic domains. This percentage of clustered compartments was below <10% for embryonic cells and almost absent in S2 cells (Fig. 4h). The latter is consistent with higher Hi-C contact frequency between H3K27me3 domains in embryos than in S2 cells (Fig. 4i). Repressive and active compartments showed different degrees of clustering (Fig. 4f–h), indicating that stochasticity can be specifically modulated by transcriptional/epigenetic states. This is likely due to the different mechanisms of clustering formation at play, such as Polycomb regrouping of repressed genes34 vs. transient interactions of active genes35, 36. It is important to note, however, that the large majority of compartments (~90%) could be accounted for by the predicted distributions of epigenomic domains, consistent with the majority of the epigenetic domains described by genome-wide methods existing at the single-cell level. These results are consistent with the cell type-specific higher-order organization of chromatin arising from stochastic contacts between chromosomal regions harboring similar epigenetic marks, likely reflecting cell type-specific transcriptional programs.

Discussion

In this work, we show that genome organization in Drosophila is not driven by stable or long-lived interactions but rather relies on the formation of transient, low-frequency contacts whose frequencies are modulated at different levels. Stochasticity is modulated locally at the TAD level by specific intra-TAD interactions, and globally at the nuclear level by interactions of TADs of the same epigenetic type. Furthermore, stochasticity is also regulated between cell types. These modulated stochasticities reveal a novel mechanism for the spatial organization of genomes. These pieces of evidence could be critical for a more accurate understanding of how different cell types interpret genomic and epigenomic states to produce different phenotypes. Dynamic measurements of chromosome organization with high coverage will be needed in future to further explore the origin of heterogeneity in chromosome architecture and to determine whether genome organization is a stationary or a fully stochastic process.

In mammals, a large proportion of consecutive TAD borders is looped by specific interactions apparently mediated by CTCF and cohesin8, 37, 38. Recent reports suggested that this mechanism may also be at play in Drosophila25, 24. Our results, however, provide compelling evidence that looping of consecutive TADs borders in Drosophila is rare at the single-cell level. These observations, supported by recent studies showing that cohesin-enriched loop anchors in Drosophila are found within TADs rather than at TAD borders39, 40, are against TAD boundaries being the bases of stable chromatin loops. Thus, the lack of frequent interactions between TAD borders could be consistent with a model where TADs arise from a dynamic balance between cohesin-mediated loop extrusion41, the blocking of that movement by architectural proteins, and factors that may load or remove cohesin42, 43. In Drosophila, however, CTCF and cohesin are not found enriched at TAD borders. Thus, we envision that other factors (e.g., Beaf-32 and CP190/chromator instead of CTCF and cohesin) could play a role at looping and dynamically extruding distant DNA fragments within the same TAD. In addition, active marks may help determine the properties of TAD boundaries25, 44 while other epigenetic marks could play a role in the formation of polycomb and inactive TADs45. Similar epigenetic mechanisms may even play a role in TAD folding in mammals, consistent with the observation that CTCF depletion leads only to minor changes in TAD organization45. Importantly, our data provide quantitative estimates of the stochasticity and absolute frequencies of interactions within TADs, imposing important constraints on any model of TAD formation in Drosophila.

Recent reports suggested that TAD borders enriched in housekeeping genes form stable 3D colocalization patterns that persist during development7. In contrast, we found that 3D contacts between TAD borders are rare and highly stochastic in all cell types investigated. These results are consistent with recent single-nucleus-Hi-C studies reporting that TAD formation is highly stochastic in mammals46, and with the rapid association and dissociation of transcription foci25 rather than with stable transcription factories.

Most current spatial models of genome architecture rely on interpreting interaction maps from chromosome conformation capture-based experiments, which seize the relative frequencies of interactions between loci at close spatial proximity. However, translation of relative contact frequencies into spatial distances is challenging. Our direct single-cell measurements of absolute contact probabilities, full distance distributions, and dissection of low-frequency events for different chromatin and cell types will complement the existing methods to refine the next generation of statistical models of genome architecture. Our results call for more sophisticated studies to reveal how a highly stochastic genome organization can ensure the maintenance of stable transcriptional programs through cell division and during development.

