Dissecting the co-segregation probability from genome architecture mapping

Lei Liu1,*, Xinmeng Cao1, Bokai Zhang2, and Changbong Hyeon3,*

1Key Laboratory of Optical Field Manipulation of Zhejiang Province, Department of Physics, Zhejiang Sci-Tech University, Hangzhou 310018, China
2School of Physical Science and Technology, Southwest University, Chongqing 400715, China
3School of Computational Sciences, Korea Institute for Advanced Study, Seoul 02455, Korea
*Correspondence: leiliu@zstu.edu.cn or hyeoncb@kias.re.kr

ABSTRACT The genome architecture mapping (GAM) is a recently developed methodology that offers the co-segregation probability of two genomic segments from an ensemble of thinly sliced nuclear profiles, enabling to probe and decipher the 3D chromatin organization. The co-segregation probability from GAM, which typically probes the length scale associated with the genomic separation greater than 1 MB, is, however, not identical to the contact probability obtained in Hi-C, and its correlation with inter-locus distance measured with FISH is not so good as the contact probability. In this study, by using a polymer-based model of chromatins, we derive a theoretical expression of the co-segregation probability as well as that of the contact probability, and carry out quantitative analyses of how they differ from each other. The results from our study, validated with in-silico GAM analysis on 3D genome structures from FISH, suggest that to attain strong correlation with the inter-locus distance, a properly normalized version of co-segregation probability needs to be calculated based on a large number of nuclear slices ($n > 10^3$).

SIGNIFICANCE By leveraging a polymer model of chromatin, we critically assess the utility of co-segregation probability captured from GAM analysis. Our polymer model, which offers analytical expressions for the co-segregation probability as well as for the contact probability and inter-locus distance, enables quantitative comparison between the data from GAM, Hi-C, and FISH. Although the plain co-segregation probabilities from GAM are not well correlated with inter-locus distances measured from FISH, properly normalized versions of the probability calculated from a large number of nuclear profiles can still reasonably represent the inter-locus distance. Our study offers instructions of how to take full advantage of GAM analysis in deciphering 3D genome organization.

INTRODUCTION Among a number of experimental methods to decipher the three-dimensional (3D) chromosome/genome structure at high resolution (1–13), a recently developed genome architecture mapping (GAM) (8, 9, 14), enabling genome-wide mapping of chromatin contacts, has gained much attention. In GAM, cells are first fixed and cryosectioned, and next processed through laser microdissection to produce ultrathin nuclear slices, called nuclear profiles (NPs). The sequencing of the DNA content in each NP allows one to identify genomic loci present in the NPs and calculate their frequencies (14) (Fig. 1A). Loci that are distant in space are expected to have smaller chance to be co-sectioned in the same NP. It has been presumed that the co-sectioning/co-segregation probability between two genomic sites ($c_{ij}$, i and j) is inversely related to their mean spatial distance ($\bar{r}_{ij}$) (14).

GAM has several key advantages over C-based techniques in that it is ligation-free (12, 15). Furthermore, in contrast to Hi-C which requires millions of sample cells, only hundreds of cells may suffice for GAM
to generate a robust genome-wide co-segregation map. Given that clinical samples are often limited in number and given in a sectioned form, GAM can be more practical than other methods when studying disease-related genome reorganization. While not easily accessible in Hi-C, a number of important properties of 3D genome, such as higher-order chromatin contact (16–21), lamin and nuclear body association (22–26), can be measured by leveraging cryoFISH-combined GAM.

Figure 1: Comparison between GAM, Hi-C, and FISH. (A) Schematics illustrating the three methods. From an ensemble of cells, Hi-C measures the cross-linking frequencies between two genomic loci; GAM measures the co-segregation frequencies in a nuclear profile; FISH measures the inter-locus spatial distances. (B-D) GAM data are compared with FISH that probed the chromosomes of mouse ESC using DNA seqFISH+ method (24), which offers the 3D coordinates of 2,460 loci spaced approximately 1 Mb apart across the whole genome in 446 cells. Heatmaps of the percentile rank (PR) of the mean inter-locus distance ($\bar{r}_{ij}^{FISH}$) on the chromosome 3 binned at 1 Mb (right, bottom corner of each panel) (24) versus (B) PR of Hi-C contact probability ($p_{ij}^{Hi-C}$), (C) PR of co-segregation probability from GAM ($c_{ij}^{GAM}$), (D) PR of normalized linkage disequilibrium (14) (NLD, $d_{ij}^{GAM}$), and (E) PR of normalized point-wise mutual information (27) (NPMI, $s_{ij}^{GAM}$), each of which is demonstrated on the top left corner of the panel. (F) The scatter plots of PRs calculated in (B)-(E). The density of data point is color-coded from blue to red. (G) The Spearman’s rank correlation coefficient ($|\rho_s| = -\rho_s$) of $p_{ij}^{Hi-C}$, $c_{ij}^{GAM}$, $d_{ij}^{GAM}$, and $s_{ij}^{GAM}$ against $\bar{r}_{ij}^{FISH}$ for the 19 autosomes of mouse ESCs.

The potential of GAM is built upon the premise that the data acquired from the mapping faithfully reflect the 3D organization of genome. To demonstrate the utility and fidelity of GAM, Beagrie et al. (14) used co-segregation frequencies (or probabilities, $c_{ij}$) to identify topologically associated domains (TADs) that are interacting with each other. Referencing to cryoFISH images, they showed that the interacting TADs have higher contact probabilities ($p_{ij}$) and smaller spatial distances ($r_{ij}$) than non-interacting TADs. More recently, by using the Strings and Binders Switch model of chromatin, Fiorillo et al. (27) reported
almost perfect correlations between \(c_{ij}\) and \(\bar{r}_{ij}\) with a Spearman’s ranking correlation coefficient of \(\rho_s < -0.98\) in four regions of genomic sizes < 7 Mb. However, given that those results were obtained with a limited number of samples, how statistically general is the inverse relationship between \(\bar{r}_{ij}\) and \(c_{ij}\) still remains as an open question. To be specific, when the co-segregation probability of a particular chromatin segment pair is greater than another pair (\(c_{ij} > c_{kl}\)), can we assert that their mean distances always satisfy the inequality \(\bar{r}_{ij} < \bar{r}_{kl}\)?

