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

Diego I. Cattoni*[1], Andrés M. Cardozo Gizzi*[1], Mariya Georgieva*[1], Marco Di Stefano[2], Alessandro Valeri[1], Delphine Chamousset[1], Christophe Houbron[1], Stephanie Déjardin[1,3], Jean-Bernard Fiche[1], Inma González[3,5], Jia-Ming Chang[3,4], Thomas Sexton[3,6], Marc A. Marti-Renom[2], Frédéric Bantignies[3], Giacomo Cavalli[3], Marcelo Nollmann [1,#]

[1] Centre de Biochimie Structurale, CNRS UMR5048, INSERM U1054, Université de Montpellier, 29 rue de Navacelles, 34090 Montpellier, France; [2] CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 4, 08028 Barcelona, Spain; Gene Regulation, Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain; Universitat Pompeu Fabra (UPF), Barcelona, Spain; ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain; [3] Institut de Génétique Humaine, CNRS UMR 9002, Université de Montpellier, 141 rue de la Cardonille, 34396 Montpellier, France; [4] Department of Computer Science, National Chengchi University, 11605 Taipei City, Taiwan; [5] Current address: Epigenetics of Stem Cells, Department of Stem Cell and Developmental Biology, Institut Pasteur, CNRS UMR3738, 25 rue du Docteur Roux, 75015 Paris, France; [6] Current address: Institut de génétique et de biologie moléculaire et cellulaire, CNRS UMR 7104 - Inserm U 964, 67404 Illkirch, France

* Co-authors

[#] To whom correspondence should be addressed:

Marcelo Nollmann: marcelo.nollmann@cbs.cnrs.fr

Abstract

At the kilo- to mega-base pair scales, eukaryotic genomes are partitioned into self-interacting modules or topologically associated domains (TADs) that associate to form nuclear compartments. Here, we combined 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 found 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 were able to visualize nanometer-sized epigenetic domains at the single-cell level. In addition, absolute contact frequencies within and between TADs were 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.

Introduction

The multi-scale organization of eukaryotic genomes defines and regulates cellular identity and tissue-specific functions [1–3]. At the kilo-megabase scales, genomes are partitioned into self-interacting modules or topologically associated domains (TADs) [4–6]. 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 [9]. 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 reported that genomes also display stochasticity in their association with the nuclear lamina [12], in the formation of chromosome territory neighborhoods [13], and in gene kissing [14]. 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 behaviour 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 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, multi-scale organization can ensure the maintenance of stable transcriptional programs through cell division and during development.

Results

Multiple, low-frequency interactions mediate TAD assembly and insulation

A major mechanism for TAD formation in mammals involves the stable looping of TAD borders [8]. Stable looping between TAD borders was also recently proposed to be relevant for the maintenance of transcriptional programs during *Drosophila* development [7]. 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 Oligopaints libraries [15] and imaged using multi-color three-dimensional structured illumination microscopy (3D-SIM [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 foci, consistent with a high degree of homologous pairing independently of the ploidy of each cell type (Supplementary Fig. 1d) [18,19]. Distances between TAD borders were Gaussian distributed for all cell types (Figs. 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

100 by the polymer properties of the chromatin fiber.

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

114 In agreement with this model, absolute contact probabilities within TADs and
115 between their borders were similar (Fig. 1f and Supplementary Fig. 1k), with
116 inactive/repressed TADs displaying higher contact probabilities than active TADs (7 ±
117 1% *vs.* 2.7 ± 1%). Contact probabilities within TADs were in all cases considerably
118 higher than with neighboring TADs (Fig. 1f), indicating that stochasticity is locally
119 modulated at the TAD level. Of note, contacts across TAD borders were not
120 uncommon (~3%, Fig. 1f), implying frequent violations of boundary insulation at TAD
121 borders. These results indicate that confinement of chromatin into TADs may require
122 only small differences in absolute contact probabilities (~2 fold). Thus, condensation
123 of chromatin into TADs may arise from a multitude of low-frequency, yet specific,
124 intra-TAD contacts.

Infrequent, long-range interactions 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 non-consecutive 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 non-consecutive TAD borders were in all cases low (<9 %, Fig. 2b) and decreased monotonically with physical and genomic distance following a power law behaviour (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-type specific transcriptional programs. To test whether this tendency held genome-wide, we calculated the ratio between normalized Hi-C contact maps of embryos and S2 cells. For all chromosomes, embryos displayed a higher relative contact frequency than S2 cells below a few Mb (Fig. 2c and Supplementary Fig. 2d), in accordance with our microscopy results. Furthermore, the frequency of contacts between non-consecutive TAD borders genome-wide was similar to that of random genomic loci for both cell types (Fig. 2d). As the large

150 majority of TAD borders in *Drosophila* contain active chromatin [22,23], our results are
151 inconsistent with stable preferential looping of active borders [7,24], and rather indicate
152 that these contacts are rare or short-lived. This interpretation is consistent with the
153 transient assembly and disassembly of transcription clusters in human cells [25].

154 Next, we sought to determine if this modulation in contact probabilities resulted
155 from cell-type specific changes in the local folding properties of the chromatin fiber.
156 To this end, we measured the end-to-end distance (R_3) for active or
157 inactive/repressed chromatin domains of varying genomic ($L_{genomic}$) lengths. For all cell
158 and chromatin types, we observed a power-law scaling behavior ($R_3 \propto L_{genomic}^{\alpha}$,
159 Fig. 2e-f and Supplementary Fig. 2e-f) with scaling exponents being higher for active
160 than for repressed domains, consistent with previous measurements in Kc_{167} cells [26].
161 Theoretical studies of polymer physics suggest that the exponent of polymers with
162 random coil behaviour is ½, while that of an equilibrium globule is ⅓ (Mirny 2011).
163 Thus, our power-law exponents situate between these two extremes, suggesting an
164 intermediate behaviour. Remarkably, scaling exponents were considerably lower in
165 embryos than in S2 cells, for both types of chromatin. TAD border localization is
166 conserved between cell types [22,27], however, our results show that TAD conformation
167 and structural heterogeneity strongly depend on cell type. This cell-type specificity in
168 TAD organization may result from the interplay between the degree of chromatin
169 compaction and the frequency of stochastic long-range interactions.

170

171 *Modulated stochasticity regulates cell-specific chromosome compaction*

172 To quantitatively dissect stochasticity at larger genomic scales, we labeled 69
173 quasi-equidistant TAD borders encompassing 90% of chromosome 3R (Fig. 3a and
174 Supplementary Fig. 3a-b). Tens of foci were resolved in embryonic and S2 cells by

175 3D-SIM (Fig. 3a). The paired probability distance distribution *p(r)* between any two
176 foci exhibited moderate single-cell variations (Fig. 3b) but were considerably different
177 between cell types (Fig. 3b and Supplementary Fig.3c). The chromosome elongation
178 and mean volume, obtained from the maximum pairwise distance (D_{max}) and the
179 radius of gyration (R_g, Fig. 3c), decreased to almost half when comparing S2 and
180 late embryonic cells, while early embryonic cells adopted intermediate values (Fig.
181 3b).

182 From the number of labeled barriers (69) and the pairing frequency of
183 homologous chromosomes (Supplementary Fig. 1c), we can estimate a maximum of
184 90-100 resolvable foci/cell in the absence of any long-range interactions
185 (Supplementary Fig. 3). Our imaging results show an average of 89 ± 28 foci/cell for
186 S2 cells (Fig. 3b), confirming our predictions and consistent with a very low frequency
187 of long-range interactions for this cell type (see discussion in Supplementary Fig. 3).
188 Surprisingly, in early and late embryos the number of observed foci was considerably
189 reduced (51 ± 20 for early and 36 ± 13, respectively, and Fig. 3b), revealing higher
190 probabilities of long-range interactions for these cell-types. The lower number of foci
191 detected was not associated with the smaller volume of embryonic cell nuclei causing
192 the probes to be closer than the resolution limit of 3D-SIM microscopy
193 (Supplementary Fig. 3e). Furthermore, for each cell type, the number of foci
194 displayed very low or nonexistent correlation with nucleus size (Supplementary Fig.
195 3f).

196 Reinforcing these findings, changes in Hi-C contact frequency of S2 *vs.* late
197 embryo for the 69 TAD borders were notable in the sub-Mb scale (200-600 kb), and
198 they extended to genomic distances as high as ~10 Mb (Fig. 3d), suggesting that
199 changes in chromosome compaction between cell types arise from an increased

200 frequency of interactions affecting all genomic scales. All in all, these data indicate
201 that chromosome folding is highly variable, with mild, cell-type specific increases in
202 the probability of long-range contacts being sufficient to produce large changes in the
203 manner in which chromosomes occupy the nuclear space (Fig. 3c).