Methods

Cell culture and embryonic tissue preparation. Drosophila S2 cells were obtained from the Drosophila Genomics Resource Center. S2 cells were grown in serum-supplemented (10%) Schneider’s S2 medium at 25 °C. Oregon-R w1118 fly stocks

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-017-01962-x

ARTICLE
were maintained at room temperature (RT) with natural light/dark cycle and raised in standard cornmeal yeast medium. Following a precollection period of at least 1 h, embryos were collected on yeast 0.4% acetic acid agar plates and incubated at 25 °C until they reached the desired developmental stage: 2–3 h or 12–14 h (total developmental time) for early and for late embryos, respectively. Embryos were mechanically broken and immediately fixed by using 4% PFA for PBS in 10 min at RT. S2 cells were allowed to adhere to a poly-l-lysine coverslip for 30 min in a covered 35-mm cell culture dish before 4% PFA fixation.

Immunostaining. Cells were permeabilized with 0.5% Triton X-100 for 10 min and blocked with 5% of bovine serum albumin (BSA) for 15 min at RT. Primary antibodies anti-H3K27me3 (pAb-195-050, Diagene and ab6002, Abcam), anti-H3K4me3 (cat#04-745, Millipore and ab1012, Abcam), anti-Polycomb47, and anti-Beaf-3248 (made from a rabbit by Eurogentec) were coupled to Alexa Fluor 647 or Cy3, as described below. Antibodies were used at a final concentration of 10 µg ml−1 in PBS and 1% BSA. Coverslips were incubated overnight at 4 °C in a humidity chamber and washed three times with PBS before introducing fiducial markers diluted at 1/4000 (Tetraspex, #10195142, FisherScientific). Coverslips were mounted on slides with 100-µl wells (#2410, Glaswarenfabrik Karl Hecht GmbH & Co KG) in dSTORM buffer composed of PBS, glucose oxidase (G7141-50KU, Sigma) at 2.5 mg ml−1, catalase at 0.2 mg ml−1 (#C3155-50MG, Sigma), 10% glucose, and 50 mM of β-mercaptoethanol (MEA, #M9768-5G, Sigma). Coverslips were sealed with duplicating silicone (Twinsil, Rotech).

Oligopaint libraries. Oligopaint libraries were constructed from the Oligopaint public database (http://genetics.med.harvard.edu/oligopaint/). All libraries consisted of 42-mer sequences discovered by Oligopaint and a list of additional sequences of 30–124 -L 42-D 1000 -t 80-T 799 -s 70 -x 35 -P 80 -m 4D “GGGGCCCTCCTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTAAAAAA” g. Oligonucleotides for libraries 1–18 and BX-C were ordered from CustomArray (Bothell, WA). The procedure used to synthesize Oligopaint probes is described below. Chr3-9 borders oligonucleotides were purchased from MYcroarray (Ann Arbour, MI). Oligopaint probes for this library were synthesized using the same procedure as for the other libraries except for the initial emulsion PCR step. Secondary, fluorescently labeled oligonucleotides were synthesized by Integrated DNA Technologies (IDT; Coralville, IA for Alexa488) and by Eurogentec (Angers, France for Cy3b). See Supplementary Table 1 for a list of Oligopaint probe sets used for libraries 1–18. Sequences for secondary oligonucleotides and PCR primers are described below (Supplementary Tables 2–4). Details for the methods used for probe synthesis are provided in Supplementary Notes.