To address the above question, we calculate \(\bar{r}_{ij}\) by using 3D imaging data of mouse embryonic stem cell (ESC) (24), and compare them with \(c_{ij}\) obtained from GAM (24) by means of their percentile rank (PR), i.e., PR(\(\bar{r}_{ij}^{\text{FISH}}\)) vs. PR(\(c_{ij}^{\text{GAM}}\)), where the superscripts “FISH”, “GAM” and “Hi-C” are added to specify the experimental data source for clarity. For the case of \(\bar{r}_{ij}\), we consider that PR is higher when \(\bar{r}_{ij}\) is smaller. As shown in Fig. 1B, PR(\(\bar{r}_{ij}^{\text{FISH}}\)) depicted in a matrix form for the chromosome 3 displays a pattern similar to the PR of the contact probability \(p_{ij}\) from Hi-C. However, for the cases of GAM-based co-segregation probability (\(c_{ij}\)) and its variants, – the normalized linkage disequilibrium (NLD, \(d_{ij}\)) (14) and the normalized point-wise mutual information (NPMI, \(s_{ij}\)) (27) (see Materials and Methods) – such similarity between the patterns is significantly weaker (Fig. 1C-E). The scatter plot of PR(\(p_{ij}\)) against PR(\(\bar{r}_{ij}\)) also displays stronger correlation than those of PR(\(c_{ij}\)), PR(\(d_{ij}\)), and PR(\(s_{ij}\)) (Fig. 1F).

The Spearman’s rank correlation of \(p_{ij}\) with \(\bar{r}_{ij}\) is significantly greater than that of GAM-based measures (\(c_{ij}\), \(d_{ij}\), \(s_{ij}\)) for all autosomes of mouse ESCs (Fig. 1G).

In the light of the general agreement between Hi-C and FISH (3, 24, 28–31), the significantly weaker correlation of co-segregation frequencies from GAM with the mean inter-locus distances from FISH, demonstrated in Fig. 1, is rather surprising. To better understand the meaning of data from GAM, we consider an analytically tractable, simple polymer model, representing chromosomes inside a nucleus, and derive co-segregation probability (\(c_{ij}\)), which shows that the \(c_{ij}\) changes with the coverage, the thickness, and the number of nuclear slices. We validate our theoretical predictions from the polymer model by carrying out in-silico GAM analysis on a publicly available 3D genome structure dataset. Our study, which enables quantitative comparison between GAM, Hi-C, FISH measurements, offers instructions of how to take full advantage of GAM analysis in deciphering 3D genome organization.

**MATERIALS AND METHODS**

**GAM, Hi-C, FISH and DamID datasets**

The genome-wide GAM data was downloaded from GEO repository (14) (GSE64881). It contains binary information, denoting either the absence or the presence of chromatin segments binned at 1 Mb, in each of 408 NPs, for all autosomes of mouse ESCs. For each chromosome, we counted the frequency of the \(i\)-th segment and the frequency of the \((i, j)\) segment pair in the same profile, which yields the segregation and co-segregation probability, \(c_i\) and \(c_{ij}\), respectively.

We used the Hi-C dataset of mouse ESC measured by Dixon et al. (5). KR-normalized intra-chromosome contact probability at 1Mb resolution, \(p_{ij}\), was calculated by using the Juicer toolbox (32).

The DNA seqFISH+ dataset was fetched from Zenodo database (https://zenodo.org/record/3735329), where 3D coordinates of genetic loci spaced approximately 1 Mb apart across the whole genome in 446 cells were deposited. The genomic coordinates of the imaged loci were converted from mm10 to our reference genome assembly mm9 with the UCSC Genome Browser utility liftOver (33). Following Takei et al. (24), we separated two homologous chromosomes in each mouse ESC based on the consensus of the spectral and hierarchical clustering of imaged loci. For each intra-chromosome loci pair \((i, j)\) binned at 1 Mb, we then calculated their spatial distances for each allele in all cells, which yields the mean inter-locus distance \(\bar{r}_{ij}\).

The DamID of Lamin B1 protein in mouse ESCs was downloaded from GSE17051 (34), where the base-2 logarithm of the fold enrichment of the interactions between chromatin loci and nuclear lamin (\(q_{L1,B1}\)) was available. A chromatin segment has a positive (negative) value of \(q_{L1,B1}\) if it has higher (lower) chance to associate with lamin than the genome-wide average level.
Two variants of co-segregation probability

Of several possible methods for normalizing GAM data (35), we consider the two most popular ones, normalized linkage disequilibrium (NLD) and normalized pointwise mutual information (NPMI).

(i) To account for the observation that different loci have different chances to be cryosectioned, the normalized linkage disequilibrium (NLD), which was originally proposed in population genetics to calculate the non-random association of two alleles at different loci (36, 37), has been employed for the analysis of GAM (14). The NLD is defined as

\[ d_{ij} = \frac{D_{ij}}{D_{ij}^{\max}}, \]  

where

\[ D_{ij} = c_{ij} - c_i c_j \]  

compares the co-segregation probability of \( i \) and \( j \)-th loci \((c_{ij})\) to the probability of statistical independence, normalized by the theoretical maximum

\[ D_{ij}^{\max} = \begin{cases} \min \left( c_i c_j, (1 - c_i)(1 - c_j) \right) & , \quad D_{ij} < 0 \\ \min \left( c_i(1 - c_j), c_j(1 - c_i) \right) & , \quad D_{ij} > 0. \end{cases} \]  

(ii) Alternatively the normalized pointwise mutual information (NPMI) has been used for the GAM analysis as well (27):

\[ s_{ij} = -\frac{\log \left( \frac{c_{ij}}{c_i c_j} \right)}{\log(c_{ij})}, \]  

where the distance of cosegregation probability of \( i \) and \( j \)-th loci is measured in logarithmic scale (\( \log c_{ij} \)) from that of the statistical independence (\( \log c_i c_j \)) and normalized by \( \log c_{ij} \).

The two quantities, \( d_{ij} \) and \( s_{ij} \), bounded between -1 and 1, are conceptually similar in that both measure the distance of the joint probability of cosegregation from the case of statistical independence (i.e., \( c_{ij} = c_i c_j \)); however, they differ from each other in that the measurement by the \( d_{ij} \), which amounts to the covariance (or correlation), is restricted to linear relationships, while \( s_{ij} \) can capture more general relationship between two random variables (38).

RESULTS AND DISCUSSIONS

Theoretical analyses

Heterogeneous loop model

We use the Heterogeneous Loop Model (21, 39–45) (HLM), to derive analytical expressions for the contact and co-segregation probabilities as well as mean inter-locus distances, and study the relations (or correlations) between them. In HLM, chromatin fibers are modeled as a linear polymer chain composed of \( N \) coarse-grained segments, each with a prescribed genomic size. We assume that the effective energy potential of chromatin can be described by a sum of harmonic restraints on the spatial distances between all segment pairs,

\[ U_K(r) = \sum_{i=0}^{N-1} \sum_{j=i+1}^{N-1} \frac{k_{ij}}{2} |\vec{r}_i - \vec{r}_j|^2, \]

\[ = \frac{1}{2} r^T \cdot L \cdot r \]  

where \( r = (\vec{r}_0, \vec{r}_1, \vec{r}_2, \cdots, \vec{r}_{N-1})^T \) specifies the 3D structure of the polymer chain, and the \( N \times N \) Laplacian matrix \( L \) is defined as \( L = D - K \) where \( K \) is a stiffness matrix of element \( k_{ij} \) and \( D \) represents a
Theoretical analysis of genome architecture mapping

A diagonal matrix with \( D_{ii} = \sum_j k_{ij} \). The probability of the chromatin to adopt a particular structure is then written as

\[
P(r) \propto e^{-U_K(r)/k_BT},
\]

where \( k_BT \) is the energy unit of the model with \( k_B \) the Boltzmann constant and \( T \) the absolute temperature. The model parameters \((K)\) for a genomic region of interest can be determined based on Hi-C data.