204 *Stochastic nanoscale organization of epigenetic marks*

205 Inter- and intra-chromosomal Hi-C maps have revealed that active and
206 repressed TADs may associate to form two types of compartments (namely A and B)
207 [9,28]. To study this higher-order level of organization in single cells and at the single-
208 molecule level, we immunolabeled active and repressive epigenetic marks (histones
209 H3K4me3 and H3K27me3, respectively) and performed multicolor direct stochastic
210 optical reconstruction microscopy (dSTORM) [29–31], a method that provides a higher
211 spatial resolution than 3D-SIM. dSTORM imaging revealed that active and repressive
212 histone marks distributed non-homogeneously across the cell nucleus, forming
213 discrete compartments of tens to hundreds of nanometers for all cell types (Fig. 4a
214 and Supplementary Fig. 4a). Repressed and active chromatin marks were strictly
215 segregated at the nanoscale for all cell types, as revealed by coordinate-based co-
216 localization analysis (aCBC[32], Fig. 4b). These findings were confirmed by
217 independent colocalization methods and by additional controls using doubly-labeled
218 nuclear factor and non-colocalizing epigenetic marks (Supplementary Figs. 4b-d).
219 Interestingly, active marks were often observed at borders of/or demarcating large
220 repressed compartments, mirroring their alternating one-dimensional genomic
221 distributions (Fig. 4c).

222 To investigate if active and repressed compartments also varied among cell
223 types and development, we resorted to one-color dSTORM using Alexa 647 as the

224 fluorophore of choice (results were similar when using other fluorophores,
225 Supplementary Fig. 4e). Compartments were detected using a Voronoi diagram-
226 based algorithm (Figs. 4d-e) [33]. In all cases, active compartments were smaller than
227 repressive compartments in agreement with two-color dSTORM observations (Figs.
228 4c-e and Supplementary Figs. 5a-b). Interestingly, for both marks, the number of
229 compartments and their sizes showed variations between single cells of the same
230 type (Supplementary Figs. 5c-d). To further evaluate if changes in compartment sizes
231 correlated with changes in local chromatin folding, we quantified the density of single
232 molecule detections in active and repressed compartments. Notably, the local density
233 of compartments was higher for both types of marks in embryonic cells than for S2
234 cells (Supplementary Fig. 6a), consistent with our previous findings (Figs. 2e-f) and
235 with compartment contact density from Hi-C counts (Supplementary Fig. 6b).

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

244 To quantify this phenomenon, we calculated the percentage of compartments
245 not accounted for by the distribution of epigenetic domains. This percentage of
246 clustered compartments was below <10% for embryonic cells and almost absent in
247 S2 cells (Fig. 4h). The latter is consistent with higher Hi-C contact frequency between
248 H3K27me3 domains in embryos that in S2 cells (Fig. 4i). Repressive and active

249 compartments showed different degrees of clustering (Figs. 4f-h), indicating that
250 stochasticity can be specifically modulated by transcriptional/epigenetic states. This
251 is likely due to the different mechanisms of clustering formation at play, such as
252 Polycomb regrouping of repressed genes [34] *vs.* transient interactions of active genes
253 [35,36]. It is important to note, however, that the large majority of compartments (~90%)
254 could be accounted for by the predicted distributions of epigenomic domains,
255 consistent with the majority of the epigenetic domains described by genome-wide
256 methods existing at the single-cell level. These results are consistent with the cell-
257 type specific higher-order organization of chromatin arising from stochastic contacts
258 between chromosomal regions harboring similar epigenetic marks, likely reflecting
259 cell-type specific transcriptional programs.

260 **Discussion**

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

274 In mammals, a large proportion of consecutive TAD borders are looped by
275 specific interactions apparently mediated by CTCF and cohesin [8,37,38]. Recent reports
276 suggested that this mechanism may also be at play in *Drosophila* [7,24]. Our results,
277 however, provide compelling evidence that looping of consecutive TADs borders in
278 *Drosophila* is rare at the single-cell level. These observations, supported by recent
279 studies showing that cohesin-enriched loop anchors in *Drosophila* are found within
280 TADs rather than at TAD borders [39,40], are against TAD boundaries being the bases
281 of stable chromatin loops. Thus, the lack of frequent interactions between TAD
282 borders could be consistent with a model where TADs arise from a dynamic balance
283 between cohesin-mediated loop extrusion [41], the blocking of that movement by
284 architectural proteins, and factors that may load or remove cohesin [42,43]. In
285 *Drosophila*, however, CTCF and cohesin are not found enriched at TAD borders.
286 Thus, we envision that other factors (e.g. Beaf-32 and CP190/chromator instead of
287 CTCF and cohesin) could play a role at looping and dynamically extruding distant
288 DNA fragments within the same TAD. In addition, active marks may help determine
289 the properties of TAD boundaries [23] while other epigenetic marks could play a role in
290 the formation of polycomb and inactive TADs [44]. Similar epigenetic mechanisms may
291 even play a role in TAD folding in mammals, consistent with the observation that
292 CTCF depletion leads only to minor changes in TAD organization [45]. Importantly, our
293 data provides quantitative estimates of the stochasticity and absolute frequencies of
294 interactions within TADs, imposing important constraints on any model of TAD
295 formation in *Drosophila*.

296 Recent reports suggested that TAD borders enriched in housekeeping genes
297 form stable 3D co-localization patterns that persist during development [7]. In contrast,
298 we found that 3D contacts between TAD borders are rare and highly stochastic in all

299 cell types investigated. These results are consistent with recent single-nucleus-Hi-C
300 studies reporting that TAD formation in highly stochastic in mammals [11], and with the
301 rapid association and dissociation of transcription foci [25] rather than with stable
302 transcription factories.

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

314

315

316 **Methods**

317 *Cell culture and embryonic tissue preparation*

318 *Drosophila* S2 cells were obtained from the *Drosophila* Genomics Resource Center.
319 S2 cells were grown in serum-supplemented (10%) Schneider's S2 medium at 25 °C.
320 Oregon-R w^{1118} fly stocks were maintained at room temperature with natural
321 light/dark cycle and raised in standard cornmeal yeast medium. Following a pre-
322 collection period of at least 1 hour, fly embryos were collected on yeasted 0.4%
323 acetic acid agar plates at and incubated at 25 C until they reached the desired
324 developmental stage: 2-3 h or 12-14 h (total developmental time) for early for late
325 embryos, respectively. Embryos were mechanically broken and immediately fixed as
326 described[46] by using 4% PFA in PBS for 10 min at room temperature (RT). S2 cells
327 were allowed to adhere to a poly-l-lysine coverslip for 30 min in a covered 35-mm cell
328 culture dish before 4% PFA fixation.

329

330 *Immunostaining*

331 Cells were permeabilized with 0.5% Triton X-100 for 10 min and blocked with 5% of
332 bovine serum albumin (BSA) for 15 min at RT. Primary antibodies anti-H3K27me3
333 (pAb-195-050, Diagenode and ab6002, Abcam), anti-H3K4me3 (cat#04-745,
334 Millipore and ab1012, Abcam), anti-Polycomb[47] and anti-Beaf-32[48] (made from rabbit
335 by Eurogentec) were coupled to Alexa Fluor 647 or Cy3b as described elsewhere[32].
336 Antibodies were used at a final concentration of 10 µg ml^{-1} in PBS and 1% BSA.
337 Coverslips were incubated overnight at 4 °C in a humidified chamber and washed 3
338 times with PBS before introducing fiducial markers diluted 1/4000 (Tetraspeck,
339 #10195142, FisherScientific). Coverslips were mounted on slides with 100 µl wells
340 (#2410, Glaswarenfabrik Karl Hecht GmbH & Co KG) in dSTORM buffer composed

341 of PBS, glucose oxidase (G7141-50KU, Sigma) at 2.5 mg ml^{-1}, catalase at 0.2 mg ml^{-1}
342 (#C3155-50MG, Sigma), 10% glucose and 50 mM of β-mercaptoethylamine (MEA,
343 #M9768-5G, Sigma). Coverslips were sealed with duplicating silicone (Twinsil,
344 Rotec).
345
346 *Oligopaint libraries*
347 Oligopaint libraries were constructed from the Oligopaint public database
348 (http://genetics.med.harvard.edu/oligopaints). All libraries consisted of 42mer
349 sequences discovered by OligoArray2.1 run with the following settings: -n 30 -l 42 -L
350 42 -D 1000 -t 80 -T99 -s 70 -x 70 -p 35 -P 80 -m 'GGGG;CCCC;TTTTT;AAAAA' -g
351 44. Oligonucleotide for libraries 1-18 and BX-C were ordered from CustomArray
352 (Bothell, WA). The procedure used to synthesize Oligopaint probes is described
353 below. Chr3R-69 borders oligonucleotides were purchased from MYcroarray (Ann
354 Arbour, MI). Oligopaint probes for this library were synthesized using the same
355 procedure as for the other libraries except for the initial emulsion PCR step.
356 Secondary, fluorescently-labeled oligonucleotides were synthesized by Integrated
357 DNA Technologies (IDT; Coralville, IA for Alexa488) and by Eurogentec (Angers,
358 France for Cy3b). See Supplementary Table 1 for a list of Oligopaint probe sets used
359 for libraries 1-18. Sequences for secondary oligonucleotides and PCR primers are
360 described below (Supplementary Tables 2-4). Details for the methods used for probe
361 synthesis are provided in Online Methods.
362
363 *Fluorescence in situ hybridization (FISH)*
364 To prepare sample slides containing fixed S2 cells for FISH, S2 cells were allowed to
365 adhere to a poly-l-lysine coverslip for 1h in a covered 35-mm cell culture dish at 25C.