Fluorescence in situ hybridization. To prepare sample slides containing fixed S2 cells for FISH, S2 cells were allowed to adhere to a poly-l-lysine coverslip for 1 h in a covered 35-mm cell culture dish at 25 °C. The slides were then washed in PBS, fixed in 4% paraformaldehyde (PFA) for 10 min, rinsed 3 times for 5 min in PBS, permeabilized for 10 min with 0.5% Triton, rinsed in PBS, incubated with 0.1% HCl for 10 min, washed 3 times in 1× saline–sodium citrate–0.1% Tween-20 (2xSSCT), and incubated in 2×SSCT/50% formamide (v/v) for at least 30 min. Then, probes were prepared by mixing 20 µl of hybridization buffer FHB (50% formamide, 10% Dextran sulfate, 2×SSC, and Salmon Sperm DNA 0.5 mg ml−1), 0.8 µl of RNAse A, 0.5 µl of RNase inhibitor, and 30 pmol of oligo-dT. 12 µl of this mix was added to a slide before adding and sealing with rubber cement the coverslips with cells onto the slide. Probes and cells are finally codenatured for 3 min at 78 °C before hybridization overnight at 37 °C. The next day, the slides were washed once in 2× SSC at 37 °C, and then three times for 5 min each at 0.1× SSC at 45 °C. Finally, they were stained with 0.5 µg ml−1 of DAPI for 10 min, washed with PBS, mounted in Vectashield, and sealed with nail polish. For more detailed protocol, see49.

Image acquisition and postprocessing of 3D-SIM data. Samples were prepared as described above and mounted on an OMX V3 microscope (Applied Precision Inc.) equipped with a 100×/1.4 oil PlanApo objective (Olympus) and three emCCD cameras. Laser lines at 405-nm, 488-nm, and 561-nm excitation were used to excite DAPI, Alexa488, and Cy3b, respectively. Each channel was acquired sequentially. A transmission image was also acquired to control for cell morphology. For each channel, a total of 1455 images made of 97 different Z-planes separated by 125 nm were acquired, in order to acquire a stack of 12 µm. Three different planes were acquired (z−z' + z''−z''), as well as five frames per second per channel, in 120 ms total acquisition time. To reconstruct 3D-SIM images using softWorX v5.0 (Applied Precision Inc.). The final voxel size was 39.5 nm in the lateral (xy) and 125 nm in the axial (z) directions, respectively, for a final 3D stack volume of ~40 × 40 × 12 µm. Multicolor TetraSpeck beads (100 nm in diameter, Invitrogen) were used to measure x, y, and z offsets, rotation about the z-axis, and magnification differences between fluorescence channels. These corrections were applied to the reconstructed images. The same beads were used to validate the reconstruction process, ensuring a final resolution of ~120 nm in xy and ~300 nm in z at 252 nm of emission wavelength. 3D-SIM raw and reconstructed images were analyzed with SIMCheck ImageJ Plug-in.29. Acquisition parameters were optimized to obtain the best signal-to-noise ratio, avoiding photobleaching between the different angular, phase, and axial acquisitions.

Three additional methods were employed as controls: pixel, Pearson, and Manders correlation. For the latter, two-color digital images were reconstructed from the localization using standard procedures29 and then used to plot the correlation between pixel intensities (pixel correlation analysis), or to calculate the Pearson or Manders correlation coefficients38,39.

Analysis of one-color dSTORM data. Single-molecule localizations are converted into a Voronoi diagram using a modified version of the Voronoi tessellation algorithm of Levet et al.31. Compartment segmentation is directly calculated from

ARTICLE NATURE COMMUNICATIONS | DOI: 10.1038/s41467-017-01962-x

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-017-01962-x
the Voronoi diagram using three steps. First, densities of each polygon are cal-
culated as the inverse of their area. Densities are then thresholded using the general
criteria of Levet et al. 33. Using this criterion, in which the threshold is determined
by the average localization density, a random distribution of localizations did not
provide any segmented polygon. Finally, polygons that have a density higher than
the threshold and that are touching each other are merged to define the com-
partment outline. Compartment sizes are obtained by interpolating each segmented
compartment on a grid of 5-nm size and calculating their equivalent diameter
using standard morphological operations. Probability density functions in
compartment-size histograms are calculated such that the area of each bar is the
using standard morphological operations. Probability density functions in