The greatest advantage of HLM is that structural and dynamic properties of chromosome can be directly derived based on Eqs. S1 and 6 along with Hi-C data. HLM and its variants have been exploited to study experimental measurements (21, 40–45). The contact probability calculated from the HLM is in excellent agreement with measurements (21, 41, 42). Specifically, despite the cell-to-cell variability of 3D genome over population (45,46), the intra-chromosomal inter-locus spatial distance distributions predicted by HLM can still be validated against those measured from DNA seqFISH+ imaging (see Fig. 1 in Ref. (21)).

Although it does not affect the discussion in this work, the HLM has limitations when it gets to a resolution greater than \( O(10^2) \) base pairs, where each monomer represents the scale of 1-2 nucleosomes, whose interactions (bending, torsion, stacking, etc.) can no longer be effectively approximated by harmonic restraints.

**Contact probability and spatial distance between two genomic segments**

After transforming \( K \) into a covariance matrix \( \Sigma \) whose matrix element is denoted as \( \gamma_{ij} \equiv \Sigma_{ij} \), one finds the mean spatial distance between the \( i \)-th and \( j \)-th segments averaged over all possible chromosome configurations as (detailed derivations are given in the Supplementary Information)

\[
\bar{r}_{ij} = \frac{2}{\sqrt{\pi \gamma_{ij}}}
\]

with \( \gamma_{ij} \equiv \frac{1}{2} \left( \sigma_{ii} + \sigma_{jj} - 2\sigma_{ij} \right)^{-1} \), and their contact probability

\[
p_{ij}(r_c) = \left( 1 + \frac{3}{2\gamma_{ij}r_c^2} \right)^{-3/2}
\]

where \( r_c \) is the effective capture radius of the cross-linking agent, and is the only tunable parameter of the contact probability in Hi-C.

In this theoretical framework, the two observables \( \bar{r}_{ij} \) and \( p_{ij} \) are related as \( p_{ij} = \left( 1 + 3\pi\bar{r}_{ij}^2/8r_c^2 \right)^{-3/2} \), and hence they are in perfect correlation, giving rise to the Spearman’s rank correlation coefficient \( \rho_s(p_{ij}, \bar{r}_{ij}) = -1 \). The consistency between Hi-C \( p_{ij} \) and FISH \( \bar{r}_{ij} \) shown in Fig. 1B and G has been demonstrated by a number of experimental studies (28–30, 47).

In addition, we trained HLM for each autosome of mouse ESCs by using their Hi-C data binned at 1 Mb, and calculated contact probability and mean inter-locus distance based on Eq. S5 and S4, respectively. As shown in Fig. S1, not only \( p_{ij} \) from HLM is highly correlated with that from Hi-C \( (\rho_s \gtrsim 0.95) \), but also \( \bar{r}_{ij} \) from FISH can be well predicted with the correlation coefficient \( \rho_s \approx 0.9 \), which is even slightly higher than the correlation between Hi-C and FISH data. These results suggest that HLM is a proper 3D model of chromosome.

**Co-segregation probability**

The probability of the \( i \)-th genomic segment being in a horizontal slice \( S^3 \) with a thickness \( \Delta \) sectioned at a height \( z \) relative to the center of mass (COM) of the chain is derived as follows (see Supplementary Information)

\[
c_i(z, \Delta) = \frac{\Delta}{\sqrt{\sigma_{ii}} + \Delta^2} e^{-\frac{1}{2}z^2/(\sigma_{ii}+\Delta^2)},
\]
Figure 2: Co-segregation of monomers in a Gaussian phantom chain. (A) Depicted are 100 chains randomly selected from an ensemble, each composed of 20 monomers. GAM data would be collected from an ensemble of slices with a thickness $\Delta$, and at a height $z$ ($|z| < R$). The terminal and the central parts of the chain are colored in red and blue, respectively. (B) Segregation and (C) co-segregation probabilities, $c_i^o$ and $c_{ij}^o$, calculated at different values of $z$ with $\Delta = 1$. (D) (Left panel) Percentile rank (PR) of the co-segregation probability (top left) compared with the PR of the mean inter-locus distance (bottom right). (Right panel) Scatter plot of (PR($c_{ij}$), PR($\bar{r}_{ij}$)). The density of data point in each pixel is color-coded from blue to red. (E) Spearman’s rank correlation coefficient ($\rho_s$) between $c_{ij}$ and $\bar{r}_{ij}$ as a function of $R$ and $\Delta$, where the co-segregation probabilities were calculated by using $n$ slices. The $|\rho_s|$ with $n \to \infty$ is obtained by using Eqs. S4 and S14. The condition of ($R^*, \Delta^*$) that gives rise to the strongest correlation $|\rho_s|$ for different $n$ is marked with the symbol $\times$.

and the co-segregation probability is formulated as

$$c_{ij}^o(z, \Delta) = \frac{\Delta^2}{\sqrt{\Psi_{ij}}} e^{-\Delta/\sqrt{\Omega_{ij}}},$$

(10)

where

$$\Psi_{ij} = \sigma_{ii} + \sigma_{jj} - 2\sigma_{ij} + 2\Delta^2$$

$$\Omega_{ij} = (\sigma_{ii} + \Delta^2)(\sigma_{jj} + \Delta^2) - \sigma_{ij}^2.$$

Eqs. S4-S11 are the general theoretical expressions derived in the framework of HLM with $\sigma_{ij}$ (or the force parameters $k_{ij}$) left unspecified. Their validity can be examined against a Gaussian phantom chain consisting of 20 monomers, which is a special case of HLM where the condition of $k_{ij} = 1$ for $|i - j| = 1$ and $k_{ij} = 0$ otherwise is assigned. By numerically generating an ensemble of 100,000 polymer chain configurations whose COM is restrained to the origin of the coordinate system (Fig. 2; see Supplementary Information for numeric details), we counted the frequency of the $i$-th monomer $c_i^o$ (or the joint frequency of the $i$- and $j$-th monomers, $c_{ij}^o$) in a horizontal slice $S^A_z$. The numerical results of (co-)segregation
Theoretical analysis of genome architecture mapping

