Three-dimensional chromatin architecture of early-stage mouse embryos reconstructed via recurrence plots

Yuki Kitanishi[1], Hiroki Sugishita[1,4], Yukiko Gotoh[1,2], and Yoshito Hirata[3,4]

[1]Graduate School of Pharmaceutical Sciences, [2]International Research Center for Neurointelligence (WPI-IRCN), The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0032, Japan

[3]Faculty of Engineering, Information and Systems, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan

[4]Corresponding authors:

Hiroki Sugishita, Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan., Email: hiroki.sugishita@mol.f.u-tokyo.ac.jp

Yoshito Hirata, Faculty of Engineering, Information and Systems, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan, Email: hirata@cs.tsukuba.ac.jp

Abstract

The chromatin conformation capture-related methods such as Hi-C have improved our understanding of nuclear architecture and organization in recent years. However, reconstruction of nuclear architecture of individual cells from single cell Hi-C (scHi-C) data has been challenging due to limited information of DNA contacts owing to the low efficiency of DNA recovery from a single cell. We have previously developed an algorithm named as "recurrence plot- based reconstruction (RPR) method" for reconstructing three dimensional (3D) genomic structure from Hi-C data of single haploid cells[1] and diploid cells [2]. This mathematical method is based on a recurrence plot, a tool of nonlinear time series analysis for visualizing temporal patterns within a time series [3,4], and enables the reconstruction of a unique 3D chromosome architecture even from sparse (low-coverage) DNA contact information. Here we applied the RPR method to analyzing published scHi-C data of diploid cells derived from early-stage F_1 hybrid embryos [5]. We found that paternal and maternal chromosomes become gradually intermingled from 1 cell to 64 cell stage and that discrete chromosome territories (CTs) are largely established between 8 cell and 64 cell stages. We also observed Rabl-like polarized distribution of chromosomes from 2 cell to 8 cell stage but this polarization becomes mostly dissolved by 64 cell stage. The formation of Rabl-like configuration precedes rod-like extension of the chromosomal shape and their parallel alignment, implicating a role of Rabl-like configuration in avoiding entanglement and promoting effective mixing of chromosomes before establishment of CTs. We also found a cell-to-cell variability in chromatin configuration. Combination of scHi-C and RPR analyses thus can characterize distinct 3D chromatin architecture of individual cells at different developmental stages during early embryogenesis.

Introduction

Spatial organization of the mammalian genome plays a key role in transcription, RNA processing and DNA replication [6–9]. The chromatin is arranged as a hierarchy ranging from loops to different scales of domains and compartments up to chromosome territories (CT), which is a foundation for gene regulation and cell function [10–12]. Classically, spatial localization of genomic region-of-interest in the nucleus has been elucidated with DNA-fluorescence in situ hybridization (FISH) [13]. However, although recent endeavors have increased the number of genomic loci to be studied [14,15], FISH-based method can examine only targeted genomic locus-of-interest. On the other hand, Hi-C, a derivative of chromosome conformation capture [16], can elucidate nontargeted and genome-wide pair-wise DNA contacts by sequencing after proximity-dependent ligation [17]. Application of Hi-C to cell populations (bulk Hi-C) has revealed principles of chromatin organization such as topology-associated domains and A/B compartments [17–20]. Bulk Hi-C per se, however, does not reveal 3D positioning of genomic loci in relation to global nuclear geography (3D genome atlas). The combination of bulk Hi-C data and other imaging and epigenetic analyses such as DNA-FISH and Lamin-DamID has contributed to our understanding of how chromatin architectures revealed by Hi-C may spatially be located and coordinated in the nucleus [21–28], although the data are based on the average of cell ensembles.

The 3D architecture of chromatin changes dramatically during cell differentiation or transitions of the cell state. Moreover, the chromatin architecture is highly dynamic, so it is highly heterogenous even between cells of the same type and state. It is thus crucial to investigate the 3D genome atlas at a single cell level. Single-cell Hi-C (scHi-C) and in situ genome sequencing (IGS) strategies have tackled this problem in a nontargeted and genome-wide manner. IGS is a method which spatially localizes genome-wide paired end sequencing in their endogenous location within intact cells. scHi-C is a modified version of widely-used Hi-C methods, which has been applied to both haploid and diploid cells [27,29–32]. In diploid scHi-C, allele-specific single nucleotide variation (SNV) sites are often used to identify the two (paternal and maternal) alleles of the chromosomes. However, reconstruction of 3D nuclear architecture from scHi-C data has been challenging due to a limited number of DNA contacts recovered from a single cell as well as the limited frequency of SNVs within the genome. Tan et al. has developed a modified scHi-C method for diploid cells called Dip-C that combines a transposon-based whole genome amplification procedure and an algorithm for imputation of the two chromosome haplotypes linked by each contact [32], although this algorithm contains some imputations that are not necessarily proven mathematically. We have reported another algorithm named as “recurrence plot- based reconstruction (RPR) method" for reconstructing the 3D chromosome structure from Hi-C data of single haploid cells [1] and diploid cells [2]. This mathematical method is based on a recurrence plot, a tool of nonlinear time series analysis for visualizing temporal patterns within a time series[1] and enables the reconstruction of a unique 3D chromosome architecture even from sparse (low-coverage) DNA contact information. This method has been mathematically proven [33,34]. In contrast to Dip-C’s

algorithm which assumes that two alleles are not spatially close to each other and do not have similar shapes, RPR algorithm only assumes that consecutive genomic regions are neighboring [1,2].

As a proof-of-concept, we analyzed 3D nuclear architecture of embryos at early preimplantation stages after fertilization with the use of RPR method and a set of scHi-C data publicly available [5], given that chromatins at this stage undergo dramatic changes as development proceeds [35]. Our results confirmed some of the previous observations obtained with FISH and IGS strategies such as gradual intermixing of paternal and maternal alleles from 1cell (zygote) to 64 cell stage. Moreover, we revealed that discrete (localized) chromosome territories (CTs) start to establish between 8 and 64 cell stages, which roughly coincides with the dissolvement of Rabl-like polarized and rod-like structure of chromosomes.

Results

Application of the RPR algorithm to scHi-C data of early stage mouse embryos

We applied the RPR method to scHi-C data on diploid cells of early stage mouse embryos (n=75 for 1 cell, n=109 for 2 cell, n=121 for 4 cell, n=125 for 8 cell, and n=201 for 64 cell stage) published by Collombet et al. [5]. The data were obtained from highly polymorphic F_1 hybrid embryos after crossing female *Mus musculus domesticus* (C57Bl/6J) with male *Mus musculus castaneus* (CAST/EiJ) (Collombet et al., GSE129029). The paternal and maternal alleles were distinguished by SNVs between these genetic backgrounds. 22.6% (1 cell), 24.4% (2 cell) 21.2% (4 cell), 23.7% (8 cell) and 22.9% (64 cell) of genomic reads were uniquely assigned to the maternal (C57Bl/6J) genome, whereas 14.6% (1 cell), 18.0% (2 cell) 16.7% (4 cell), 17.9% (8 cell) and 17.7% (64 cell) of genomic reads were uniquely assigned to the paternal (CAST/EiJ) genome (Fig. 1c). For performing RPR, we divided the genome into 1 Mb segments and used contact information of pairs of the genomic segments both of which harbor at least one SNV. A brief summary of the RPR method is as follows (Fig. 1a). First, we constructed a weighted network from a contact map obtained from a single diploid cell Hi-C dataset. In this network, each node corresponds to a segment of an allele-distinguished chromosome. We connected two nodes if there is a contact between these corresponding chromosomal segments both of which have their alleles identified with SNVs. We also connected two nodes if they are on the same allele of the same chromosome within 10 Mb. Let i and j denote a node in the network, respectively. Then, we assigned the weight for the edge between nodes i and j by using the portion of common nodes against the nodes which are connected with either i, j or both (this definition coincides with that of the Jaccard coefficient [36]. More exactly, the weight is the portion of common nodes subtracted from 1. Then, this weight can be regarded as the local distance between two corresponding segments i and j. The latter parts are similar to Lesne et al., Hirata et al., and Tenenbaum et al. [37–39]: we calculated the shortest distances between every pair of nodes in the network, followed by applying the classical multidimensional scaling to find a set of point arrangements preserving the shortest distances so that top three components become our