Analysis of genome-wide data. Chromatin states were defined according to the
enrichment in the percentages of H3K4me3 and H3K27me3/PC, as described in
Supplementary Fig. 6. Calculation of the genomic size distributions of H3K27me3
and H3K4me3 domains (Supplementary Fig. 6) was performed as follows: (1) ChiP-clip-seq computed peaks were downloaded from ModEncode (ftp://data.
modencode.org/D3.melanogaster). Data sets used are described in Supplementary
Table 3. (2) Peak positions and intensities were used to resample the data and
produce a continuous signal as a function of genomic position. (3) The threshold
was thresholded with a threshold of 0.1 of the log of the maximum intensity signal,
ensuring that even peaks with very low intensity were retained. (4) Domains were
defined as continuous segments extending for more than 2 bp with nonzero
intensity. (5) Domains that were closer than 1 kb were fused together. This pro-
cedure was robust to calculate size distributions above 3 kb (Supplemen-
tary Fig. 6). (6) Finally, we estimated physical domain sizes from their genomic
length as follows. The size of each genomic domain in bp was converted into
nanometers using the empirical power law that relates genomic sizes to physical
nanometers using the empirical power law that relates genomic sizes to physical

In situ Hi-C data processing and normalization. Hi-C data were processed using an in-house pipeline based on TADbit. First, the quality of the reads was checked
using the quality_plot() function in TADbit, which is similar to the tests performed
by the FastQC program with adaptations for Hi-C data sets. Next, the reads are
mapped following a fragment-based strategy, as implemented in TADbit where
each side of the sequenced read was mapped in full length to the reference genome
(dm3). After this step, if a read was not uniquely mapped, we assumed that the read
was chimeric due to ligation of several DNA fragments. We next searched for
ligation sites, discarding those reads in which no ligation site was found. The
remaining reads were split as often as ligation sites were found. Individual split
read fragments were then mapped independently. Next, we used the TADbit-
filtering module to remove noninformative contacts and to create contact matrices.
From the resulting contact matrices, low-quality bins (those presenting low contact
numbers) were removed, as implemented in TADbit’s filter_columns() function.
Next, the matrices were normalized using the ICG algorithm 35. The normalization
iterations stopped when the biases were diverting less than 10% of the previous
values or a max of 10 iterations. Finally, all matrices were corrected to achieve an
average content of one interaction per cell. All parameters in TADbit were kept at
default values.

The resulting late embryo and S2 Hi-C interaction maps (at 10 kb resolution) of
the different replicates for each experiment were highly correlated (correlation
coefficients from genomic distances ranging from 10 kb to 20 Mb were 0.99 to 0.75
and 0.95 to 0.45, respectively) and thus were further merged into the final data sets
with more than 282-million and 210-million valid pairs each (Supplementary Table 6).

Data availability. The Hi-C data reported in this study are available in the Gene
Expression Omnibus (GEO) repository under accession code GSE104961. Compu-
ter code and other data that support the findings of this study are available from
the corresponding author upon reasonable request.

Received: 19 June 2017 Accepted: 29 October 2017
Published online: 24 November 2017