366 Slides were then washed in PBS, fixed 4% paraformaldehyde (PFA) for 10 min,
367 rinsed 3 times for 5 min in PBS, permeabilized 10min with 0.5% Triton, rinsed in PBS,
368 incubated with 0.1M HCl for 10min, washed in 3 times for 1 min with 2X saline-
369 sodium citrate - 0.1% Tween-20 (2XSSCT) and incubated in 2XSSCT/50%
370 formamide (v/v) for at least 30min. Then, probes were prepared by mixing 20µl of
371 hybridization buffer FHB (50% Formamide, 10% Dextransulfat, 2X SSC, Salmon
372 Sperm DNA 0.5 mg ml^{-1}), 0.8 µL of RNAse A, 30 pmol of primary probe and 30 pmol
373 of secondary oligo. 12µl of this mix were added to a slide before adding and sealing
374 with rubber cement the coverslips with cells onto the slide. Probes and cells are
375 finally co-denaturated 3 min at 78 □C before hybridization overnight at 37 □C. The
376 next day, the slides were washed for 3 times for 5 min in 2X SSC at 37 □C, then for 3
377 times for 5 min in 0.1X SSC at 45 □C. Finally, they were stained with 0.5 µg ml^{-1} of
378 DAPI for 10 min, washed with PBS, mounted in Vectashield and sealed with nail
379 polish. For a more detailed protocol, see [49].
380

381 *Image acquisition and post-processing of 3D-SIM data*

382 Samples were prepared as described above and mounted on an OMX V3
383 microscope (Applied Precision Inc.) equipped with a 100X/1.4 oil PlanSApo objective
384 (Olympus) and three emCCD cameras. 405 nm, 488 nm and 561 nm excitation
385 lasers lines were used to excite DAPI, Alexa488 and Cy3b, respectively. Each
386 channel was acquired sequentially. A transmission image was also acquired to
387 control for cell morphology. For each channel, a total of 1455 images made of 97
388 different Z-planes separated by 125 nm were acquired, in order to acquire a stack of
389 12 µm. Three different angles (60°, 0° and +60°) as well as five phase steps were
390 used to reconstruct 3D-SIM images using softWoRx v5.0 (Applied Precision Inc.).

391 Final voxel size was 39.5 nm in the lateral (xy) and 125 nm in the axial (z) directions,
392 respectively, for a final 3D stack volume of ~40 x 40 x 12 μm. Multicolor TetraSpeck
393 beads (100 nm in diameter, Invitrogen) were used to measure x, y and z offsets,
394 rotation about the z-axis and magnification differences between fluorescence
395 channels. These corrections were applied to the reconstructed images. The same
396 beads were used to validate the reconstruction process ensuring a final resolution of
397 ~120 nm in xy and ~300 nm in z at 525 nm of emission wavelength. 3D-SIM raw and
398 reconstructed images were analyzed with SIMCheck ImageJ Plug-in [50]. Acquisition
399 parameters were optimized to obtain the best signal-to-noise ratio avoiding
400 photobleaching between the different angular, phase, and axial acquisitions.
401

402 *3D nuclei segmentation from 3D-SIM data*

403 3D-SIM images were analyzed employing homemade software written in Matlab. In
404 order to identify nuclear shells, nuclei are first segmented by manually selecting
405 rectangular regions of interest (ROIs) of the DAPI signal in the XY-plane and keeping
406 all the Z-planes and then a low-pass filter is applied to the DAPI intensities so that
407 only the large-scale information (i.e. nuclear shape) is kept. For each plane of the 3D
408 ROIs, an intensity threshold is computed as described by Snell et al. [51] in order to
409 distinguish voxels inside or outside the nucleus. The average intensity threshold
410 calculated from the threshold of the single planes is used to identify the complete
411 nuclear shell. After nuclei segmentation, foci were identified by calculating, for each
412 channel separately, the maximum entropy threshold of the fluorescence intensities in
413 the 3D ROIs. By using the intensity thresholds the 3D ROIs are finally binarized
414 (voxels above threshold are set to 1 while the others to 0) and the different foci
415 identified as groups of connected voxels. From the group of connected voxels the

416 center of mass was estimated with subpixel resolution. Distance between TBs was
417 estimated as the linear distance between the closest foci imaged in two different
418 emission channels.
419
420 *Image acquisition of two-color dSTORM data*
421 Super-resolution experiments were carried out in a custom-made inverted
422 microscope employing an oil-immersion objective (Plan-Apocromat, 100x, 1.4NA oil
423 DIC, Zeiss) mounted on a z-axis piezoelectric stage (P-721.CDQ, PICOF, PI). For 2D
424 imaging, a 1.5x telescope was used to obtain a final imaging magnification of 150 fold
425 corresponding to a pixel size of 105 nm. Three lasers were used for excitation/photo-
426 activation: 405 nm (OBIS, LX 405-50, Coherent Inc.), 488 nm (OBIS, LX 488-50,
427 Coherent Inc.), 561 nm (OBIS, LX 561-50, Coherent Inc.), and 640 nm (OBIS, LX
428 640-100, Coherent Inc.). Laser lines were expanded, and coupled into a single beam
429 using dichroic mirrors (427, 552 and 613 nm LaserMUXTM, Semrock). An acousto-
430 optic tunable filter (AOTFnc-400.650-TN, AA opto-electronics) was used as to
431 modulate laser intensity. Light was circularly polarized using an achromatic quarter
432 wave plate (QWP). Two achromatic lenses were used to expanded the excitation
433 laser and an additional dichroic mirror (zt405/488/561/638rpc, Chroma) to direct it
434 towards the back focal plane of the objective. Fluorescence light was spectrally
435 filtered with emission filters (ET525/50m, ET600/50m and ET700/75m, Chroma
436 Technology) and imaged on an EMCCD camera (iXon X3 DU-897, Andor
437 Technologies). The microscope was equipped with a motorized stage (MS-2000,
438 ASI) to translate the sample perpendicularly to the optical axis. To ensure the stability
439 of the focus during the acquisition, a home-made autofocus system was built. An 785
440 nm laser beam (OBIS, LX 785-50, Coherent Inc.) was expanded twice and directed

441 towards the objective lens by a dichroic mirror (z1064rdc-sp, Chroma). The reflected
442 IR beam was redirected following the same path than the incident beam and guided
443 to a CCD detector (Pixelfly, Cooke) by a polarized beam splitter cube (PBS). Camera,
444 lasers and filter wheel were controlled with software written in Labview [52].
445 For image acquisition, on average of 30,000 frames (per detection channel in two-
446 color acquisitions) were recorded at a rate of 50 ms/frame. Continuous excitation and
447 activation was employed for all fluorophores employed in this work with powers as
448 follows: 1 kW cm^{-2} at 641 nm (for AF647), 0.8–1.2 kW cm^{-2} at 561 nm (for mEos2),
449 and 0–0.1 kW cm^{-2} at 405 nm for activation. The intensity of activation was
450 progressively increased throughout the acquisition to ensure a constant amount of
451 simultaneously emitting fluorophores within the labeled structures. These excitation
452 powers were optimized to ensure single-molecule detection, despite the large nuclear
453 density of epigenetic compartments. More technical details and the method used to
454 ensure single-molecule detection are described elsewhere [32,53].
455

456 *Post-processing and Analysis of two-color dSTORM data*

457 Unless stated otherwise, all homemade software and routines were developed in
458 Matlab. Before further processing, super-resolution image quality was quantitatively
459 assessed by using NanoJ-SQUIRREL [54]. Next, single-molecule localizations were
460 obtained by using Multiple Target Tracing (MTT) [55]. Localization coordinates were
461 further processed using SMLM_2C, a custom software written in Matlab [32].
462 Fluorescent beads were used to correct for drift and chromatic aberrations. Lateral
463 drift was corrected with 5 ± 3 nm precision as previously described [52]. Chromatic
464 aberration correction was performed as described in [32,56]. Samples with abnormal
465 drift or lesser precision of drift or chromatic aberration correction were discarded.