3D reconstruction of chromosomes. The strength of our method is that (i) we can reconstruct local structures more finely by reconstructing the local distances by the Jaccard coefficient appropriately compared with that of Lesne et al. [39], (ii) we can apply the method with sparser data because the proportion of common nodes is a robust statistic under such conditions, and (iii) we can use the information of pairs of chromosome segments in the Hi-C data, in which at least one of the segments does not harbor a SNV and thus were not used for 3D reconstruction, to evaluate the accuracy of reconstruction.

We defined the inaccuracy score for each reconstructed point by using the root mean square distance between each reconstructed point and its corresponding neighbor (the contact pair found in the Hi-C data) (Fig. 1b). Here, the neighbor was defined as a 1 Mb segment containing the contact pair. The more accurate the reconstruction is, the closer the neighbor is expected to be to the reconstructed point in the 3D genome atlas. With the RPR method, the center of the nucleus tends to be more accurate compared to the peripheral (Ext Fig. 1a). The overall results indicate that a majority of reads, even those with no overlapping SNV, to be assigned to the predicted spatial location through co-localization with haplotype-resolved reads in the same location. In this analysis, however, some cells showed relatively high averaged inaccuracy scores (with greater standard deviations (s.d.) as shown in Fig 1d and Ext. Fig. 1b), which also tended to show an irregular shape of the nucleus (for example, see Fig. 1e, left). We thus excluded the cells with s.d. of the inaccuracy score greater than 0.15 (Ext. Fig. 1c) from the following analyses. We found no significant correlation between the inaccuracy score and the number of contact pairs detected by Hi-C (r = -0.101, Ext. Fig. 1d), suggesting that the high inaccuracy score is not solely due to few reads of Hi-C contact pairs.

Spatial intermixing of paternal and maternal genomes during early stage development

At the 1 cell stage, the parental genomes are separated in two haploid pronuclei. The spatial mixing of the parental genomes after this stage has previously been investigated, but the time and extent of their mixing has been somewhat controversial [40,41]. We thus investigated the extent of mixing of the parental genomes based on the RPR of 3D genomic architecture. We could observe the gradual spatial intermixing of paternal alleles from 1 to 64 cell stages (Fig. 2a). To quantitively measure the extent of intermixing between the parental genomes, we first calculated the distance of all pairs of 1 Mb genomic segments. We then defined proximity pairs (the genomic segments located in a proximity) as the top (nearest) 5% of all pairs (Extended Figure 2). In order to focus on inter-chromosomal arrangements, we only used the information about the genomic segments which showed at least one proximity pair with a genomic segment outside of its own chromosome. We then calculated the fraction of proximity pairs in which the genomic segments are located in different alleles (maternal and paternal alleles) in each cell. This fraction ("allele intermixing score") was close to zero at the 1 cell stage, as expected, but gradually increased at the 2, 4, 8 and 64 stages (Fig. 2c). The mean intermixing score significantly increased between 1 and 2 cell stages, 2 and 4 cell stages, 4 and 8 cell stages and 8 and 64 cell stages (Mann Whitney U test with Bonferroni correction, $p < 10^{-27}$, $p < 10^{-24}$, $p < 10^{-3}$, $p < 10^{-4}$) (Fig. 2b), indicating that the parental genome

intermingling starts at the 2 cell stage and completes at least after 8 cell stage. Given that the allele intermixing score became close to 0.5 at the 64 cell stage (Fig. 2b), chromosomes appear to be positioned irrespectively to their alleles at this stage. Interestingly, however, the extent of allelic intermingling appears to vary between cells at all stages (see examples of 64 cell stage, Fig. 2c), suggesting a cell-to-cell variety of chromosomal organization.

Formation and dissolvement of Rabl-like configuration during early stage development

Mouse chromosomes are acrocentric. In early stage mouse embryos, chromosomes are organized in a Rabl-like configuration in which centromeres and their nearby telomeres are confined to one part of the nuclei while the remaining telomeres are clustered in the other part [41–44]. However, the extent of this polarized configuration at each developmental stage has been controversial. We thus examined the polarity of chromosomal arrangement along the centromere-distal telomere axis. When we labeled the short arm and long arm termini of each chromosomes in the reconstructed 3D map, the polarization of chromosomes was visible from 2 to 8 cell stages (Fig. 3a). To quantitatively analyze the extent of polarization, we calculated the mean position along the chromosomal axis (the short arm terminus = 0, the long arm terminus = 1) of proximity pairs for each genomic segment (proximity pairs within the same chromosome of the genomic segment were excluded as described in Fig. 2b). We then compared the position of each genomic segment along the chromosomal axis and the mean chromosomal position of its proximity pairs, and found their greater correlation at 2, 4 and 8 cell stages compared to 1 and 64 cell stages (Fig. 3b). This difference in correlation indicates that chromosomes become polarized to a Rabl-like configuration from 1 to 2 cell stage as reported by Payne et al [41] and continue to be polarized up to 8 cell stage, but this configuration becomes mostly dissolved from 8 to 64 cell stage. The dissolvement of polarization appears to start between 4 and 8 cell stages, given the angle of the correlation curve becomes shallower (Fig. 3b).

Establishment of discrete chromosome territories during early stage development

Within the nucleus, each chromosome is confined to a discrete region, referred to as a chromosome territory (CT) [10–12]. The spatial organization of CTs is not random and critical for gene regulation and genome stability [12]. It remains unclear, however, when discrete (patch-like) CTs emerge during early stage development. We thus evaluated the extent of intra-chromosomal interactions (compared to that of inter-chromosomal interactions) at each stage based on the RPR of 3D genomic architecture. We again defined proximity pairs as the top (nearest) 5% of all pairs and calculated the fraction of proximity pairs in which the genomic segments are located in the same chromosome in each cell. This fraction ("the CT score") slightly increased from 1 cell to 8 cell stage, and greatly increased from 8 to 64 cell stage (Fig. 4a and 4b). This great increase suggests that discrete CTs are largely established between 8 and 64 cell stages. The tendency of a steep increase of the CT score between 8 and 64 cell stages was found in all chromosomes but to a lesser extent for shorter chromosomes especially chromosome 19 (Fig. 4c), probably due to their already packed (nonelongated) configuration at 8 cell stage.

We also determined the distance between the termini of short and long arms of each chromosome as an indicative of overall shape of the chromosome. Interestingly, the average distance greatly increased from 2 to 4 cell stage and reduced from 4 to 64 cell stages (Fig. 5a, b). This is consistent with the extended (rod-like) shape of chromosomes found particularly at 4 cell stage (Fig. 3a). This result also supports the establishment of discrete (patch-like, nonelongated) CTs from 8 to 64 cell stage.