References

1. Sexton, T. & Cavalli, G. The role of chromosome domains in shaping the
functional genome. Cell 160, 1049–1059 (2015).
2. Gilbert, D. M. & Fraser, P. Three dimensional organization of the nucleus:
adding DNA sequences to the big picture. Genome Biol. 16, 181 (2015).
3. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers
and their regulatory landscapes. Nature 502, 499–506 (2013).
4. Dixon, J. R. et al. Topological domains in mammalian genomes identified by
analysis of chromatin interactions. Nature 485, 376–380 (2012).
5. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-
inactivation centre. Nature 485, 381–385 (2012).
6. Sexton, T. et al. Three-dimensional folding and functional organization
principles of the Drosophila genome. Cell 148, 458–472 (2012).
7. Hug, C. B., Grimaldi, A. G., Kruse, K. & Vaquerizas, J. M. Chromatin
architecture emerges during zygotic genome activation independent of
transcription. Cell 169, 216–228.e19 (2017).
8. Rao, S. S. P. et al. A 3D Map of the human genome at kilobase resolution reveals
principles of chromosome folding. Cell 162, 687–688 (2015).
9. Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Exploring the three-dimensional
organization of genomes: interpreting chromatin interaction data. Nat. Rev.
Genet. 14, 390–403 (2013).
10. Giorgetti, L. et al. Predictive polymer modeling reveals coupled fluctuations in
chromosome conformation and transcription. Cell 157, 950–963 (2014).
11. Flyamer, I. M. et al. Single-nucleus Hi-C reveals unique chromatin
reorganization at oocyte-to-zygote transition. Nature 544, 110–114 (2017).
12. Kind, J. & van Steensel, B. Stochastic genome-nuclear lamina interactions:
modulating roles of Lamin A and BAF. Nuclear 5, 124–130 (2014).
13. Cremer, T. & Cremer, M. Chromosome territories. Cold Spring Harb. Perspect.
Biol. 2, a003889 (2010).
14. LeMaitre, C., Chretin, T., Cremer, M., Cavalli, G. & Cremer, T. Dynamic genome
architecture in the nuclear space: regulation of gene expression in three
dimensions. Nat. Rev. Genet. 8, 104–115 (2007).
15. Beliveau, B. J. et al. Single-molecule super-resolution imaging of chromosomes
and in situ haplotype visualization using Oligopaint FISH probes. Nat.
Commun. 6, 1–13 (2015).
16. Gustafsson, M. G. L. Surpassing the lateral resolution limit by a factor of two
using structured illumination microscopy. J. Microsc. 198, 82–80 (2000).
17. Gustafsson, M. G. L. et al. Three-dimensional resolution doubling in wide-field
fluorescence microscopy by structured illumination. Biophys. J. 94, 4957–4970
(2008).
18. Eng, J. C., Marshall, W. F.,邓, A., Agard, D. A. & Sedat, J. W. Homologous
chromosome pairing in Drosophila melanogaster proceeds through multiple
independent initiations. J. Cell Biol. 141, 5–20 (1998).
19. Zhang, Y. et al. Expression in aneuploid Drosophila S2 cells. PLoS Biol.
8, e1003200 (2010).
20. Ramirez, F. et al. High-resolution TADs reveal DNA sequences underlying
genome organization in flies. Preprint at: bioRxiv 115063 (2017).
21. Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin domains: the unit of
chromosome organization. Mol. Cell 62, 668–680 (2016).
22. Hou, C., Li, L., Qin, Z. S. & Corces, V. G. Gene density, transcription, and
insulators contribute to the partition of the Drosophila genome into physical
domains. Mol. Cell 48, 471–482 (2012).
23. Ulanov, S. V. et al. Active chromatin and transcription play a key role in
chromosome partitioning into topologically associating domains. Genome
Res. 26, 70–84 (2016).
24. Li, L. et al. Widespread rearrangement of 3D chromatin organization underlies
polycomb-mediated stress-induced silencing. Mol. Cell 58, 216–231 (2015).
25. Cisse, I. et al. Real-time dynamics of RNA polymerase II clustering in live
human cells. Science 341, 664–667 (2013).
26. Boettiger, A. N. et al. Super-resolution imaging reveals distinct chromatin
folding for different epigenetic states. Nature 529, 418–422 (2016).
27. Dekker, J. & Heard, E. Structural and functional diversity of topologically
associating domains. FEBS Lett. 589, 2877–2884 (2015).
28. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions
reveals folding principles of the human genome. Science 326, 289–293 (2009).
29. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic
optical reconstruction microscopy (STORM). Nat. Methods 3, 793–795 (2006).
30. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-
resolution imaging by stochastic optical reconstruction microscopy. Science
336, 810–813 (2008).
31. Heilmann, M. et al. Subdiffraction-resolution fluorescence imaging with
conventional fluorophores. Angew. Chem. Int. Ed. 47, 6172–6176 (2008).