466 Clustering of localizations was performed as described in Cattoni et al. [53].
467 Colocalization of single-molecule detections was performed by using a custom
468 implementation of the Coordinate-based colocalization (CBC) analysis [57] adapted for
469 whole-cell automated analysis [32]. Three additional methods were employed as
470 controls: pixel, Pearson, and Manders correlation. For the latter, two-color digital
471 images were reconstructed from the localization using standard procedures [52] and
472 then used to plot the correlation between pixel intensities (pixel correlation analysis),
473 or to calculate the Pearson or Manders correlation coefficients [58,59].

474

475 *Analysis of one-color dSTORM data*

476 Single-molecule localizations are converted into a Voronoi diagram using a
477 modified version of the Voronoi tesselation algorithm of Levet et al. [33]. Compartment
478 segmentation is directly calculated from the Voronoi diagram using three steps. First,
479 densities of each polygon are calculated as the inverse of their area. Densities are
480 then thresholded using the general criteria of Levet et al. [33]. Using this criterion, in
481 which the threshold is determined by the average localization density, a random
482 distribution of localizations did not provide any segmented polygon. Finally, polygons
483 that have a density higher than the threshold and that are touching each other are
484 merged to define the compartment outline. Compartment sizes are obtained by
485 interpolating each segmented compartment on a grid of 5 nm size and calculating
486 their equivalent diameter using standard morphological operations. Probability
487 density functions in compartment size histograms are calculated such that the area of
488 each bar is the relative number of observations and that the sum of the bar areas is
489 equal to 1.

490

491 *Analysis of genome-wide data*

492 Chromatin states were defined according to the enrichment in the percentages
493 of H3K4me3 and H3K27me3/PC as described in Supplementary Fig. 1. Calculation
494 of the genomic size distributions of H3K27me3 and H3K4me3 domains
495 (Supplementary Fig. 6) was performed as follows: (1) ChIP-chip/seq computed peaks
496 were downloaded from ModEncode (ftp://data.modencode.org/D.melanogaster/) [60].
497 Datasets used are described in the Online Methods (Supplementary Table 5). (2)
498 Peak positions and intensities were used to resample the data and produce a
499 continuous signal as a function of genomic position. (3) This signal was thresholded
500 with a threshold of 0.1 of the log of the maximum intensity signal, ensuring that even
501 peaks with very low intensity were retained. (4) Domains were defined as continuous
502 segments extending for more than 2 bp with non-zero intensity. (5) Domains that
503 were closer than 1 kb were fused together. This procedure was robust to calculate
504 domain size distributions above 3kb (Supplementary Fig. 6). (6) Finally, we estimated
505 physical domain sizes from their genomic length as follows. The size of each
506 genomic domain in bp was converted into nanometers using the empirical power law
507 that relates genomic sizes to physical distances (Figs. 2e-f). The parameters of the
508 power-law depended on chromatin type (active or repressed) and on cell type (S2,
509 early or late embryo), and are shown in the insets of Figs. 2e-f. After repeating this
510 process for all genomic domain sizes, we obtained the distribution of domains sizes
511 in nm for a specific chromatin type and cell type (Fig. 4f-g).

512 Clustering of domains of different epigenetic marks was defined as the ratio
513 between the number of clusters of sizes larger than 150 nm obtained from Chip-seq
514 *vs.* microscopy imaging. Changes in this threshold did not affect our main
515 conclusions.

516

517 *In-situ Hi-C data processing and normalization*

518 Hi-C data was processed using an in-house pipeline based on TADbit [61]. First, quality
519 of the reads was checked using the *quality_plot()* function in TADbit, which is similar
520 to the tests performed by the FastQC program with adaptations for Hi-C datasets.
521 Next, the reads are mapped following a fragment-based strategy as implemented in
522 TADbit where each side of the sequenced read was mapped in full length to the
523 reference genome (dm3). After this step, if a read was not uniquely mapped, we
524 assumed the read was chimeric due to ligation of several DNA fragments. We next
525 searched for ligation sites, discarding those reads in which no ligation site was found.
526 Remaining reads were split as often as ligation sites were found. Individual split read
527 fragments were then mapped independently. Next, we used the TADbit filtering
528 module to remove non-informative contacts and to create contact matrices. From the
529 resulting contact matrices, low quality bins (those presenting low contacts numbers)
530 were removed as implemented in TADbit's *filter_columns()* function. Next, the
531 matrices were normalized using the ICE algorithm [62]. The normalization iterations
532 stopped when the biases were diverting less than 10% of the previous values or a
533 max of 10 iterations. Finally, all matrices were corrected to achieve an average
534 content of one interaction per cell. All parameters in TADbit were kept at default
535 values.

536 The resulting late-embryo and S2 Hi-C interaction maps (at 10kb resolution) of the
537 different replicates for each experiment were highly correlated (correlation coefficient
538 from genomic distances ranging from 10Kb to 20Mb were 0.99 to 0.75 and 0.95 to
539 0.45, respectively) and thus were further merged into the final datasets with more
540 than 282 and 210 million valid pairs each (Supplementary Table 6).

Data availability

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

548 **References**

549 1. Sexton, T. & Cavalli, G. The Role of Chromosome Domains in Shaping the
550 Functional Genome. *Cell* **160,** 1049–1059 (2015).
551 2. Gilbert, D. M. & Fraser, P. Three Dimensional Organization of the Nucleus:
552 adding DNA sequences to the big picture. *Genome Biol.* **16,** 181 (2015).
553 3. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers
554 and their regulatory landscapes. *Nature* **502,** 499–506 (2013).
555 4. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by
556 analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
557 5. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-
558 inactivation centre. *Nature* **485,** 381–385 (2012).
559 6. Sexton, T. *et al.* Three-dimensional folding and functional organization principles
560 of the Drosophila genome. *Cell* **148,** 458–472 (2012).
561 7. Hug, C. B., Grimaldi, A. G., Kruse, K. & Vaquerizas, J. M. Chromatin Architecture
562 Emerges during Zygotic Genome Activation Independent of Transcription. *Cell*
563 **169,** 216–228.e19 (2017).
564 8. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution
565 Reveals Principles of Chromatin Looping. *Cell* **162,** 687–688 (2015).
566 9. Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Exploring the three-dimensional
567 organization of genomes: interpreting chromatin interaction data. *Nat. Rev.*
568 *Genet.* **14,** 390–403 (2013).
569 10. Giorgetti, L. *et al.* Predictive polymer modeling reveals coupled fluctuations in
570 chromosome conformation and transcription. *Cell* **157,** 950–963 (2014).
571 11. Flyamer, I. M. *et al.* Single-nucleus Hi-C reveals unique chromatin reorganization
572 at oocyte-to-zygote transition. *Nature* (2017). doi:10.1038/nature21711

573 12. Kind, J. & van Steensel, B. Stochastic genome-nuclear lamina interactions:
574 modulating roles of Lamin A and BAF. *Nucleus* **5,** 124–130 (2014).
575 13. Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harb. Perspect.*
576 *Biol.* **2,** a003889 (2010).
577 14. Lanctôt, C., Cheutin, T., Cremer, M., Cavalli, G. & Cremer, T. Dynamic genome
578 architecture in the nuclear space: regulation of gene expression in three
579 dimensions. *Nat. Rev. Genet.* **8,** 104–115 (2007).
580 15. Beliveau, B. J. *et al.* Single-molecule super-resolution imaging of chromosomes
581 and in situ haplotype visualization using Oligopaint FISH probes. *Nat. Commun.*
582 **6,** 1–13 (2015).
583 16. Gustafsson, M. G. L. Surpassing the lateral resolution limit by a factor of two
584 using structured illumination microscopy. *J. Microsc.* **198,** 82–87 (2000).
585 17. Gustafsson, M. G. L. *et al.* Three-dimensional resolution doubling in wide-field
586 fluorescence microscopy by structured illumination. *Biophys. J.* **94,** 4957–4970
587 (2008).
588 18. Fung, J. C., Marshall, W. F., Dernburg, A., Agard, D. A. & Sedat, J. W.
589 Homologous chromosome pairing in Drosophila melanogaster proceeds through
590 multiple independent initiations. *J. Cell Biol.* **141,** 5–20 (1998).
591 19. Zhang, Y. *et al.* Expression in aneuploid Drosophila S2 cells. *PLoS Biol.* **8,**
592 e1000320 (2010).
593 20. Ramirez, F. *et al.* High-resolution TADs reveal DNA sequences underlying
594 genome organization in flies. *bioRxiv* 115063 (2017). doi:10.1101/115063
595 21. Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin Domains: The Unit of
596 Chromosome Organization. *Mol. Cell* **62,** 668–680 (2016).
597 22. Hou, C., Li, L., Qin, Z. S. & Corces, V. G. Gene density, transcription, and