Discussion

Bulk Hi-C can give us contact information between genomic loci based on proximity ligation, but this method is limited to measuring simultaneous contacts between a small number (~2-3) of genomic regions and thus cannot show how multiple genomic regions simultaneously organize within the nucleus [45]. Other methods called genome architecture mapping (GAM) [46], which combines ultrathin cryo-sectioning and DNA sequencing followed by the mathematical model SLICE, as well as split-pool recognition of interactions by tag extension (SPRITE) [47], which identifies proximal genomic regions by barcoding them during split-pool cycles, can reveal multiple-way contacts across the genome, although the contacts are estimated as an average of cell ensembles. Accordingly, these methods cannot generate a global picture of genome organization at a single cell level. The RPR of scHi-C data as well as the recently-reported method IGS, on the other hand, can estimate the 3D coordinates of all genomic regions at a single cell level. This estimation performance means that, with the use of RPR, we can theoretically identify simultaneous neighboring of multiple genomic regions as well as the distance between any genomic regions of interest even without direct contact. In this study, we applied the RPR method to reveal 3D genome architecture in early-stage embryos at a single cell level. We also newly implemented the inaccuracy score to evaluate the accuracy of 3D positioning of each genomic segment to the reconstructed 3D atlas. We used the averaged inaccuracy score of all genomic segments to exclude cells which do not show reliable 3D reconstruction from the following analyses in this study. Possible reasons which may explain high inaccuracy scores include too few Hi-C contact reads, particular cell cycle phase or genomic recombination. We did not observe any obvious relation to the mitotic phase, however. A very weak negative correlation between the inaccuracy score and the number of Hi-C contact reads was found, but this may not be the main reason. Future studies need to clarify this point to improve the accuracy of the Hi-C based RPR method.

After fertilization and breakdown of the pronuclear envelope, the initially separate parental genomes gradually intermingle and undergo major architectural reorganization to prime the embryos for zygotic genome activation, which mostly begins at the 2 cell stage in mice [35] and for cell fate determination into distinct lineages. The process of parental genome intermingling has been visualized by DNA-FISH and IGS. A DNA-FISH analysis which separately visualized parental centromeres in mouse interspecific hybrids (as well as an

analysis of sperm-specific BrdU labeling) has shown that topological separation of the parental genomes is preserved up to the 4 cell stage and then gradually disappears [40]. By contrast, a recent IGS analysis has quantitatively shown the intermingling of paternal and maternal genomes from 1 cell to 4 cell stages [41]. In this study, by the use of the RPR method, we confirmed the observation of Payne et al.[41] regarding the paternal and maternal genome intermingling from 1 to 4 cell stages. Moreover, our results quantitatively support the conclusion of Mayer et al.[40] on significant intermingling of parental genomes between 4 and 8 cell stages. It is interesting that a fraction of cells at 64 cell stage still show separation of the parental genomes, which is consistent with the previous report indicating that 10-15% of cells at the 64 cell stage show a nonrandom distribution of the two parental centromere sets [40].

Previous studies have reported a polarized, Rabl-like configuration of chromosomes in early stage embryos. A 3D FISH analysis of centromeric, pericentromeric and telomeric DNA sequences showed rather constant polarity from 2 to 32 cell stages [44]. However, an IGS analysis showed an increase of polarity from 1 to 4 cell stages [41]. Our study on the RPR of scHi-C data indicated that the Rabl-like chromosomal polarity increases from 1 to 4 cell stages, which is consistent with the study by Payne et al. [41]. Interestingly, another DNA FISH analysis showed reduction of polarity from 4 to 8 cell stage [40]. Our study indicated that the Rabl-like polarity may start to decrease from 4 to 8 cell stage, but rather dramatically reduce between 8 and 64 cell stages. This sharp reduction appears to contribute to the establishment of discrete CTs from 8 to 64 cell stages.

The chromosomes are organized into localized CTs that are an essential layer of the higher-order nuclear architecture. We estimated the extent of CT formation (the CT score) in this study as the ratio of intra-chromosomal interactions among all chromosomal interactions between 1 Mb genomic segments estimated by the RPR method. This estimation suggested that discrete CTs are mostly established between 8 and 64 cell stages during early embryonic development. This timing appears to coincide with the dissolvement of Rabl-like polarized configuration as mentioned above. However, this dissolvement may not be the only cause of increased CT scores, given that the CT score increased even from 1 to 4 cell stage when the chromosomes became more polarized.

The biological meaning of Rabl-like polarized configuration has remained enigmatic. It has been proposed to be just a result from the anaphase orientation of chromosomes without particular cellular roles. Of note, although Rabl-like polarized configuration (which involves separate clustering of telomere and centromere termini) starts to form at 2 cell stage (Fig. 3), the shape of chromosomes starts to elongate only at 4 cell stage (Fig. 5). And both Rabl-like polarization and rod-like extension of chromosomes disappear around the time of CT formation at 64 cell stage. This sequence of events implies a possible role of Rabl-like polarization in rod-like extension and parallel alignment of chromosomes, which may in turn avoid entanglement and thus promote effective mixing of chromosomes, as proposed in yeast [48,49]. The resultant mixing of chromosomes, especially those from different parental

origins, may be necessary for establishing proper CTs and therefore should be completed before discrete CTs start to give rise.

In future studies, it would be of great interest to overlay other chromatin-related information onto the 3D genome architecture revealed by the RPR analysis at a single cell level. The RPR analysis in this study was performed at a 1 Mb resolution. Analyses at a higher resolution would be necessary for identification of contacts between gene promoters and enhancers and their relations to epigenetic information such as DNA and histone modifications as well as to TADs and compartments.

Material and methods

scHi-C data processing

Processed single cell Hi-C (scHi-C) data of early stage mouse embryos were obtained from Collombet et al. [5] (GSE129029). Thus, sequence data processing (i.e. quality control, Hi-C data mapping and Allele assignment of sequencing reads) is according to the provided data. The detail information about each result is on supplemental files.

Reconstruction of 3D chromosome structure from a single diploid cell Hi-C dataset

Suppose that for each allele of each chromosome, the genomic sequence is divided into the segments of length 1Mb. For each segment, we will prepare a representative point in the three-dimensional space. We array the representative points for maternal alleles from Chr 1 to Chr X, then the representative points for paternal alleles from Chr 1 to Chr X for female cells, from Chr 1 to Chr Y for male cells to form a one-dimensional axis.

Assume that a set of phased contacts is given. First, we map this set of phased contacts to the corresponding contact map C with the resolution of 1Mb, whose element $C(i,j)$ is 1 if segments i and j have at least a phased contact in the set. Otherwise, we set $C(i,j) = 0$. In addition, we define $C(i,i) = 1$ for every segment i for the notational convenience. Therefore, the matrix C becomes symmetric, i.e., $C(i,j) = C(j,i)$.

Second, we prepare a recurrence plot denoted by a matrix R based on the contact map C. The matrix R is in the same size as C. We define $R(i,j) = 1$ if there exists a pair of (k,l) which corresponds to the same pair of the same alleles for the same chromosomes as (i,j),

as well as satisfy $|k-i| \leq 10$, $|j-l| \leq 10$, and $C(k,l)=1$. Therefore, the matrix R is also symmetric, i.e., $R(i,j)=R(j,i)$. This kind of definition for R means that the consecutive segments for the corresponding chromosomes are also spatial neighbors, facilitating below the reconstruction of 3D chromosome structure [1,50].

Third, we construct a network based on R. In this network, each segment corresponds to a node in the network. We assign an edge between i and j if $R(i,j)=1$. For each edge between i and j, we assign the following local distance d:

$$d(i,j)=\frac{\sum_k R(i,k)+\sum_k R(j,k)-2\sum_k R(i,k)R(j,k)}{\sum_k R(i,k)+\sum_k R(j,k)-\sum_k R(i,k)R(j,k)}.$$

This local distance is the "one-complement" of the Jaccard coefficient [36], which is literally the ratio of unshared neighbors between i and j against neighbors for either of i or j. From this third step, we follow Hirata et al. [1,38].