Figure 3: Co-segregation probability from the HLM-based structural ensemble at a resolution of 50 kb for a 1-Mb genomic region on chromosome 5 (chr5:97,800,000-98,800,000) of GM12878 cells. (A) The contact probability ($p_{ij}$) measured by Rao et al. using Hi-C (6) (top-left) is compared with that predicted by HLM (bottom-right). Details about the determination of model parameters based on Hi-C data and reconstruction of 3D chromosome structures can be found in the Refs. (21). (B) An ensemble of chromatin chains ($N = 30$) randomly selected from the most populated state of the structural ensemble. Color-coded are three distinct domains. (C) An ensemble of 100 randomly selected chromatin structures without alignment. Co-segregation was determined from an ensemble of slices of a thickness $\Delta$, and a stochastic height $z$ satisfying $|z| < R$. (D) Segregation and (E) co-segregation probabilities, $c_i^\Delta$ and $c_{ij}^\Delta$, calculated at different values of $z$ with $\Delta = 1$. (F) (Left panel) PR of the co-segregation probability (top left) compared with that of the mean inter-locus distance (bottom right). (Right panel) Scatter plot of (PR($c_{ij}$), PR($\bar{r}_{ij}$)). (G) Spearman’s rank correlation coefficient ($\rho_s$) of $c_{ij}$ against $\bar{r}_{ij}$ as a function of $R$ and $\Delta$. The symbol $\times$ marks the condition of $(R^*, \Delta^*)$ that gives rise to the strongest correlation at different $n$.

The probability obtained from explicit 3D structures (Fig. 2A) are in perfect agreement with our analytic expressions in Eqs. S10 and S11 (Figs. 2B and C).

Eqs. S10 and S11 make it explicit that the (co-)segregation probability not only depends on the thickness, but also on the location of the slice. As expected, $c_i^\Delta$ and $c_{ij}^\Delta$ increases with the thickness of the slice. But, their variation with $z$ is nontrivial. The radius of the polymer ensemble depicted in Fig. 2A is $R \approx 5$. Thus, in a NP sectioned at $z = 4$, the chain segments located at the terminals have a higher odds to be sliced than those around the center, and this trend is reversed in another NP sectioned at $z = 0$ (Fig. 2B).
Figure 4: In silico GAM on the DNA seqFISH+ dataset (24). (A) A typical 3D image of the genome of mouse ESC where different chromosomes are distinguished by different colors. Shown on the right is a slice, in which the sectioned loci are color-coded by their lamin association propensities ($q_{LB1}$). (B) PR of $c_{ij}$ sectioned by slices with $\Delta = 0.5 \mu m$ and at $z = 0$ and 7 $\mu m$, shown in the top-left and bottom-right panel, respectively. The Spearman’s rank correlation coefficient between the two $c_{ij}$’s is $\rho_s = 0.83$. The Spearman’s rank correlations ($|\rho_s|$) of $\bar{r}_{FISH}^{ij}$ against (C) $c_{ij}^{FISH}$, (D) $d_{ij}^{FISH}$ (NLD), and (E) $s_{ij}^{FISH}$ (NPMI) as a function of the slice range of nuclei ($R$) and slice width ($\Delta$) with varying sample size ($n$), where the “x” symbol marks the value of ($R, \Delta$) that maximizes $|\rho_s|$. The heatmaps in (C-E) quantify the variations of correlation over the full range of their respective data.

As shown in Fig. 2A, the chain segments colored based on their position along the chain contour (blue for the segments around the center and red for the segments at the two ends) visualize the origin of the $z$-dependencies. Due to the geometrical restraint on the COM of an individual chain, the chain terminals lie at the periphery, and hence the distribution of chain segments is radially non-uniform. The ranking order of $c_{ij}^{\alpha}$ demonstrates qualitative difference with $z$ as well (Fig. 2C).

In GAM, NPs are collected from many samples, i.e., random slices in position and orientation over many nuclei (cells). Provided that the slice range of nuclei is $[-R, R]$, the mean segregation probability averaged over the uniform distribution of $z \in [-R, +R]$, which corresponds to data resulting from the GAM analysis for a large number of slices ($n \gg 1$), would be obtained as

$$c_i(R, \Delta) = \frac{1}{2R} \int_{-R}^{R} c_i^{\alpha}(z, \Delta) dz$$

$$= \sqrt{\frac{\pi \Delta}{2R}} \text{erf}\left(\frac{R}{\sqrt{2(\sigma_{ii} + \Delta^2)}}\right)$$

(11)
with \( \text{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x dt e^{-t^2} \). Similarly, the mean co-segregation probability is obtained as

\[
c_{ij}(R, \Delta) = \frac{\bar{\Psi}_{ij}}{2} \Psi_{ij}^{-1/2} \text{erf} \left( \sqrt{\frac{\Psi_{ij}}{2\Omega_{ij}}} \right).
\]  

Eq. S14 suggests that \( c_{ij} \) is decided by the range of the nuclei being sliced, namely \([-R, +R]\). Unlike \( \Psi_{ij} \), which is a monotonically decreasing function of the mean distance, \( \Psi_{ij} = (2\gamma_{ij})^{-1} + 2\Delta^2 = \pi R_{ij}^2/8 + 2\Delta^2, \Omega_{ij} \) changes non-monotonically with \( \bar{r}_{ij} \). Thus, the presumption of a monotonic relation between \( c_{ij} \) and \( \bar{r}_{ij} \) does not hold in the outcomes derived from HLM. To illustrate this, in Fig. 2D, \( \text{PR}(c_{ij}) \) clearly differs from \( \text{PR}(\bar{r}_{ij}) \).

To demonstrate the effect of the sample size (\( n \)) on the Spearman’s rank correlation between \( c_{ij} \) and \( \bar{r}_{ij} \), we divided the ensemble of polymer chain structures into 100,000/\( n \) replicas, calculated \( c_{ij} \) for each replica, and plotted the replica-averaged value of \(|\rho_x|\) as a function of \( R \) and \( \Delta \) in Fig. 2E. As \( n \) increases from \( n = 10^3 \) to \( n \to \infty \), the value \((R^*, \Delta^*)\) that maximizes \(|\rho_x|\) shifts towards large \( R \) and small \( \Delta \).

The segregation (Eq. 11) and co-segregation probabilities (Eq. S14) were derived by assuming that the sectioning was made with Gaussian probability at height \( z \) and that the NPs were collected uniformly over the volume nucleus. They can be derived by assuming other models of the sectioning probability and slice position profile; however, the resulting probabilities remain qualitatively the same (see the text and Figs. S2 and S3 in the Supplementary Information).

**Co-segregation probability of HLM-based chromosome model**

The behaviors of co-segregation probability obtained from Gaussian phantom chain are sufficiently general, but can be made more realistic. HLM of 1-Mb genomic region on chromosome 5 of GM12878 cells, trained in reference to its Hi-C data (Fig. 3A), produces an ensemble of structures characterized with three distinct domains (Fig. 3B). As found from the Gaussian phantom chain model (Fig. 2), we confirm the dependence of the (co-)segregation probability on the position of the slice (i.e., the dependence of \( c_{ij}^o \) and \( c_{ij}^f \) on \( z \)); see Figs. 3D, E), the nonuniform radial distribution of chromosome (Figs. 3C, D, E), and the improved correlation between \( c_{ij} \) and \( \tilde{r}_{ij} \) at greater \( n \) (Figs. 3F, G).