598 insulators contribute to the partition of the Drosophila genome into physical
599 domains. *Mol. Cell* **48,** 471–484 (2012).
600 23. Ulianov, S. V. *et al.* Active chromatin and transcription play a key role in
601 chromosome partitioning into topologically associating domains. *Genome Res.*
602 **26,** 70–84 (2016).
603 24. Li, L. *et al.* Widespread rearrangement of 3D chromatin organization underlies
604 polycomb-mediated stress-induced silencing. *Mol. Cell* **58,** 216–231 (2015).
605 25. Cisse, I. I. *et al.* Real-time dynamics of RNA polymerase II clustering in live
606 human cells. *Science* **341,** 664–667 (2013).
607 26. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin folding
608 for different epigenetic states. *Nature* **529,** 418–422 (2016).
609 27. Dekker, J. & Heard, E. Structural and functional diversity of Topologically
610 Associating Domains. *FEBS Lett.* **589,** 2877–2884 (2015).
611 28. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions
612 reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).
613 29. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic
614 optical reconstruction microscopy (STORM). *Nat. Methods* **3,** 793–795 (2006).
615 30. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-Dimensional Super-
616 Resolution Imaging by Stochastic Optical Reconstruction Microscopy. *Science*
617 **319,** 810–813 (2008).
618 31. Heilemann, M. *et al.* Subdiffraction-Resolution Fluorescence Imaging with
619 Conventional Fluorescent Probes. *Angew. Chem. Int. Ed.* **47,** 6172–6176 (2008).
620 32. Georgieva, M. *et al.* Nanometer resolved single-molecule colocalization of
621 nuclear factors by two-color super resolution microscopy imaging. *Methods* **105,**
622 44–55 (2016).

623 33. Levet, F. *et al.* SR-Tesseler: a method to segment and quantify localization-
624 based super-resolution microscopy data. *Nat. Methods* **12,** 1065–1071 (2015).
625 34. Bantignies, F. & Cavalli, G. Polycomb group proteins: repression in 3D. *Trends*
626 *Genet.* **27,** 454–464 (2011).
627 35. Chuang, C.-H. & Belmont, A. S. Close encounters between active genes in the
628 nucleus. *Genome Biol.* **6,** 237 (2005).
629 36. Fraser, P. & Bickmore, W. Nuclear organization of the genome and the potential
630 for gene regulation. *Nature* **447,** 413–417 (2007).
631 37. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and
632 domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci.*
633 *U. S. A.* **112,** E6456–65 (2015).
634 38. Vietri Rudan, M. *et al.* Comparative Hi-C reveals that CTCF underlies evolution
635 of chromosomal domain architecture. *Cell Rep.* **10,** 1297–1309 (2015).
636 39. Cubeñas-Potts, C. *et al.* Different enhancer classes in Drosophila bind distinct
637 architectural proteins and mediate unique chromatin interactions and 3D
638 architecture. *Nucleic Acids Res.* (2016). doi:10.1093/nar/gkw1114
639 40. Eagen, K. P., Aiden, E. L. & Kornberg, R. D. Polycomb-mediated chromatin
640 loops revealed by a subkilobase-resolution chromatin interaction map. *Proc.*
641 *Natl. Acad. Sci. U. S. A.* (2017). doi:10.1073/pnas.1701291114
642 41. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion.
643 *Cell Rep.* **15,** 2038–2049 (2016).
644 42. Haarhuis, J. H. I. *et al.* The Cohesin Release Factor WAPL Restricts Chromatin
645 Loop Extension. *Cell* **169,** 693–707.e14 (2017).
646 43. Barrington, C., Finn, R. & Hadjur, S. Cohesin biology meets the loop extrusion
647 model. *Chromosome Res.* **25,** 51–60 (2017).

648 44. Jost, D., Carrivain, P., Cavalli, G. & Vaillant, C. Modeling epigenome folding:
649 formation and dynamics of topologically associated chromatin domains. *Nucleic*
650 *Acids Res.* **42,** 9553–9561 (2014).
651 45. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and
652 gene expression in human cells. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 996–1001
653 (2014).
654 46. Cheutin, T. & Cavalli, G. Progressive polycomb assembly on H3K27me3
655 compartments generates polycomb bodies with developmentally regulated
656 motion. *PLoS Genet.* **8,** e1002465 (2012).
657 47. Martinez, A.-M., Colomb, S., Déjardin, J., Bantignies, F. & Cavalli, G. Polycomb
658 group-dependent Cyclin A repression in Drosophila. *Genes Dev.* **20,** 501–513
659 (2006).
660 48. Emberly, E. *et al.* BEAF regulates cell-cycle genes through the controlled
661 deposition of H3K9 methylation marks into its conserved dual-core binding sites.
662 *PLoS Biol.* **6,** 2896–2910 (2008).
663 49. Bantignies, F. & Cavalli, G. Topological organization of Drosophila hox genes
664 using DNA fluorescent in situ hybridization. in **1196,** 103–120 (Springer New
665 York, 2014).
666 50. Ball, G. *et al.* SIMcheck: a Toolbox for Successful Super-resolution Structured
667 Illumination Microscopy. *Sci. Rep.* **5,** 15915 (2015).
668 51. Snell, V., Kittler, J. & Christmas, W. Segmentation and Shape Classification of
669 Nuclei in DAPI Images. *British Machine Vision Conference Workshop* (2011).
670 52. Fiche, J.-B. *et al.* Recruitment, assembly, and molecular architecture of the
671 SpoIIIE DNA pump revealed by superresolution microscopy. *PLoS Biol.* **11,**
672 e1001557 (2013).

673 53. Cattoni, D. I., Fiche, J.-B., Valeri, A., Mignot, T. & Nollmann, M. Super-resolution
674 imaging of bacteria in a microfluidics device. *PLoS One* **8,** e76268 (2013).
675 54. Culley, S., Albrecht, D., Jacobs, C., Pereira, P. M. & Leterrier, C. NanoJ-
676 SQUIRREL: quantitative mapping and minimisation of super-resolution optical
677 imaging artefacts. *bioRxiv* (2017).
678 55. Serge, A., Bertaux, N., Rigneault, H. & Marguet, D. Dynamic multiple-target
679 tracing to probe spatiotemporal cartography of cell membranes. *Nat. Methods* **5,**
680 687–694 (2008).
681 56. Churchman, L. S. & Spudich, J. A. Colocalization of fluorescent probes: accurate
682 and precise registration with nanometer resolution. *Cold Spring Harb. Protoc.*
683 **2012,** 141–149 (2012).
684 57. Tarancón Díez, L. *et al.* Coordinate-based co-localization-mediated analysis of
685 arrestin clustering upon stimulation of the C-C chemokine receptor 5 with
686 RANTES/CCL5 analogues. *Histochem. Cell Biol.* **142,** 69–77 (2014).
687 58. McDonald, J. H. & Dunn, K. W. Statistical tests for measures of colocalization in
688 biological microscopy. *J. Microsc.* **252,** 295–302 (2013).
689 59. Manders, E., Verbeek, F. J. & Aten, J. A. Measurement of co-localization of
690 objects in dual-colour confocal images. *J. Microsc.* **169,** 375–382 (1993).
691 60. Contrino, S. *et al.* modMine: flexible access to modENCODE data. *Nucleic Acids*
692 *Res.* **40,** D1082–8 (2012).
693 61. Serra, F., Baù, D., Filion, G. & Marti-Renom, M. A. Structural features of the fly
694 chromatin colors revealed by automatic three-dimensional modeling. (2016).
695 doi:10.1101/036764
696 62. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of
697 chromosome organization. *Nat. Methods* **9,** 999–1003 (2012).

698

Acknowledgements

We thank Brian Beliveau, Hien Hoang and Ting Wu for help with oligopaints design. This research was supported by funding from the European Research Council under the 7th Framework Program (FP7/2010-2015, ERC grant agreement 260787 to M.N. and FP7/2007-2013, ERC grant agreement 609989 to M.A.M-R.). M.A.M-R and G.C. acknowledge support from the European Union's Horizon 2020 research and innovation programme under grant agreement 676556. This work has benefited also 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 contribution

D.I.C, A.M.C.G. M.G. and M.N. designed experiments and conducted research. A.V. J.B.F., M.G. and M.N. developed software for image analysis. D.I.C, A.M.C.G., A.V. and F.B. designed Oligopaint probes. F.B. performed fly handling. M.D. and M.A.M-R. performed Hi-C bioinformatics analysis. D.C., C.H. and S.D. synthesized and purified oligopaints libraries and performed S2 cells handling. D.I.C, A.M.C.G. M.G., G.C. and M.N. wrote the manuscript. All the authors reviewed and commented the data.

723 **Competing financial interests**

724 The authors declare no competing financial interests.

725

726

727 **Figure legends**

728 **Figure 1. TAD organization arises from modulation of stochasticity.**

729 **(a)** Top, region of Hi-C contact matrix of chromosome 2L. Black dotted line
730 demarcates a single TAD and pink and cyan boxes represent the Oligopaint
731 labelled TAD borders (TB). Chromatin epigenetic state is indicated at the
732 bottom using the color code of panel b. Bottom, representative three-color 3D-
733 SIM image in two different orientations. DAPI, TB2 and TB3 are shown in gray,
734 pink and cyan respectively. Scalebar = 1 μm for main image. 5x amplification
735 of selected region is depicted in the insets.