Fourth, we find the shortest distance for every pair of nodes in this network to obtain a set of global distances. For this step, we here use Johnson's method [51].

Fifth, we apply the multidimensional scaling to the set of global distances. The multidimensional scaling is a method for finding a set of point arrangements by preserving as a set of distances among given points. We can regard the top three components of the resulting series as our reconstruction of 3D chromosome structure. Let $\{x(i) \in \mathcal{R}^3\}$ be our reconstruction of 3D chromosome structure.

Inaccuracy score

We define an inaccuracy score for each reconstruction point $x(c)$ as follows: First, find the set N_c of the neighboring points in the recurrence plot, i.e. $N_c=\{i: R(i,c)=1\}$. Then, the inaccuracy score for $x(c)$ can be defined as

$$\sqrt{\frac{1}{|N_c|}\sum_{i\in N_c}||x(c)-x\big(n_c(i)\big)||^2}.$$

Namely, in this inaccuracy score, we evaluate how spread the neighboring points in the recurrence plot are in our reconstruction.

Visualization of RPR

We visualized the 3D chromosome structure regarding the top three components obtained from RPR analysis as 3D coordinates. Each segment was visualized in 1 Mb resolution and the overall structure was displayed as beads-on-string structure by Pymol (version 2.4.0) software.

Exclusion of cell using the inaccuracy score

Standard deviation (s.d.) of the inaccuracy score was calculated to exclude cells with inaccurate structure as depicted in Fig. 1e. We set the threshold of the inaccuracy score s.d. as 0.15 and the cells over this value were removed from further analyses.

Calculation of maternal-paternal intermixing score

Euclid distance between all 1 Mb segments versus were calculated. The top 5% nearest distance was calculated and used as a threshold to decide whether two segments were proximally located. “Allele intermixing score” for each 1 Mb segment was calculated as the proportion of neighboring segments with different allele out of total neighboring segments. Neighboring segments in the same chromosome were excluded from this calculation to focus on inter-chromosomal arrangements. The cell average score was determined by averaging the overall maternal-paternal intermixing score of all 1 Mb segments.

Calculation of mean chromosomal position of proximal pairs

In this analysis, we followed the method of Payne et al. [41] with some modifications. “Chromosomal position” for each 1Mb segment was calculated as the relative position in the chromosomal axis, which took continuous value from 0 to 1 (the short arm terminus = 0, the long arm terminus = 1). Mean chromosomal position of the proximal pairs (“Proximal position score”) was calculated excluding pairs on the same chromosome as the same as the calculation of allele mixing score. Finally, all chromosomes were divided into 100 segments and average proximal position score for each segment was compared with their chromosomal position.

Calculation of chromosome territory score

“Chromosome territory score (CT score)” was calculated as the proportion of neighboring segments on the same chromosome out of all neighboring segments. The cell average score

was determined by averaging the overall CT score of all 1 Mb segments.

Data Availability

Datasets analyzed during the current study are available from the corresponding author on reasonable request.

Declaration of interests

Y.H. has a pending Japanese patent application related to this manuscript (its patent application number is 2016-023214). The authors declare no other competing interests.

Author contribution

Conception: Y.K., H.S, Y.G, Y.H., Investigation and data analysis: Y.K. and Y.H, Writing: Y.K., H.S., Y.G and Y.H., Funding acquisition: Y.G. and Y.H., Supervision: H.S., Y.G. and Y.H.

Acknowledgments

This study was supported by KAKENHI grants from the Ministry of Education, Culture, Sports, Science, and Technology of Japan and the Japan Society for the Promotion of Science (JP16H06481, JP16H06279, JP16H06479 and JP15H05773 to Y.G.) as well as by AMED-CREST of the Japan Agency for Medical Research and Development (JP21gm1310004) by the Uehara Memorial Foundation, and by the International Research Center for Neurointelligence (WPI-IRCN), The University of Tokyo Institutes for Advanced Study.

References

1. Hirata, Y., Oda, A., Ohta, K. & Aihara, K. Three-dimensional reconstruction of single-cell chromosome structure using recurrence plots. *Sci. Rep.* **6**, 34982 (2016).
2. Hirata, Y., Oda, A. H., Motono, C., Shiro, M. & Ohta, K. Reconstructions of three-dimensional chromosome structure of single diploid cells with self-checking of validity: techniques inspired by nonlinear time series analysis. *Submitted* (2021).
3. Eckmann, J.-P., Kamphorst, O. S. & Ruelle, D. Recurrence Plots of Dynamical Systems. *EPL* **4**, 973–977 (1987).
4. Marwan, N., Carmen Romano, M., Thiel, M. & Kurths, J. Recurrence plots for the analysis of complex systems. *Phys. Rep.* **438**, 237–329 (2007).
5. Collombet, S. *et al.* Parental-to-embryo switch of chromosome organization in early embryogenesis. *Nature* **580**, 142–146 (2020).
6. Zheng, H. & Xie, W. The role of 3D genome organization in development and cell differentiation. *Nat. Rev. Mol. Cell Biol.* **20**, 535–550 (2019).
7. Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 661–678 (2016).
8. Jordan Rowley, M. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19**, 789–800 (2018).
9. Burton, A. & Torres-Padilla, M.-E. Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 723–735 (11/2014).
10. Bolzer, A. *et al.* Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes. *PLoS Biol.* **3**, e157 (2005).
11. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292–301 (4/2001).
12. Marella, N. V., Seifert, B., Nagarajan, P., Sinha, S. & Berezney, R. Chromosomal rearrangements during human epidermal keratinocyte differentiation. *J. Cell. Physiol.* **221**, 139–146 (10/2009).

13. Levsky, J. M. & Singer, R. H. Fluorescence in situ hybridization: past, present and future. *J. Cell Sci.* **116**, 2833–2838 (2003).

14. Nguyen, H. Q. *et al.* 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing. *Nat. Methods* **17**, 822–832 (2020).

15. Su, J.-H., Zheng, P., Kinrot, S. S., Bintu, B. & Zhuang, X. Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* **182**, 1641-1659.e26 (2020).

16. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing Chromosome Conformation. *Science* **295**, 1306–1311 (2002).

17. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).

18. Sexton, T. *et al.* Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell* **148**, 458–472 (2012).

19. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).

20. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

21. Wang, S. *et al.* Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598–602 (2016).

22. Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, (2018).

23. Nir, G. *et al.* Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS Genet.* **14**, e1007872 (2018).

24. Szabo, Q. *et al.* TADs are 3D structural units of higher-order chromosome organization in Drosophila. *Science Advances* **4**, eaar8082 (2018).

25. Finn, E. H. *et al.* Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* **176**, 1502-1515.e10 (2019).

26. Mateo, L. J. *et al.* Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54 (2019).

27. Stevens, T. J. *et al.* 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).

28. Zheng, X. *et al.* Lamins Organize the Global Three-Dimensional Genome from the Nuclear Periphery. *Mol. Cell* **71**, 802-815.e7 (2018).

29. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).

30. Nagano, T. *et al.* Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).

31. Ramani, V. *et al.* Massively multiplex single-cell Hi-C. *Nat. Methods* **14**, 263–266 (2017).

32. Tan, L., Xing, D., Chang, C.-H., Li, H. & Sunney Xie, X. Three-dimensional genome structures of single diploid human cells. *Science* **361**, 924–928 (2018).

33. Khor, A. & Small, M. Examining k-nearest neighbour networks: Superfamily phenomena and inversion. *Chaos* **26**, 043101 (2016).