**Correlation of Hi-C and GAM against FISH**

**In silico GAM analysis of FISH data**

The interactions between chromatin segments can be mediated by many nuclear compositions such as nucleoli, lamins, granules, and so forth. Although such contributions may implicitly be accounted in the energy potential of HLM (Eq. S1), efficacy of our theory for the 3D structure modeling of whole genome remains unexplored. To this end, we next performed in-silico GAM analysis on the 3D genome structures measured by Takei et al. (24).

We randomly sampled 100,000 genome structures with replacement from the mouse ESC dataset of the DNA seqFISH+ experiment, and then translated the COM of each structure to the origin, and rotated the structure to a random orientation. Horizontal slices with a thickness \( \Delta \) were obtained from the structures by randomly sampling the position along the \( z \)-axis in the range of \( z \in [-R, +R] \) (Fig. 4A). The resulting ensemble of slices were partitioned into replicas, each containing \( n \) slices.

First, as demonstrated by the different lamin colocalization propensities of loci shown in Fig. 4A, the chromatin segments inside the cell nucleus has a non-uniform radial distribution. Second, we confirm that co-segregation probabilities \( (c_{ij}^o) \) calculated at the different positions of the nuclear slice \((z = 0 \text{ and } 7 \mu m)\) are qualitatively different from each other (Fig. 4B). More pronounced dependence of \( c_{ij}^o \) on \( z \) is found for other chromosomes (Figs. S4A and S5A). Third, the Spearman’s rank correlation of \( c_{ij}^\text{FISH} \) with \( p_{ij}^\text{FISH} \) follows the same trend with \( R \), \( \Delta \), and \( n \) (Fig. 4C) in accord with our theoretical analyses using the Gaussian polymer chain (Fig. 2E) and HLM for chromosome 5 of GM12878 cell (Fig. 3G). For \( n \gg 1 \), the correlation of three co-segregation probabilities with \( \bar{r}_{ij} \) is maximized at a large \( R \) and a small \( \Delta \). However, it is noteworthy that even with \( n = 100,000 \), the best correlation between \( c_{ij}^\text{FISH} \) and \( p_{ij}^\text{FISH} \) is about \(|\rho_x| \approx 0.63 \), which is significantly weaker than the correlation of \( p_{ij} \) of Hi-C with \( \bar{r}_{ij} \) \(|\rho_x| \approx 0.9 \);
see Fig. 1G). The other two \( c_{ij} \)-related quantities, NLD and NPMI, show better correlations with \( \bar{r}_{ij} \) than \( c_{ij} \); however, the number of NPs should be greater than 10^3 for the maximal values of \( |p_{ij}| \) for NLD and NPMI to exceed 0.9 (Fig. 4D and E). Similar conclusions are drawn from chromosomes 12 (Fig. S4) and 18 (Fig. S5).

**Stratified comparison of GAM with FISH**

Thus far, all intra-chromosome segment pairs were taken into account to calculate the \( \rho_s \). To see how the fidelity of GAM data in representing the chromosome architecture changes at different genomic scales, we re-calculated \( \rho_s \) against the mean spatial distance \( (\bar{r}_{ij}) \) from FISH (24) by restricting our analysis to the loci pairs separated by a certain genomic distance (48, 49). Despite large variations among different chromosomes (Fig. 5A and Fig. S6), at short genomic separation \( (|i - j| < 40 \text{ Mb}) \), Hi-C contact probability \( (p_{ij}) \) is significantly better correlated with \( \bar{r}_{ij} \) than the GAM-based co-segregation probability and its variants \( (c_{ij}, d_{ij}, s_{ij}) \). Only for the extremely long-range segment pairs \( (|i - j| > 75 \text{ Mb}) \), the GAM dataset marginally outperforms the \( p_{ij} \) from Hi-C. Given that \( p_{ij}^{\text{Hi-C}}(< 0.001) \) is typically orders of magnitude smaller than the corresponding \( c_{ij}^{\text{GAM}}(> 0.1) \) for long-range pairs, the bad performance of Hi-C might be an outcome of its relatively higher NSR (Eq. 14).

![Figure 5: Stratified comparison of GAM (14) and Hi-C (5) against FISH data (24).](image)

(A) Autosome-averaged Spearman’s rank correlation coefficient \( (\rho_s) \) between \( \bar{r}_{ij}^{\text{FISH}} \) and Hi-C contact probability \( (p_{ij}) \), GAM co-segregation probability \( (c_{ij}) \) and its two variants \( (d_{ij}, s_{ij}) \) as a function of the genomic separation between the two sites \( (i, j) \). The mean and the standard deviation of \( \rho_s \) averaged over all autosomes are plotted with solid lines and shades, respectively. (B) Comparison between the distance distributions of one intra-chromosomal pair \( (i, j) \) (blue) and another \( (k, l) \) based on FISH. \( P \)-value to reject the hypothesis \( (r_{ij} < r_{kl}) \) calculated from Mann-Whitney U test is shown on the top. The left (right) panel shows an example with a small (large) \( P \)-value. (C) Four possible ratios of a segment pair \( (i, j) \) to another \( (k, l) \) versus their \( P \)-value. The two examples in (B) are marked with the symbol \( (\times) \). The vertical line depicts \( P = 1.0 \times 10^{-5} \).

Lastly, for two intra-chromosomal pairs on the chromosome 3 separated by 4 Mb, say \( (i, j) \) and \( (k, l) \), we calculate a \( P \)-value between the distributions of two inter-locus distances from FISH (24) to assess the statistical significance of the statement that the spatial distance of a particular pair is shorter than another \( (r_{ij} < r_{kl}) \) instead of merely comparing their means \( (\bar{r}_{ij} \text{ versus } \bar{r}_{kl}) \). When the \( P \)-value is smaller, it is more likely that the two pairs display different distances (Fig. 5B); if \( r_{ij} < r_{kl} \), then \( p_{ij} > p_{kl} \) is highly likely. However, according to our explicit calculation summarized in Fig. 5C, the above statement regarding the relation between inter-locus distance and contact probability does not nicely translate into the GAM-based co-segregation probabilities. In Fig. 5C which plots the data points \( x_{ij} / x_{kl} \), where \( x_{ij} \) denotes either the contact \( (p_{ij}) \) from Hi-C or one of the GAM-related co-segregation probabilities \( (c_{ij}, d_{ij}, s_{ij}). \)
or \( s_{ij} \) as a function of their \( P \)-value, we find that the number of data points satisfying \( x_{ij}/x_{kl} > 1 \) among those with \( P < P^* = 1 \times 10^{-5} \), i.e.,

\[
    f_s = \frac{n[(x_{ij}/x_{kl} > 1) \text{ and } (P < P^*)]}{n[P < P^*]},
\]

(13)
is \( f_p = 0.93, f_c = 0.65, f_d = 0.84, \) and \( f_s = 0.87, \) leading to \( f_p > f_s \geq f_d \gg f_c \). The \( p_{ij} \) from Hi-C reflects the inter-locus distance more faithfully than the GAM-related co-segregation probabilities. Although the plain co-segregation probability from GAM (\( c_{ij} \)) is poorly correlated with \( r_{ij} \), the properly normalized versions of co-segregation probabilities, NLD (\( d_{ij} \)) and NPMI (\( s_{ij} \)), can still reasonably represent the inter-locus distances (\( r_{ij} \)).