736 **(b)** Oligopaint libraries in chromosomes 2L and 3R employed in this study (TB1-
737 16 at TAD borders and IT17-19 within TADs). Colored boxes display the
738 chromatin type of TADs as defined in Supplementary Fig. 1a-b. Red: active,
739 blue: repressed, black: inactive. Dotted colored lines indicate the combinations
740 of libraries measured.

741 **(c)** 3D distance distributions between TB2-TB2 and TB2-TB3. Mean colocalization
742 resolution, estimated from two-color labelling of a single border (40 nm,
743 vertical blue dashed line). Blue and black solid lines represent Gaussian
744 fittings. Absolute contact probability between libraries was obtained from the
745 integral of the area of the Gaussian fitting (shaded gray) below 120 nm (see
746 Supplementary Fig. 1e). N = 161 and 556 for TB2-TB2 and TB2-TB3
747 respectively, from more than three biological replicates.

748 **(d)** Absolute contact probability between consecutive borders *vs.* genomic
749 distance. Chromatin state of TADs is color-coded as defined in panel 1b. Error
750 bars represent SEM.

751 **(e)** Normalized Hi-C counts between consecutive TAD borders (circles) and
752 random loci (solid gray line) as a function of genomic distance for S2 and late
753 embryonic cells. Matrix resolution = 10kb. Two biological replicates for each
754 cell-type were performed.
755 **(f)** Schematic representation of contact probability between and within TADs
756 (solid colored lines) for late embryo and S2 cells at the chromosomal region
757 shaded in panel b. Sizes of TADs (grey shaded triangles) is proportional to
758 genomic length (scale bar on top). Chromatin type is indicated at the bottom.
759 Thickness of the lines and color indicate absolute contact probability. Dotted
760 lines indicate inter-TADs contacts. Early embryo measurements are depicted
761 in Supplementary Fig. 1j. Numbers of cells for each combination is provided in
762 Supplementary Fig. 1f-h.
763
764 **Figure 2. Long-range absolute contact probability is specifically modulated for**
765 **each cell type**
766 **(a)** Left, a schematic representation of pairwise distance measurements between
767 consecutive and non-consecutive borders, with color code and positions as in
768 Fig. 1b. Right, normalized Hi-C counts *vs*. microscopy absolute contact
769 probability for consecutive and non-consecutive domain borders for embryo
770 and S2 cells. Solid black and red lines represent exponential and a power-law
771 fits respectively. Matrix resolution = 10kb. N for microscopy pairwise
772 measurements is provided in Supplementary Figs. 1f-h. N = 2 for Hi-C data,
773 from at least three and two biological replicates, respectively.
774 **(b)** Absolute contact probability *vs.* mean physical distance between probes for
775 consecutive and non-consecutive TAD borders (filled circles). Solid lines

776 represent power-law fittings with scaling exponent described in Supplementary
777 Fig. 2b. Triangles represent measurements within TADs.

778 **(c)** Matrix of relative frequency of normalized Hi-C counts for late embryo *vs.* S2
779 cells for chromosome 2L. Contact frequency ratio is color coded according to
780 scale bar. Matrix resolution = 50kb. N=4, biological replicates.

781 **(d)** Log-log plot of normalized Hi-C counts between TADs borders *vs.* genomic
782 distance for embryo and S2 cells. Solid lines represent the average contact
783 frequency for randomly chosen positions in the genome. Matrix resolution =
784 10kb. N=2, biological replicates.

785 **(e-f)** Log-log plot of the mean physical distance vs. genomic length for (e) active
786 and (f) inactive/repressed chromatin domains for different cell types. Mean
787 distance values were normalized by the pre-exponential factor from the
788 power-law fit of each dataset (Supplementary Figs. 2d-e). Solid lines show the
789 power-law fits, with the scaling exponents β shown in the panel. Circles and
790 triangls are depicted as described in panel 2b. Error bars represent SEM. N >
791 140 for each data point, from more than three biological replicates (see
792 Supplementary Fig. 1)

793

794 **Figure 3. Cell-type specific frequency of long-range contacts defines**
795 **chromosome folding in 3D space.**

796 **(a)** Left, schematic representation of 69 domain borders labelled by a single
797 Oligopaint library (Lib-69) in Chr. 3R. Each probe spanned ~20 kb, probes
798 were separated by 320 kb in average (see Supplementary Fig. 3a-b). Right,
799 representative two-color 3D-SIM images for all studied cell types. DAPI signal
800 (white) and Lib-69 (pink) are shown. Scalebar = 200 nm.

801 **(b)** Left panel, single-cell probability distance distribution *p(r)* between all pairs of
802 foci imaged by 3D-SIM. White line represents the population averaged *p(r)*
803 frequency. Detailed R_g and D_{max} values are shown in Supplementary Fig. 3.
804 D_{max} is defined as the distance that comprises <97% of the area under the *p(r)*
805 function. Right panel, number of foci per cell for each condition with mean
806 population values shown as solid vertical lines and indicated above. N = 180,
807 from more than three biological replicates.
808 **(c)** Schematic representation of the chromosome structure for each cell type.
809 Solid grey line represent the chromatin fiber and pink circles represent domain
810 borders with sizes proportional to the number of regrouped borders.
811 **(d)** Hi-C contact frequencies of S2 *vs.* late embryo cells for all the pairwise
812 combinations of the 69 borders. Solid red line represents the relation expected
813 if frequencies of interactions between the 69 borders were equal between cell-
814 types. Insets depicts chromosome 3R and different combinations of genomic
815 distances and frequencies of interaction between borders. Matrix resolution =
816 50kb. N=4, from at least three biological replicates.
817

818 **Figure 4. Chromatin reorganization between cell types is modulated by**
819 **stochastic clustering between epigenetic domains.**
820 **(a)** Two-color dSTORM image of active (H3K4me3, blue) and repressive
821 (H3K27me3, red) chromatin marks in a representative S2 cell. Images of early
822 and late embryo are displayed in Supplementary Fig. 4a and panel c. Scalebar
823 = 1 μm

824 **(b)** Quantification of co-occurrence (CA>0.5) between active and repressive
825 chromatin using aCBC [32]. Boxplots of CAs for H3K4me3 and H3K27me3 are
826 shown in the upper panel and lower panels, respectively.

827 **(c)** Representative zoomed images of two-color dSTORM for the three cell-types
828 investigated. Black arrows indicate the localization of small active chromatin
829 domains in the periphery of large repressive domains. Lower panel display
830 active and repressive marks Chip-Seq enrichment profiles for late embryo.
831 Scalebar = 200 nm.

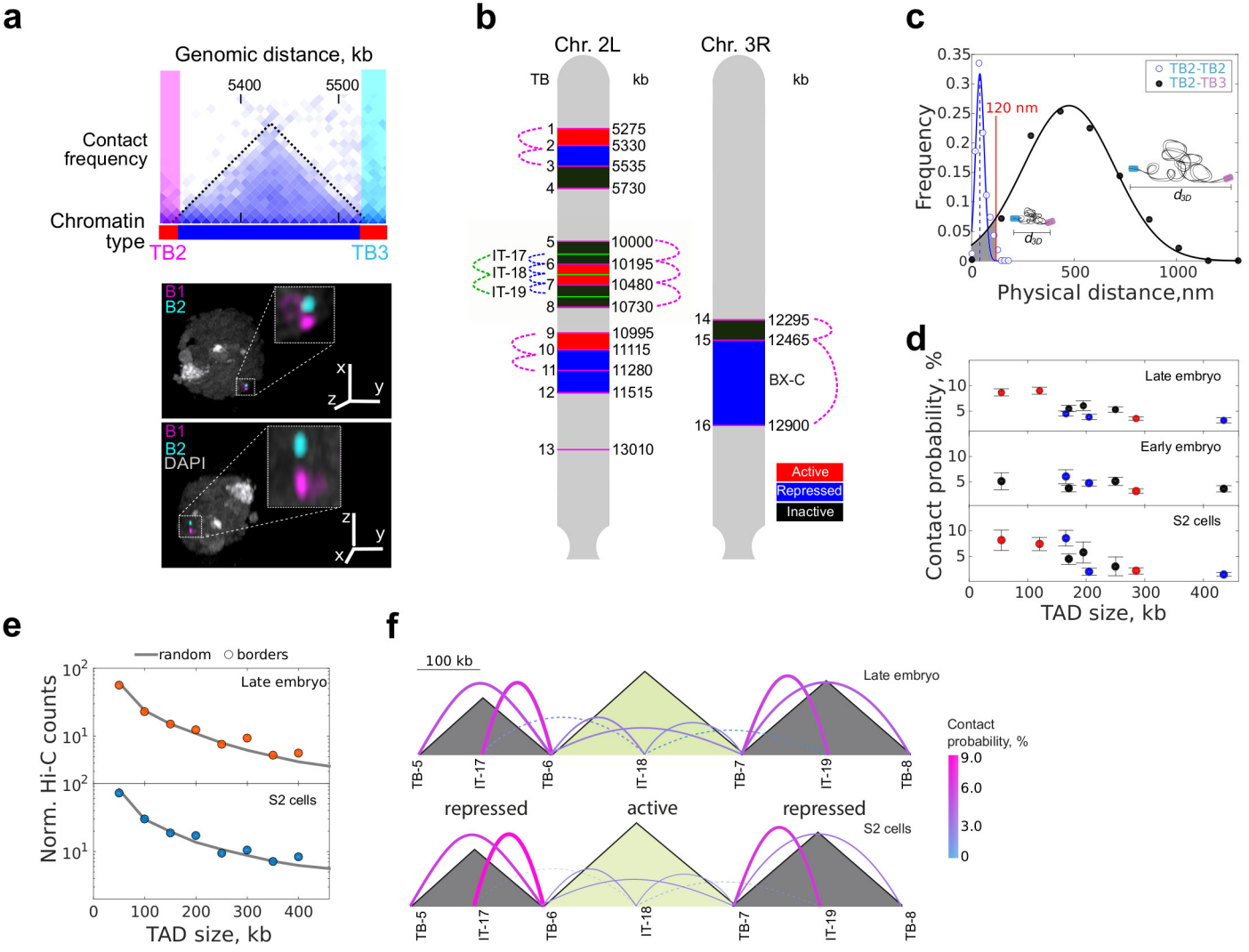
832 **(d-e)** dSTORM rendered images of Alexa-647-labelled **(d)** H3K27me3 and **(e)**
833 H3K4me3. Images show density maps computed from the area of the
834 polygons obtained from the Voronoï diagram with scale defined on top.
835 Scalebar = 1 µm. Zoomed regions display detected compartments (highlighted
836 with different colors). Scalebar = 200 nm Additional, images for all cell types
837 and chromatin marks are displayed in Supplementary Fig. 5a-b.