34. Hirata, Y., Komuro, M., Horai, S. & Aihara, K. Faithfulness of Recurrence Plots: A Mathematical Proof. *Int. J. Bifurcat. Chaos* **25**, 1550168 (2015).

35. Schultz, G. A. Molecular Biology of the Early Mouse Embryo. *Biol. Bull.* **171**, 291–309 (1986).

36. Levandowsky, M. & Winter, D. Distance between Sets. *Nature* **234**, 34–35 (1971).

37. Tenenbaum, J. B., de Silva, V. & Langford, J. C. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* **290**, 2319–2323 (2000).

38. Hirata, Y., Horai, S. & Aihara, K. Reproduction of distance matrices and original time series from recurrence plots and their applications. *Eur. Phys. J. Spec. Top.* **164**, 13–22 (2008).

39. Lesne, A., Riposo, J., Roger, P., Cournac, A. & Mozziconacci, J. 3D genome reconstruction from chromosomal contacts. *Nat. Methods* **11**, 1141–1143 (2014).

40. Mayer, W., Smith, A., Fundele, R. & Haaf, T. Spatial Separation of Parental Genomes in Preimplantation Mouse Embryos. *J. Cell Biol.* **148**, 629–634 (2000).

41. Payne, A. C. *et al.* In situ genome sequencing resolves DNA sequence and structure in intact biological samples. *Science* **371**, (2021).

42. Merico, V. *et al.* Epigenomic differentiation in mouse preimplantation nuclei of biparental, parthenote and cloned embryos. *Chromosome Res.* **15**, 341–360 (2007).

43. Probst, A. V. *et al.* A strand-specific burst in transcription of pericentric satellites is required for chromocenter formation and early mouse development. *Dev. Cell* **19**, 625–638 (2010).

44. Aguirre-Lavin, T. *et al.* 3D-FISH analysis of embryonic nuclei in mouse highlights several abrupt changes of nuclear organization during preimplantation development. *BMC Dev. Biol.* **12**, 30 (2012).

45. O'Sullivan, J. M., Hendy, M. D., Pichugina, T., Wake, G. C. & Langowski, J. The statistical-mechanics of chromosome conformation capture. *Nucleus* **4**, 390–398 (2013).

46. Beagrie, R. A. *et al.* Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).

47. Quinodoz, S. A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744-757.e24 (2018).

48. Pouokam, M. *et al.* The Rabl configuration limits topological entanglement of chromosomes in budding yeast. *Sci. Rep.* **9**, 1–10 (2019).

49. Chikashige, Y. *et al.* Telomere-led premeiotic chromosome movement in fission yeast. *Science* **264**, 270–273 (1994).

50. Tanio, M., Hirata, Y. & Suzuki, H. Reconstruction of driving forces through recurrence plots. *Phys. Lett. A* **373**, 2031–2040 (2009).

51. Johnson, D. B. Efficient Algorithms for Shortest Paths in Sparse Networks. *J. ACM* **24**, 1–13 (1977).

Figure Legend

Figure 1: Implementation of the inaccuracy score to the recurrence plot-based reconstruction (RPR) method

a. Schematic illustration of the recurrence plot-based reconstruction (RPR). **b**. Schematic illustration of the inaccuracy score. Suppose that we have a contact map on the left side and its 3D reconstruction on the right side. Now we would like to obtain the inaccuracy score for the fifth point shown at the red dot. In the fifth column, there are three other points, which are the fourth point, the sixth point, and the tenth point. These three points are neighbors for the fifth point specified by the contact map. The fifth point is at the red cross on the 3D reconstruction, and the neighbors correspond to the magenta circles. First, we obtain the squared distances of the current fifth point to each of the corresponding neighboring points (fourth, sixth, and tenth points) in the 3D reconstruction. Then, we take their average and calculate its square root, which is the inaccuracy score here. **c**. The proportion of uniquely assigned reads among the total Hi-C contact reads at each cell stage. Standard deviation (s.d.) at each cell stage is shown. Data are represented as mean ± s.d.. **d.** Bar plots showing s.d. of the inaccuracy score of individual samples at 1 cell, 2 cell, 4 cell, 8 cell and 64 cell stages. The order is sorted by the s.d. of the inaccuracy score. **e**. Two examples of three-dimensional reconstruction of a nucleus at 2 cell stage (cell2_91 and cell2_56) by the RPR method. The inaccuracy score is shown by the heatmap color. A cell with a relatively high s.d. of inaccuracy score (left, cell2_91) and a cell with a relatively low s.d. of the inaccuracy score (right, cell2_56) are shown.

Figure 2: Spatial intermixing of paternal and maternal genomes during early stage development

a. Three-dimensional reconstruction of representative single nuclei at different developmental stages. Paternal and maternal alleles are labeled in different colors (maternal, red; paternal, blue). **b**. Box plots showing the paternal-maternal allele intermixing score at different developmental stages. Each dot indicates the score of individual cells. The p value was calculated by Mann Whitney U test with Bonferroni correction. **c**. Three-dimensional reconstruction of representative single nuclei at 64 cell stage. Paternal and maternal alleles are labeled in different colors (maternal, red; paternal, blue), as in **a**. Note that the cell in the right side shows less intermixing of parental alleles than that in the left side.

Figure 3: Formation and dissolvement of Rabl-like configuration during early stage development

a. Three-dimensional reconstruction of representative single nuclei at different developmental stages. The terminus of the short arm and that of the long arm are highlighted in red and blue, respectively. **b**. The mean chromosomal position (the short arm terminus = 0, the long arm terminus = 1) of proximity pairs of each genomic segment ("Proximal Position Score"; y-axis). The chromosomal position of segment along the chromosomal axis (x-axis, the short arm terminus = 0, the long arm terminus = 1, all chromosomes were divided). s.d. for each segment of the averaged chromosome is shown.

Figure 4: Establishment of discrete chromosome territories during early stage development

a. Box plots showing the chromosome territory score at different developmental stages. Each dot indicates the chromosome territory score of individual cells. The p value was calculated by Mann Whitney U test with Bonferroni correction. **b**. Three-dimensional reconstruction of a representative single nucleus at different developmental stages. Individual chromosomes are labeled in different colors. **c**. Box plots showing the chromosome territory score of each chromosome at different developmental stages.

Figure 5: Changes in the chromosomal shape during early stage development

a. Box plots showing the averaged distance between the termini of each chromosome at different developmental stages. **b**. Box plots showing the distance between the termini of each chromosome at different developmental stages.

Extended Figure 1

a. The inaccuracy score of Fig. 1d (right) is shown with a different range of the heatmap color. **b**. Histograms showing the number of cells at each range of s.d. of the inaccuracy score at each stage. **c**. The distribution of s.d. of the inaccuracy score of each cell at all stages. The order is sorted by the s.d. of the inaccuracy score. We excluded cells with s.d. greater than 0.15 from the following analyses in this study. **d**. Scatter plots showing the correlation between s.d. of the inaccuracy score and the number of Hi-C contact reads in each cell.

Extended Figure 2

Definition of a proximal set of segments. We defined proximity pairs (the genomic segments located in a proximity) as the top (nearest) 5% of all pairs. These proximity pairs were used for further analyses (Figures 2, 3 and 4).

Figure 1

Kitanishi et al.