32. Georgieva, M. et al. Nanometer resolved single-molecule colocalization
of nuclear factors by two-color super-resolution microscopy imaging. Methods
105, 44–55 (2016).
33. Levet, F. et al. SR-Tesseler: a method to segment and quantify localization-
based super-resolution microscopy data. Nat. Methods 12, 1065–1071 (2015).
34. Bantignies, F. & Cavalli, G. Polyclomb group proteins: repression in 3D. Trends
Genet. 27, 454–464 (2011).
35. Chuang, C.-H. & Belmont, A. S. Close encounters between active genes in the nucleus. *Genome Biol.* 6, 237 (2005).
36. Fraser, P. & Bickmore, W. Nuclear organization of the genome and the potential for gene regulation. *Nature* 447, 413–417 (2007).
37. Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad. Sci. USA* 112, E6456–E6465 (2015).
38. Vietri Rudan, M. et al. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* 10, 1297–1309 (2015).
39. Cubéñas-Potts, C. et al. Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic Acids Res.* 45, 1714–1730 (2016).
40. Eagen, K. P., Aiden, E. L. & Kornberg, R. D. Polycomb-mediated chromatin loop extension by a subkilobase-resolution chromatin interaction map. *Proc. Natl Acad. Sci. USA* 114, 8764–8767 (2017).
41. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* 15, 2038–2049 (2016).
42. Haarhuis, J. H. I. et al. The cohesin release factor WAPL restricts chromatin loop extension. *Cell* 169, 693–707.e14 (2017).
43. Barrington, C., Finn, R. & Hadjur, S. Cohesin biology meets the loop extrusion model. *Chromosoma* 25, 51–60 (2017).
44. Jost, D., Carrivain, P., Cavalli, G. & Vaillant, C. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res.* 42, 9533–9561 (2014).
45. Zain, J. et al. Cohesin and CTCF differentially affect chromatin architecture and expression in human cells. *Proc. Natl Acad. Sci. USA* 111, 996–1001 (2014).
46. Cheutin, T. & Cavalli, G. Progressive polycomb assembly on H3K27me3 compartments generates polycomb bodies with developmentally regulated motion. *PLoS Genet.* 8, e1002463 (2012).
47. Martinez, A.-M., Colomb, S., Déjardin, J., Bantignies, F. & Cavalli, G. Polycomb group-dependent cyclin A repression in Drosophila. *Genes Dev.* 20, 501–513 (2006).
48. Emberly, E. et al. BEAF regulates cell-cycle genes through the controlled deposition of H3K9 methylation marks into its conserved dual-core binding sites. *PLoS Biol.* 6, 2896–2910 (2008).
49. Bantignies, F. & Cavalli, G. Topological organization of Drosophila hox genes using DNA fluorescence in situ hybridization. *Methods Mol. Biol.* 1196, 103–120 (2014).
50. Ball, G. et al. SIMcheck: a toolbox for successful super-resolution structured illumination microscopy. *Sci. Rep.* 5, 15915 (2015).
51. Snell, V., Kittler, J. & Christmas, W. Segmentation and shape classification of nuclei in DAPI images. In *The 22nd British Machine Vision Conference* (University of Dundee, Dundee, UK, 2011).
52. Fiche, J.-B. et al. Recruitment, assembly, and molecular architecture of the SpoIIIE DNA pump revealed by superresolution microscopy. *PLoS Biol.* 11, e1001557 (2013).
53. Cattoni, D. L., Fiche, J.-B., Valeri, A., Mignot, T. & Nollmann, M. Super-resolution imaging of bacteria in a microfluidics device. *PLoS One* 8, e70268 (2013).
54. Culley, S., Albrecht, D., Jacobs, C., Pereira, P. M. & Leterrier, C. NanoJ-SQUIRREL: quantitative mapping and minimisation of super-resolution optical imaging artefacts. Preprint at: bioRxiv 10.1101/158279. (2017).
55. Serge, A., Bertaux, N., Rigneault, H. & Marguet, D. Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nat. Methods* 5, 687–694 (2008).
56. Churchman, L. S. & Spudich, J. A. Colocalization of fluorescent probes: accurate and precise registration with nanometer resolution. *Cold Spring Harb. Protoc.* 2012, 141–149 (2012).
57. Tarancón Díez, L. et al. Coordinate-based co-localization-mediated analysis of arrestin clustering upon stimulation of the C-C chemokine receptor 5 with RANTES/CCL5 analogues. *Histochem. Cell Biol.* 142, 69–77 (2014).
58. McDonald, J. H. & Dunn, K. W. Statistical tests for measures of colocalization in biological microscopy. *J. Microsc.* 252, 295–302 (2013).
59. Manders, E., Verbeek, F. J. & Aten, J. A. Measurement of co-localization of objects in dual-colour confocal images. *J. Microsc.* 169, 375–382 (1993).
60. Contrino, S. et al. modMime: flexible access to modENCODE data. *Nucleic Acids Res.* 40, D1082–D1088 (2012).
61. Serra, F., Bai, D., Filion, G. & Marti-Renom, M. A. Structural features of the fly chromatin colors revealed by automatic three-dimensional modeling. Preprint at: 10.1101/036764 (2016).
62. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* 9, 999–1002 (2013).