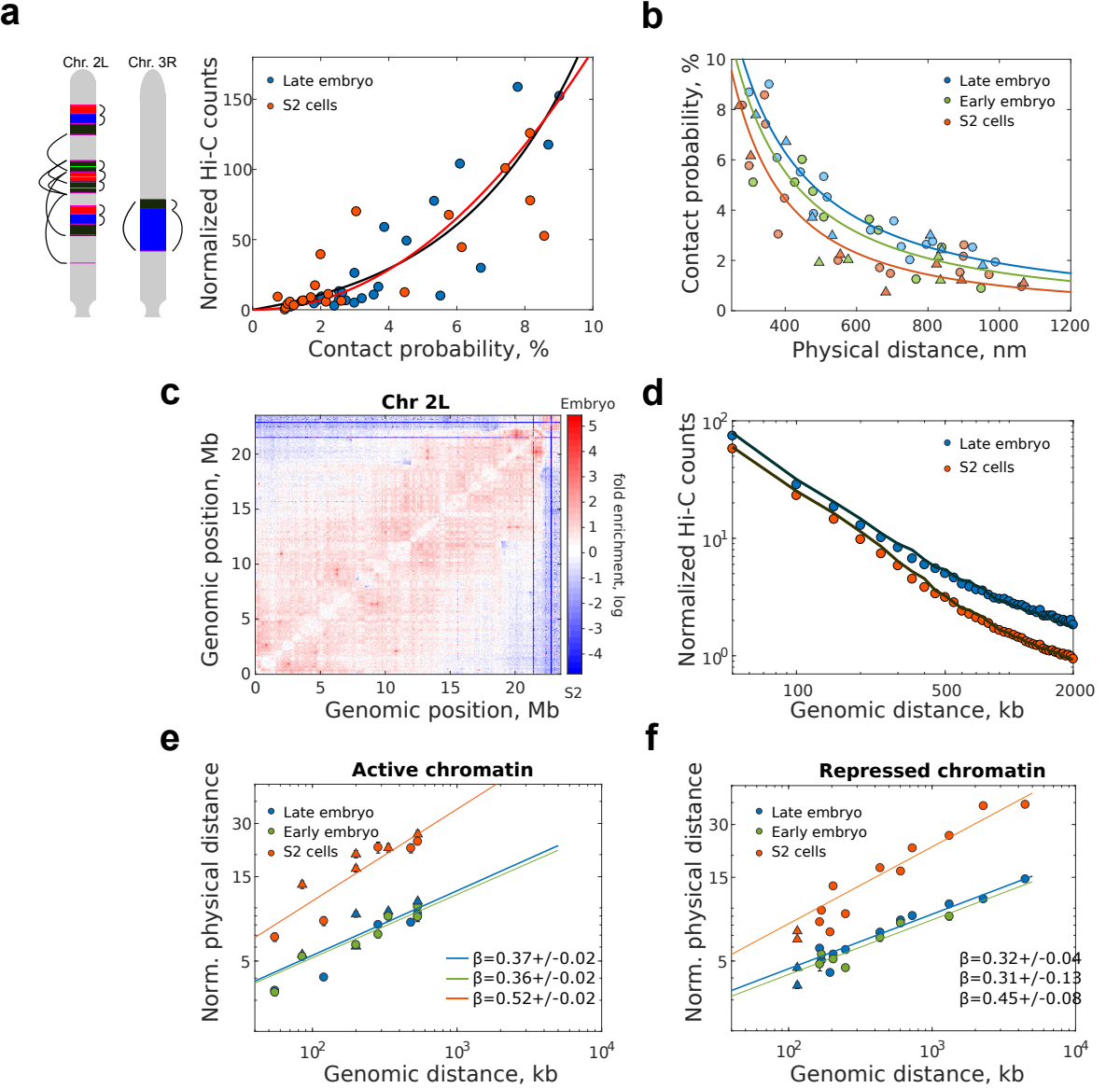
838 **(f-g)** Population based distribution of epigenetic domains sizes as obtained from
839 dSTORM and predicted from ChiP-seq data for H3K27me3 **(f)** and H3K4me3
840 **(g)**. PDF is probability density function. Single cell distributions of physical
841 sizes and Chip-Seq data are shown in Supplementary Figs. 5c-d and 6b,
842 respectively. N=60, from two to three biological replicates in microscopy
843 imaging.

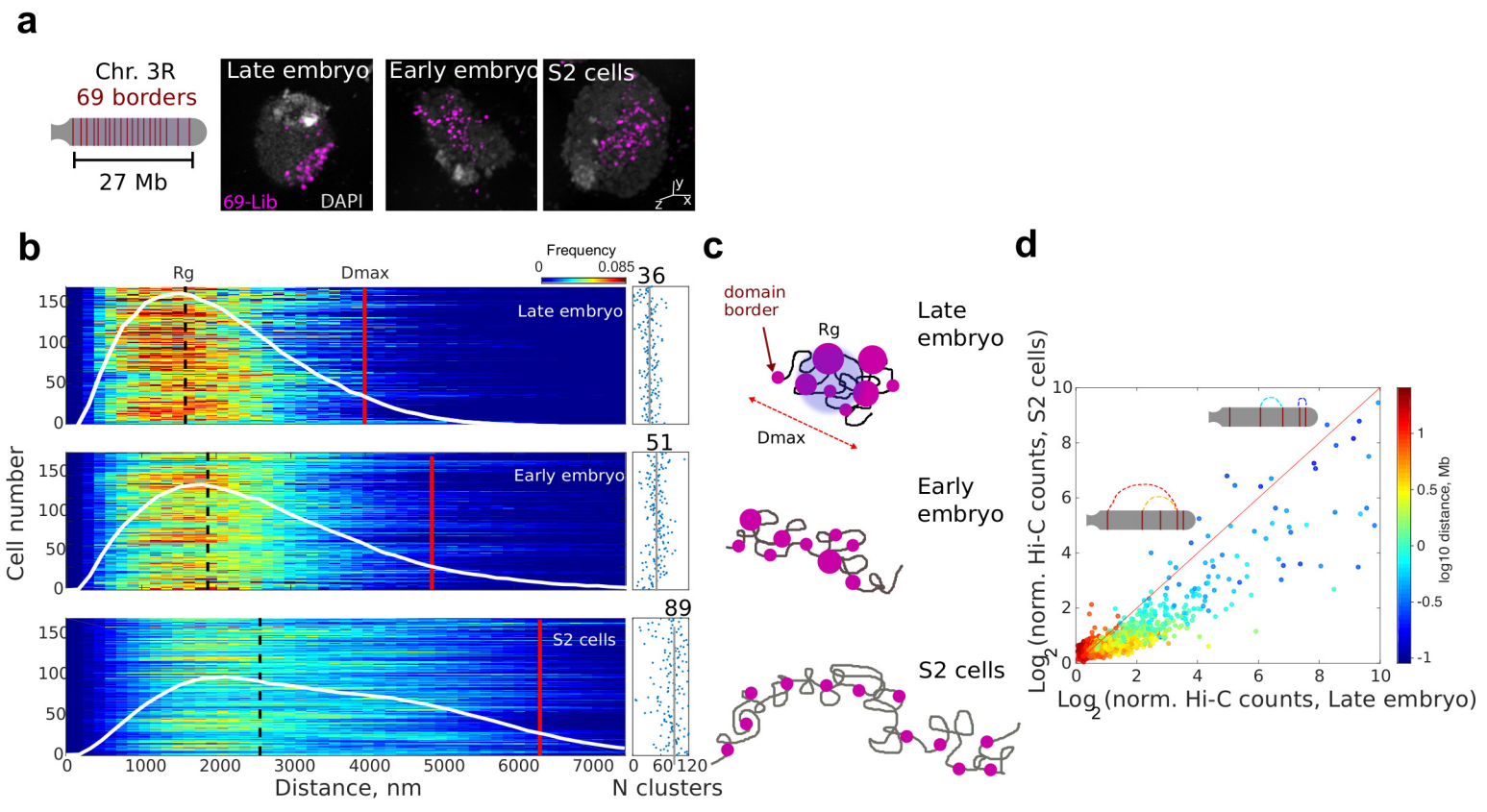
844 **(h)** Percentage of clustering for active and inactive chromatin marks for each cell
845 type. Error bars = SD. One-sample t-test p-values: *p<0.01; **p<0.001.