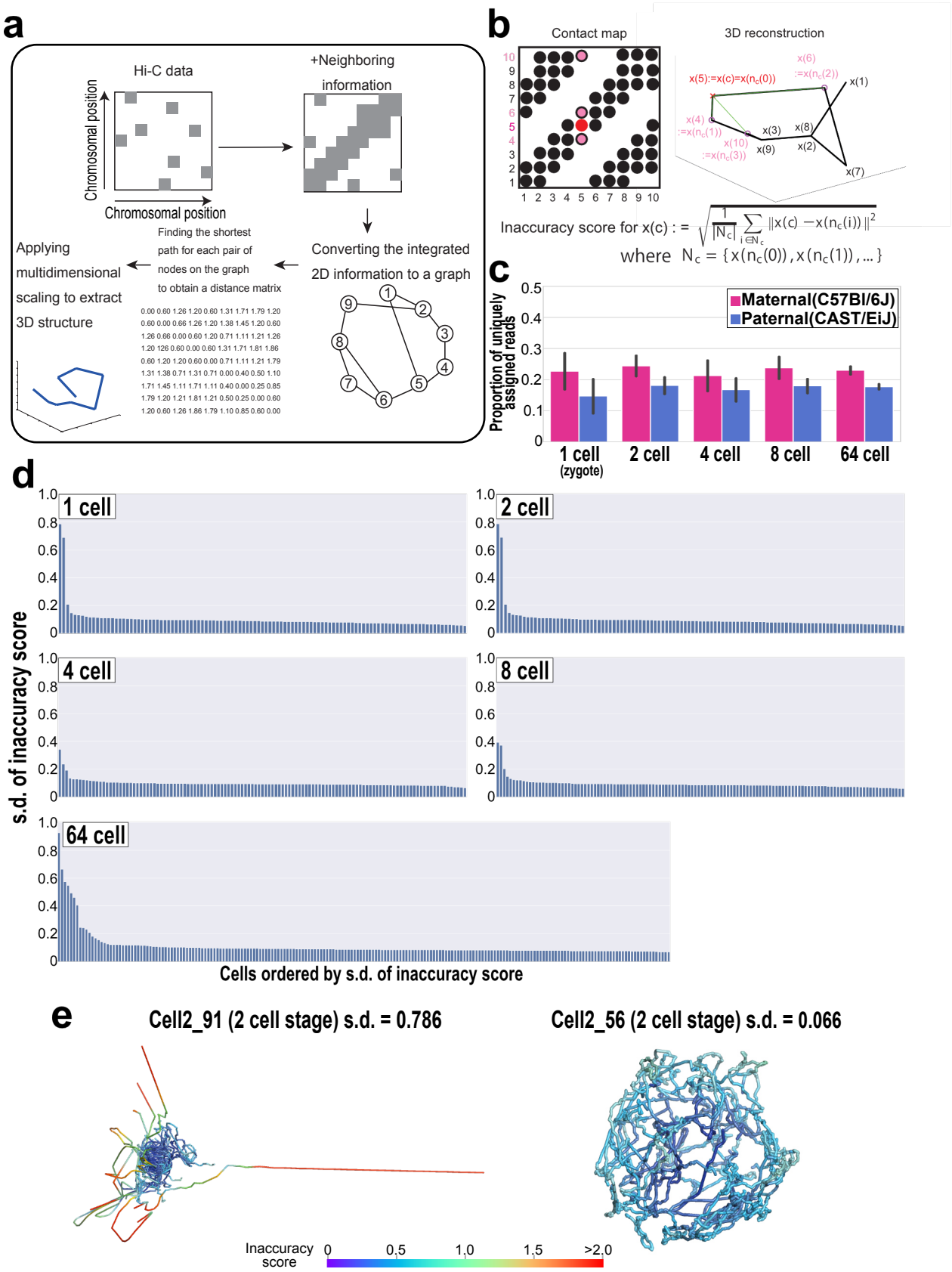


Figure 2

Kitanishi et al.

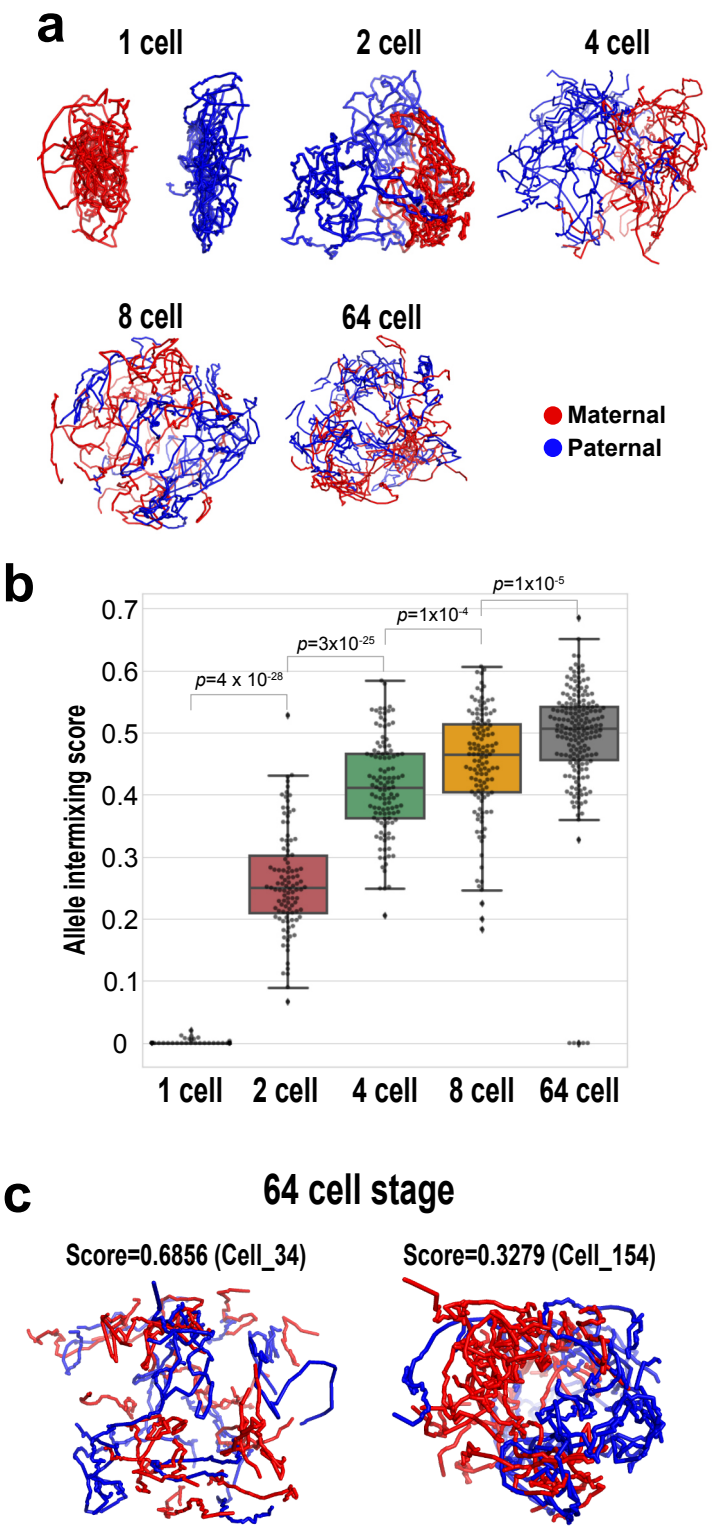


Figure 3

Kitanishi et al.