Acknowledgements

We thank Brian Beliveau, Hien Hoang, and Ting Wu for their help with oligopaint design. This research was supported by funding from the European Research Council under the 7th Framework Program (FP7/2010–2015, ERC grant agreement 266787 to M.N. and FP7/2007–2013, and ERC grant agreement 609889 to M.A.M.-R.), M.A.M.-R. and G.C. acknowledge support from the European Union’s Horizon 2020 research and innovation program under grant agreement 676556. This work has also benefited from support by the Labex EpiGenMed, an “Investments for the future” program, reference ANR-10-LABX-12-01, the Spanish Ministry of Economy and Competitiveness (BFU2013-47736-P to M.A.M.-R.), and from “Centro de Excelencia Severo Ochoa 2013–2017,” SEV-2012-0208 to the CRG. 3D-SIM experiments were performed at Montpellier Resource Imaging. We acknowledge the France-BioImaging infrastructure supported by the French National Research Agency (ANR-10-INBS-04, “Investments for the future”).

Author contributions

D.I.C., A.M.C.G. M.G. and M.N. designed the experiments and conducted the research. A.V. J.-B.F., M.G. and M.N. developed the software for image analysis. D.I.C, A.M.C.G., M.G. and M.N. designed the experiments and conducted the research. This research was supported by funding from the European Research Council under the 7th Framework Program (FP7/2010–2015, ERC grant agreement 266787 to M.N. and FP7/2007–2013, and ERC grant agreement 609889 to M.A.M.-R.), M.A.M.-R. and G.C. acknowledge support from the European Union’s Horizon 2020 research and innovation program under grant agreement 676556. This work has also benefited from support by the Labex EpiGenMed, an “Investments for the future” program, reference ANR-10-LABX-12-01, the Spanish Ministry of Economy and Competitiveness (BFU2013-47736-P to M.A.M.-R.), and from “Centro de Excelencia Severo Ochoa 2013–2017,” SEV-2012-0208 to the CRG. 3D-SIM experiments were performed at Montpellier Resource Imaging. We acknowledge the France-BioImaging infrastructure supported by the French National Research Agency (ANR-10-INBS-04, “Investments for the future”).

Additional information

Supplementary Information accompanies this paper at 10.1038/s41467-017-01962-x.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://www.nature.com/reprintsandpermissions/

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

© The Author(s) 2017