CONCLUDING REMARKS

In practice, the sample size (\( n \)) in GAM is finite. With the mean (\( c_{ij} \)) and variance (\( \sigma_{c_{ij}}^2 = c_{ij}(1 - c_{ij})/n \)) of the co-segregation probability, the corresponding noise-to-signal ratio (NSR) is given by

\[
    \text{NSR}_{\text{GAM}} = \frac{\sigma_{c_{ij}}}{c_{ij}} = \frac{1 - c_{ij}}{c_{ij}n},
\]

(14)

which decreases with \( n \) as \( 1/\sqrt{n} \) (27). It is of note that the NSR of contact probability (\( p_{ij} \)) has a form identical to Eq. 14 when \( c_{ij} \) is replaced with \( p_{ij} \). For the contact and co-segregation probabilities, which display power-law decays (\( p_{ij}, c_{ij} \sim |i - j|^{-\alpha} \) (50), measurements are less reliable (or large NSR) for long-range segment pairs (Eq. 14). Furthermore, since the variations of the co-segregation probability with respect to the variations of \( \Delta \) and \( R \) satisfy \( \partial c_{ij}/\partial \Delta < 0 \) and \( \partial c_{ij}/\partial R > 0 \) for \( n \gg 1 \) (Eq. S14), the NSR_{GAM} is reduced for \( \delta \Delta < 0 \) and \( \delta R > 0 \), i.e.,

\[
    \delta(\text{NSR}_{\text{GAM}}) = -\frac{(\partial c_{ij}/\partial \Delta)\delta \Delta + (\partial c_{ij}/\partial R)\delta R}{(2\pi c_{ij}^2 \text{NSR}_{\text{GAM}})} < 0.
\]

(15)
The reduced NSR_{GAM} at larger \( R \) and smaller \( \Delta \) accounts for better correlation with \( r_{ij} \), which manifests itself as the enhanced \( |p_{ij}| \) for \( n \gg 1 \) under the same condition (Figs. 2E and 3G).

Meanwhile, HLM predicts that the inter-locus distance (\( r_{ij} \)) measured by FISH has a mean and variance of \( r_{ij} = 2/\sqrt{\pi} \gamma_{ij} \) and \( \sigma_{r_{ij}}^2 = (3\pi - 8)/(2\pi \gamma_{ij}) \), respectively. Thus, NSR of \( r_{ij} \) from FISH is

\[
    \text{NSR}_{\text{FISH}} = \frac{\sigma_{r_{ij}}}{r_{ij}} = \sqrt{\frac{3\pi - 8}{8n}}.
\]

(16)

Comparison of Eq. 16 with Eq. 14 (\( \text{NSR}_{\text{FISH}} < \text{NSR}_{\text{GAM}} \)) suggests that for a given sample size, the distance measurement using FISH is more precise and reliable than GAM (or Hi-C) when the co-segregation probability \( c_{ij} \) (or contact probability, \( p_{ij} \)) is less than \( 8/3\pi \approx 0.85 \).

To recapitulate, our theory predicts that unlike the contact probability, the co-segregation probability change non-monotonically with the spatial distance between two loci, which accounts for the discrepancy between GAM and FISH data on chromosome scales (Fig. 1). We demonstrate by using Gaussian phantom chain (Fig. 2), HLM of chromosome (Fig. 3), and in-silico analysis of experimental genome structures (Fig. 4) that the co-segregation probability depends on the slice range of nuclei, the thickness and the number of nuclear sections. The correlation between the co-segregation probability from GAM and the spatial distance between two genomic segments is not perfect, but it is only moderate both at chromosome-wide (Fig. 1) and short genomic range (Fig. 5). As a result, GAM data is not always straightforward to interpret. Thus, the findings made with GAM analysis, such as inter-chromosome contact, higher-order interaction, and associations with lamina and nucleolus, are usually supplemented with other measurements (9, 14). Our theoretical analysis of GAM based on HLM shows that if (co)-segregation probabilities from NPs with slice thickness of \( \sim 0.5 \) \( \mu \)m are only available data, then as long as the data are collected from a sufficient number (\( n \geq 10^3 \)) of NPs NPMI (\( s_{ij} \)), albeit not perfect, is slightly better suited than NLD (\( d_{ij} \)) for faithful interpretation of the 3D chromatin organization.
AUTHOR CONTRIBUTIONS
L.L. and C.H. designed the research. L.L. carried out all simulations. L.L., X.C., B.Z., and C.H. analyzed the data and wrote the article.

DECLARATION OF INTEREST
The authors declare no competing interests.

ACKNOWLEDGMENTS
L.L. thanks Xue Sun for helpful discussions. This work was financially supported by Zhejiang Provincial Natural Science Foundation, National Natural Science Foundation of China, Zhejiang Sci-Tech University (LQ22B040001, 12104404, 20062226-Y to L.L.) and Korea Institute for Advanced Study (CG035003 to C.H.). We thank the Center for Advanced Computation in KIAS for providing computing resources.

REFERENCES
1. Szabo, Q., D. Jost, J.-M. Chang, D. I. Cattoni, G. L. Papadopoulos, B. Bonev, T. Sexton, J. Gurgo, C. Jacquier, M. Nollmann, F. Bantignies, and G. Cavalli, 2018. TADs are 3D structural units of higher-order chromosome organization in Drosophila. Sci. Adv. 4:eaar8082.

2. Su, Q. P., Z. W. Zhao, L. Meng, M. Ding, W. Zhang, Y. Li, M. Liu, R. Li, Y.-Q. Gao, X. S. Xie, and Y. Sun, 2020. Superresolution imaging reveals spatiotemporal propagation of human replication foci mediated by CTCF-organized chromatin structures. Proc. Natl. Acad. Sci. USA 117:15036–15046.

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

4. Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker, 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–293.

5. Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren, 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485:376–380.

6. Rao, S. S. P., M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden, 2014. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell 159:1665–1680.

7. Krietenstein, N., S. Abraham, S. V. Venev, N. Abdennur, J. Gibcus, T.-H. S. Hsieh, K. M. Parsi, L. Yang, R. Maehr, L. A. Mirny, J. Dekker, and O. J. Rando, 2020. Ultrastructural Details of Mammalian Chromosome Architecture. Molecular Cell 78:554–565.e7.