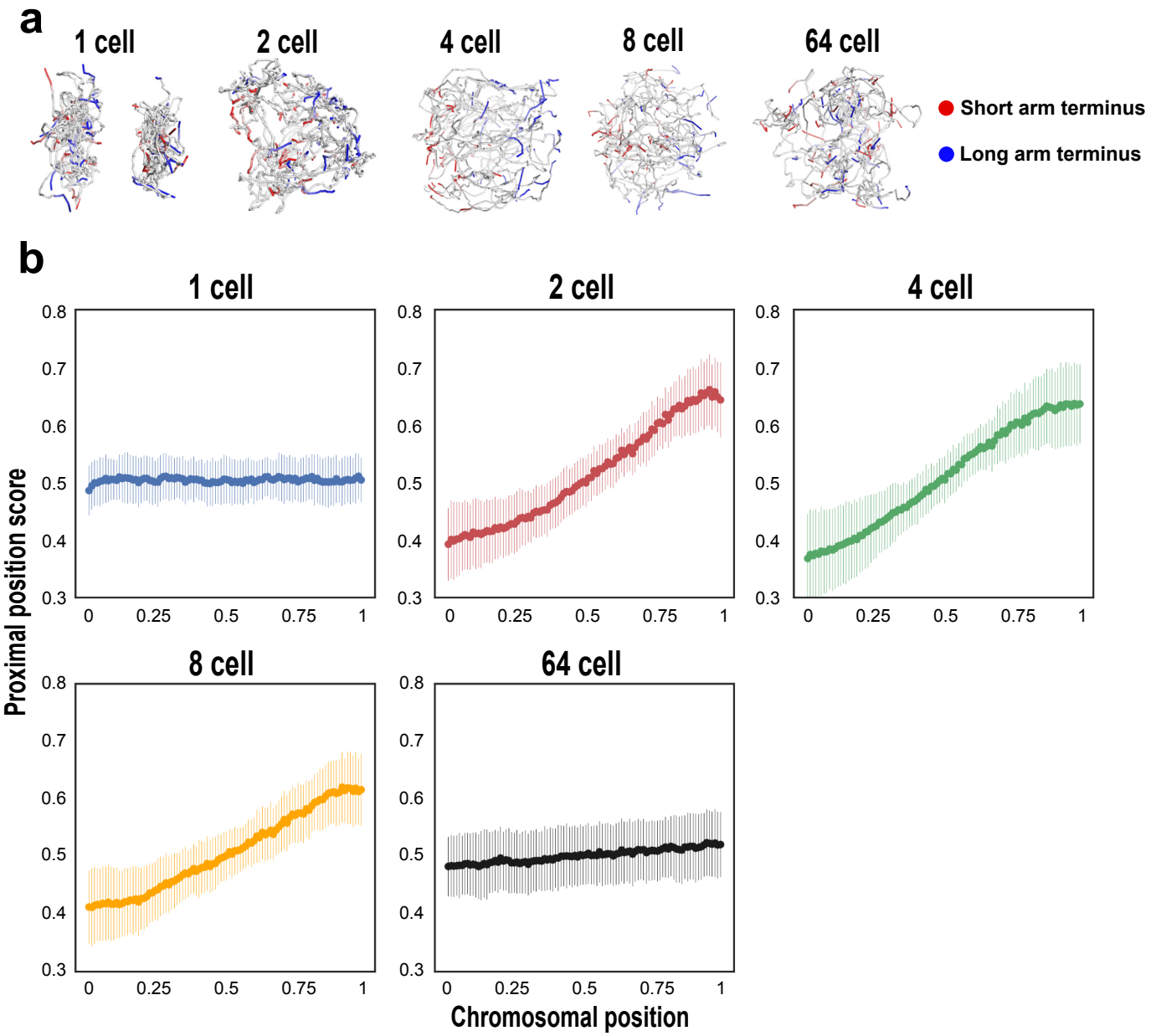


Figure 4

Kitanishi et al.

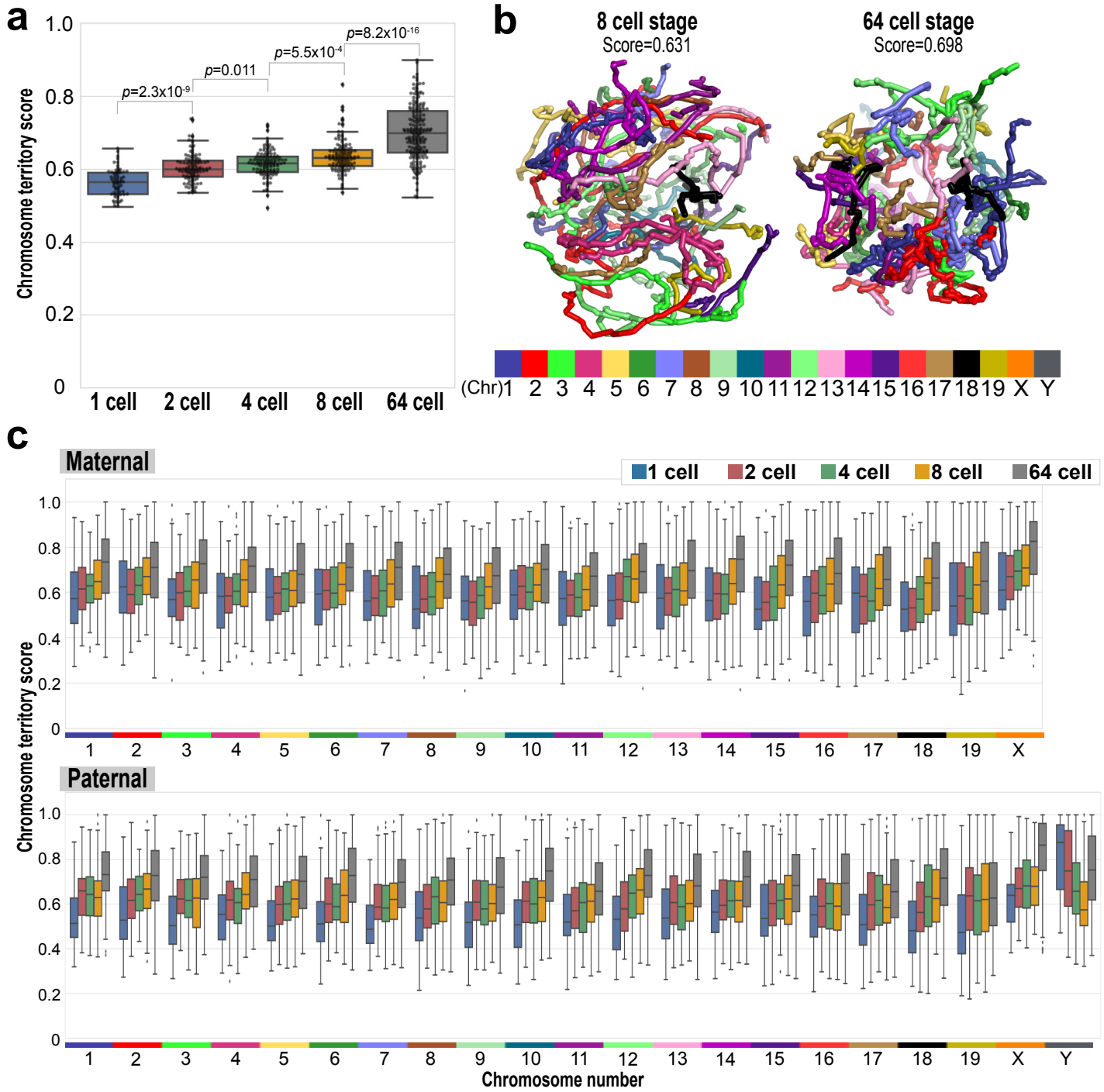
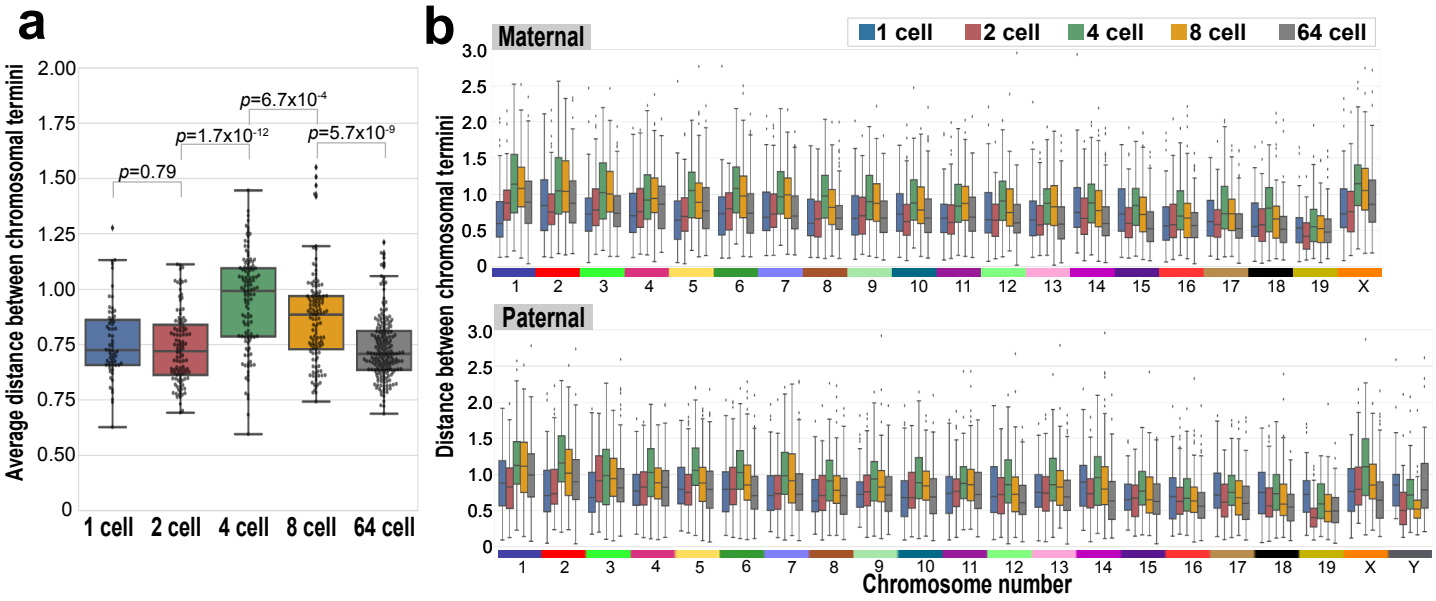