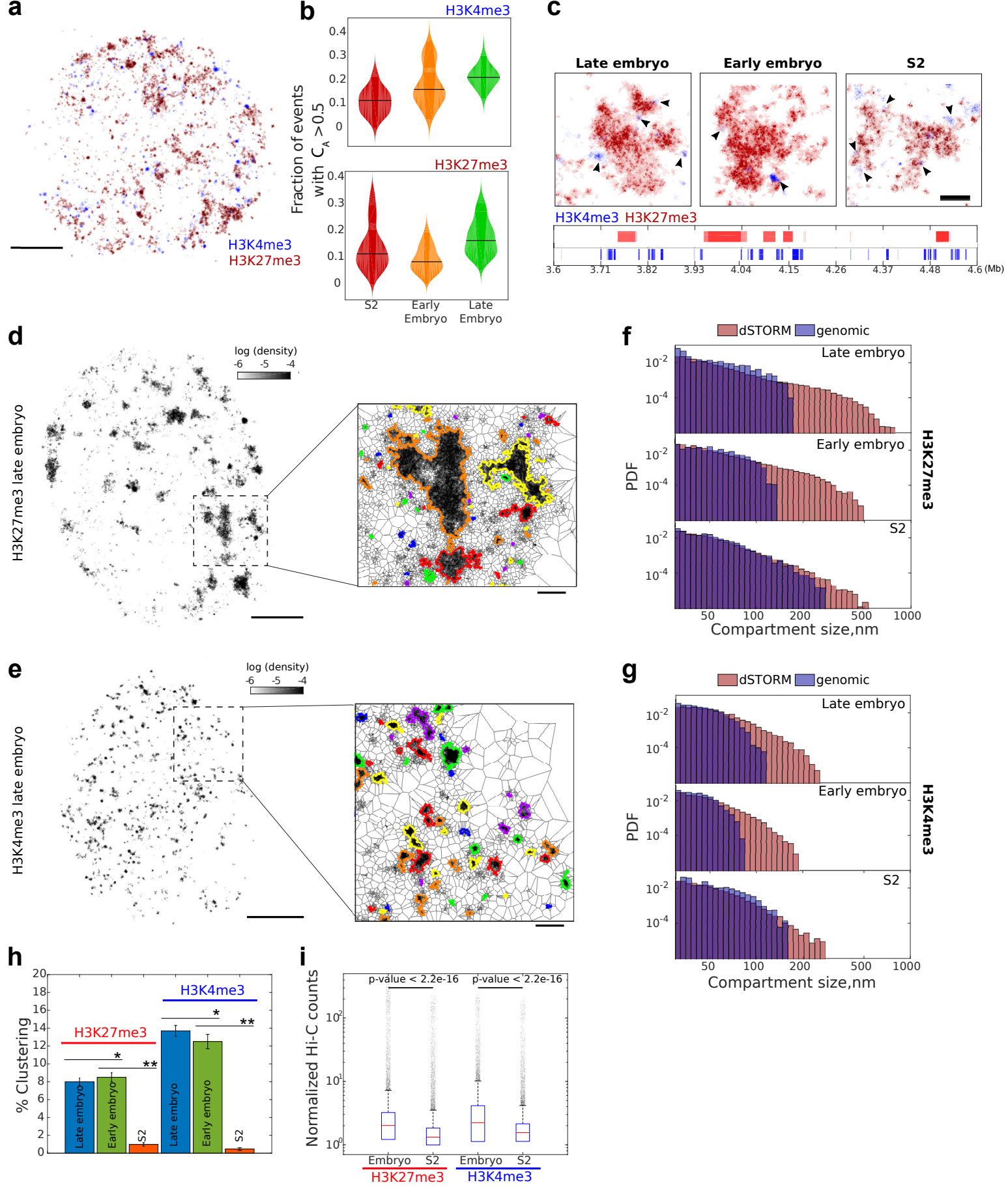
846 **(i)** Box plots of the distributions of normalised Hi-C counts between chromatin
847 domains of H3K27me3 or H3K4me3 in embryos and S2 cells. Results were
848 independent of matrix resolution (10, 20 and 50 kb). Boxes contain 50% of the

data (0.67σ), and red lines mark the median values. Outliers (>3.3σ away from the mean values) are shown as black dots. p-values were calculated using the Welch t-test.


a
Genomic distance, kb
5400
5500
Contact frequency
Chromatin type
TB2
TB3
B1
B2
DAPI
b
Chr. 2L
Chr. 3R
TB
kb
IT-17
IT-18
IT-19
BX-C
Active
Repressed
Inactive
c
TB2-TB2
TB2-TB3
120 nm
Frequency
Physical distance,nm
d
Late embryo
Early embryo
S2 cells
Contact probability, %
TAD size, kb
e
random
borders
Late embryo
S2 cells
Norm. Hi-C counts
TAD size, kb
f
100 kb
Late embryo
S2 cells
repressed
active
repressed
Contact probability, %
TB-5
IT-17
TB-6
IT-18
TB-7
IT-19
TB-8


a
Chr. 2L
Chr. 3R
Late embryo
S2 cells
Normalized Hi-C counts
Contact probability, %
b
Late embryo
Early embryo
S2 cells
Contact probability, %
Physical distance, nm
c
Chr 2L
Embryo
S2
fold enrichment, log
Genomic position, Mb
Genomic position, Mb
d
Late embryo
S2 cells
Normalized Hi-C counts
Genomic distance, kb
e
Active chromatin
Late embryo
Early embryo
S2 cells
β=0.37+/-0.02
β=0.36+/-0.02
β=0.52+/-0.02
Norm. physical distance
Genomic distance, kb
f
Repressed chromatin
Late embryo
Early embryo
S2 cells
β=0.32+/-0.04
β=0.31+/-0.13
β=0.45+/-0.08
Norm. physical distance
Genomic distance, kb


a
Chr. 3R
69 borders
27 Mb
Late embryo
Early embryo
S2 cells
69-Lib
DAPI
b
Frequency
0
0.085
Rg
Dmax
36
Late embryo
51
Early embryo
89
S2 cells
Cell number
Distance, nm
N clusters
c
domain border
Rg
Dmax
Late embryo
Early embryo
S2 cells
d
Log2 (norm. Hi-C counts, S2 cells)
Log2 (norm. Hi-C counts, Late embryo)
log10 distance, Mb


a
H3K4me3
H3K27me3
b
H3K4me3
H3K27me3
Fraction of events with C_A >0.5
S2
Early Embryo
Late Embryo
c
Late embryo
Early embryo
S2
H3K4me3 H3K27me3
3.6 3.71 3.82 3.93 4.04 4.15 4.26 4.37 4.48 4.6 (Mb)
d
log (density)
-6 -5 -4
H3K27me3 late embryo
e
log (density)
-6 -5 -4
H3K4me3 late embryo
f
dSTORM genomic
Late embryo
Early embryo
S2
PDF
H3K27me3
Compartment size,nm
g
dSTORM genomic
Late embryo
Early embryo
S2
PDF
H3K4me3
Compartment size,nm
h
% Clustering
H3K27me3
H3K4me3
Late embryo
Early embryo
S2
i
Normalized Hi-C counts
p-value < 2.2e-16
p-value < 2.2e-16
Embryo S2 Embryo S2
H3K27me3
H3K4me3