8. Beagrie, R. A., C. J. Thieme, C. Annunziatella, C. Baugher, Y. Zhang, M. Schueler, A. Kukalev, R. Kemper, A. M. Chiariello, S. Bianco, et al., 2021. Multiplex-GAM: genome-wide identification of chromatin contacts yields insights not captured by Hi-C. bioRxiv 2020–07.

9. Winick-Ng, W., A. Kukalev, I. Harabula, L. Zea-Redondo, D. Szabó, M. Meijer, L. Serebreni, Y. Zhang, S. Bianco, A. M. Chiariello, I. Irastorza-Azcarate, C. J. Thieme, T. M. Sparks, S. Carvalho, L. Fiorillo, F. Musella, E. Irani, E. T. Triglia, A. A. Kołodziejczyk, A. Abentung, G. Apostolova, E. J. Paul, V. Franke, R. Kemper, A. Akalin, S. A. Teichmann, G. Dechant, M. A. Ungless, M. Nicodemi, L. Welch, G. Castelo-Branco, and A. Pombo, 2021. Cell-type specialization is encoded by specific chromatin topologies. Nature 599:684–691.
Theoretical analysis of genome architecture mapping

10. Ou, H. D., S. Phan, T. J. Deerinck, A. Thor, M. H. Ellisman, and C. C. O’Shea, 2017. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* 357:eaa0025.

11. Dekker, J., A. S. Belmont, M. Guttman, V. O. Leshyk, J. T. Lis, S. Lomvardas, L. A. Mirny, C. C. O’Shea, P. J. Park, B. Ren, J. C. R. Politz, J. Shendure, S. Zhong, and the 4D Nucleome Network, 2017. The 4D nucleome project. *Nature* 549:219–226.

12. Kempfer, R., and A. Pombo, 2020. Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* 21:207–226.

13. Jerković, I., and G. Cavalli, 2021. Understanding 3D genome organization by multidisciplinary methods. *Nat. Rev. Mol. Cell Biol.* 22:511–528.

14. Beagrie, R. A., M. Scialdone, Antonioand Schueler, D. C. A. Kraemer, M. Chotalia, S. Q. Xie, M. Barbieri, I. de Santiago, L.-M. Lavitas, M. R. Branco, J. Fraser, J. Dostie, L. Game, N. Dillon, P. A. W. Edwards, M. Nicodemi, and A. Pombo, 2017. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* 543:519–524.

15. Finn, E. H., and T. Misteli, 2017. Genome Architecture from a Different Angle. *Dev. Cell* 41:3–4.

16. Olivares-Chauvet, P., Z. Mukamel, A. Lifshitz, O. Schwartzman, N. O. Elkayam, Y. Lubling, G. Deikus, R. P. Sebra, and A. Tanay, 2016. Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature* 540:296–300.

17. Darrow, E. M., M. H. Huntley, O. Dudchenko, E. K. Stamenova, N. C. Durand, Z. Sun, S.-C. Huang, A. L. Sanborn, I. Machol, M. Shamim, A. P. Seberg, E. S. Lander, B. P. Chadwick, and E. L. Aiden, 2016. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc. Natl. Acad. Sci. U. S. A.* 113:E4504–E4512.

18. Oudelaar, A. M., J. O. J. Davies, L. L. P. Hanssen, J. M. Telenius, R. Schwessinger, Y. Liu, J. M. Brown, D. J. Downes, A. M. Chiariello, S. Bianco, M. Nicodemi, V. J. Buckle, J. Dekker, D. R. Higgs, and J. R. Hughes, 2018. Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nat. Genet.* 50:1744–1751.

19. Allahyar, A., C. Vermeulen, B. A. M. Bouwman, P. H. L. Krijger, M. J. A. M. Verstegen, G. Geeven, M. van Kranenburg, M. Pieterse, R. Straver, J. H. I. Haarhuis, K. Jalink, H. Teunissen, I. J. Renkens, W. P. Kloosterman, B. D. Rowland, E. de Wit, J. de Ridder, and W. de Laat, 2018. Enhancer hubs and loop collisions identified from single-allele topologies. *Nat. Genet.* 50:1151–1160.

20. Quinodoz, S. A., N. Ollikainen, B. Tabak, A. Palla, J. M. Schmidt, E. Detmar, M. M. Lai, A. A. Shishkin, P. Bhat, Y. Takei, V. Trinh, E. Aznauryan, P. Russell, C. Cheng, M. Jovanovic, A. Chow, L. Cai, P. McDonel, M. Garber, and M. Guttman, 2018. Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* 174:744 – 757.e24.

21. Liu, L., B. Zhang, and C. Hyeon, 2021. Extracting multi-way chromatin contacts from Hi-C data. *PLoS Comput. Biol.* 17:e1009669.

22. van Steensel, B., and A. S. Belmont, 2017. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* 169:780–791.

23. Chen, Y., Y. Zhang, Y. Wang, L. Zhang, E. K. Brinkman, S. A. Adam, R. Goldman, B. van Steensel, J. Ma, and A. S. Belmont, 2018. Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J. Cell Biol.* 217:4025–4048.

24. Takei, Y., J. Yun, S. Zheng, N. Ollikainen, N. Pierson, J. White, S. Shah, J. Thomassie, S. Suo, C.-H. L. Eng, M. Guttman, G.-C. Yuan, and L. Cai, 2021. Integrated spatial genomics reveals global architecture of single nuclei. *Nature* 590:344–350.
25. Takei, Y., S. Zheng, J. Yun, S. Shah, N. Pierson, J. White, S. Schindler, C. H. Tischbirek, G.-C. Yuan, and L. Cai, 2021. Single-cell nuclear architecture across cell types in the mouse brain. *Science* 374:586–594.

26. Kamat, K., Y. Qi, Y. Wang, J. Ma, and B. Zhang, 2021. Genome Compartmentalization with Nuclear Landmarks: Random yet Precise. *bioRxiv*.

27. Fiorillo, L., F. Musella, M. Conte, R. Kempfer, A. M. Chiariello, S. Bianco, A. Kukalev, I. Irastorza-Azcarate, A. Esposito, A. Abraham, A. Prisco, A. Pombo, and M. Nicodemi, 2021. Comparison of the Hi-C, GAM and SPRITE methods using polymer models of chromatin. *Nat. Methods* 18:482–490.

28. Bintu, B., L. J. Mateo, J.-H. Su, N. A. Sinnott-Armstrong, M. Parker, S. Kinrot, K. Yamaya, A. N. Boettiger, and X. Zhuang, 2018. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* 362:eaau1783.

29. Fiorillo, L., F. Musella, M. Conte, R. Kempfer, A. M. Chiariello, S. Bianco, A. Kukalev, I. Irastorza-Azcarate, A. Esposito, A. Abraham, A. Prisco, A. Pombo, and M. Nicodemi, 2021. Comparison of the Hi-C, GAM and SPRITE methods using polymer models of chromatin. *Nat. Methods* 18:482–490.