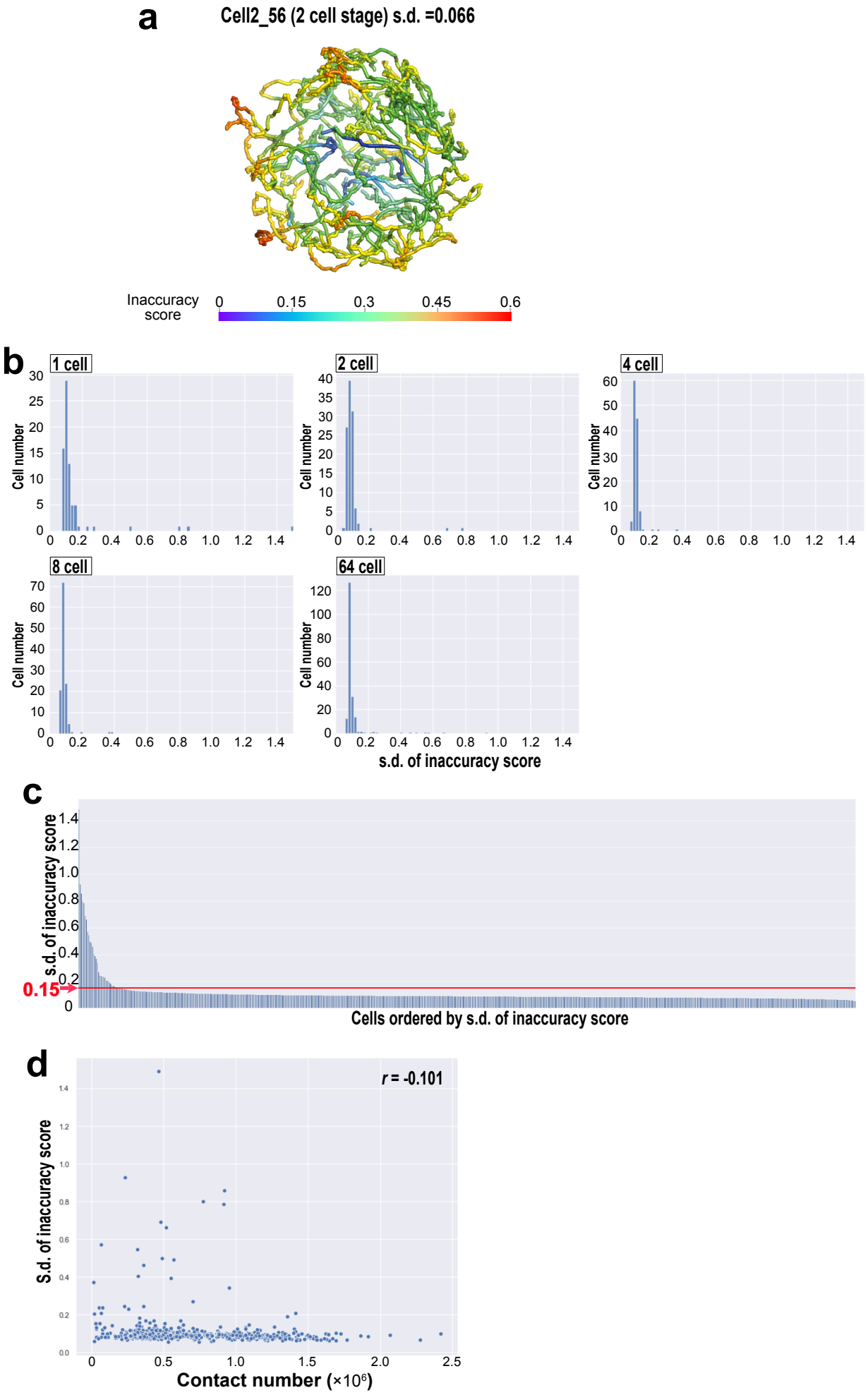
Figure 5

Kitanishi et al.



Extended Figure 1

Kitanishi et al.

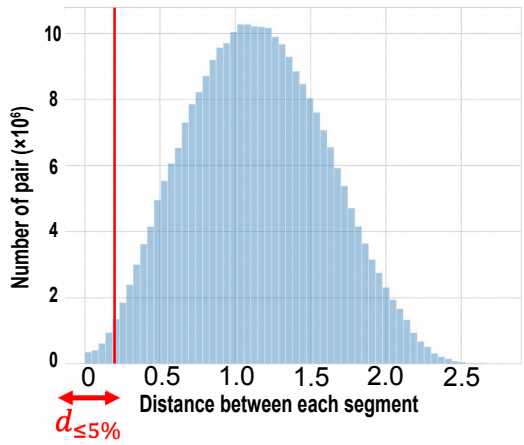


Extended Figure 2

Kitanishi et al.

Histogram of distance between every pair $x(i), x(j)$ of 1 Mb genomic segments

(e.g. Cell64_105)



Distribution of $||x(i) - x(j)||$

$i \neq j$

$i, j \in \{0,1, \ldots, N\}$

N: total number of 1Mb segments

A set of (i,j) satisfying

$||x(i) - x(j)|| \leq d_{5\%}$

was defined as a proximal set of segments