30. Liu, M., Y. Lu, B. Yang, Y. Chen, J. S. D. Radda, M. Hu, S. G. Katz, and S. Wang, 2020. Multiplexed imaging of nucleome architectures in single cells of mammalian tissue. *Nat. Comm.* 11:2907.

31. Takei, Y., S. Zheng, J. Yun, S. Shah, N. Pierson, J. White, S. Schindler, C. H. Tischbirek, G.-C. Yuan, and L. Cai, 2021. Single-cell nuclear architecture across cell types in the mouse brain. *Science* 374:586–594.

32. Durand, N. C., M. S. Shamim, I. Machol, S. S. Rao, M. H. Huntley, E. S. Lander, and E. L. Aiden, 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Systems* 3:95–98.

33. Kuhn, R. M., D. Haussler, and W. J. Kent, 2013. The UCSC genome browser and associated tools. *Brief Bioinform* 14:144–161.

34. Peric-Hupkes, D., W. Meuleman, L. Pagie, S. W. Bruggeman, I. Solovei, W. Brugman, S. Gräf, P. Flicek, R. M. Kerkhoven, M. van Lohuizen, M. Reinders, L. Wessels, and B. van Steensel, 2010. Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. *Mol. Cell* 38:603 – 613.

35. Liu, T., and Z. Wang, 2019. normGAM: an R package to remove systematic biases in genome architecture mapping data. *BMC Genomics* 20:1006.

36. Lewontin, R. C., 1964. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics* 49:49–67.

37. Slatkin, M., 2008. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genetics* 9:477–485.

38. Razak, F. A., 2014. The derivation of mutual information and covariance function using centered random variables. *In AIP Conference Proceedings*. American Institute of Physics, volume 1635, 883–889.

39. Bohn, M., D. W. Heermann, and R. van Driel, 2007. Random loop model for long polymers. *Phys. Rev. E* 76:051805.

40. Le Treut, G., F. Képès, and H. Orland, 2018. A polymer model for the quantitative reconstruction of chromosome architecture from HiC and GAM data. *Biophys. J.* 115:2286–2294.
41. Liu, L., M. H. Kim, and C. Hyeon, 2019. Heterogeneous loop model to infer 3D chromosome structures from Hi-C. *Biophys. J.* 117:613–625.

42. Liu, L., and C. Hyeon, 2020. Revisiting the organization of Polycomb-repressed domains: 3D chromatin models from Hi-C compared with super-resolution imaging. *Nucleic Acids Res.* 48:11486–11494.

43. Shinkai, S., M. Nakagawa, T. Sugawara, Y. Togashi, H. Ochiai, R. Nakato, Y. Taniguchi, and S. Onami, 2020. PHi-C: deciphering Hi-C data into polymer dynamics. *NAR Genom. Bioinf.* 2:lqaa020.

44. Shinkai, S., T. Sugawara, H. Miura, I. Hiratani, and S. Onami, 2020. Microrheology for Hi-C Data Reveals the Spectrum of the Dynamic 3D Genome Organization. *Biophys. J.* 118:2220–2228.

45. Shi, G., and D. Thirumalai, 2021. From Hi-C Contact Map to Three-dimensional Organization of Interphase Human Chromosomes. *Phys. Rev. X* 11:011051.

46. Shi, G., and D. Thirumalai, 2019. Conformational heterogeneity in human interphase chromosome organization reconciles the FISH and Hi-C paradox. *Nature Commun.* 10:1–10.

47. Jia, B. B., A. Jussila, C. Kern, Q. Zhu, and B. Ren, 2022. A spatial genome aligner for multiplexed DNA-FISH. *bioRxiv*.

48. Bianco, S., D. G. Lupiáñez, A. M. Chiariello, C. Annunziatella, K. Kraft, R. Schöpflin, L. Wittler, G. Andrey, M. Vingron, A. Pombo, S. Mundlos, and M. Nicodemi, 2018. Polymer physics predicts the effects of structural variants on chromatin architecture. *Nature Genet.* 50:662–667.

49. Yang, T., F. Zhang, G. G. Yardimci, F. Song, R. C. Hardison, W. S. Noble, F. Yue, and Q. Li, 2017. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* 27:1939–1949.

50. Liu, L., and C. Hyeon, 2016. Contact statistics highlight distinct organizing principles of proteins and RNA. *Biophys. J.* 110:2320–2327.
SUPPLEMENTAL INFORMATION

The expressions of mean inter-locus distance and the contact probability between chromatin segment pairs of Heterogeneous Loop Model (HLM) are derived in Sec. 1 and 2, respectively (39–43), followed by the detailed derivation of the co-segregation probability in Sec. 3. In the last section, we describe numerical generation of a 3D structural ensemble of the model.

1 MEAN INTER-LOCUS DISTANCE

The chromatin fiber in a genomic region of interest was modeled as a linear polymer chain composed of \( N \) coarse-grained monomers each representing a chromatin segment with a prescribed genomic length. The interaction energy of the fiber has a form of

\[
U(r) = \sum_{i=0}^{N-1} \sum_{j=i+1}^{N-1} \frac{k_{ij}}{2} |\vec{r}_i - \vec{r}_j|^2 = \sum_{\alpha} \frac{1}{2} r_{\alpha}^T L r_{\alpha} \tag{S1}
\]

where \( r_{\alpha} = (r_{0,\alpha}, r_{1,\alpha}, r_{2,\alpha}, \ldots, r_{N-1,\alpha})^T \) and the subscript \( \alpha \) represents \( x, y \) and \( z \). The Laplacian matrix \( L \) is defined as \( L = D - K \), where \( K \) is a stiffness matrix of elements \( k_{ij} \) and \( D \) is a diagonal matrix of elements \( D_{ii} = \sum_j k_{ij} \).

Upon translating the center of mass (COM) of the polymer to the origin of the coordinate system, \( L \) can be transformed to a new matrix,

\[
\Sigma = Q \text{ diag} \left( \lambda_1, \lambda_2, \ldots, \lambda_{N-1} \right) Q^T, \tag{S2}
\]

where the \( i \)-th column of the orthogonal matrix \( Q \) and \( \lambda_i \) are the corresponding \( i \)-th eigenvector and eigenvalue of \( L \), respectively. By using the notation of

\[
\gamma_{ij} = \frac{1}{2} (\sigma_{ii} + \sigma_{jj} - 2\sigma_{ij})^{-1} \tag{S3}
\]

where \( \sigma_{ij} \) is the \((i, j)\)-th element of \( \Sigma \), the mean distance and the mean squared distance between the \( i \)- and \( j \)-th sites in 3D space, that can be measured by FISH imaging, are obtained as

\[
\bar{r}_{ij} = \frac{2}{\sqrt{\gamma_{ij}}} \left(1 + \frac{3}{2\gamma_{ij}}\right)^{-1/2}
\]

\[
\bar{r}_{ij}^2 = \frac{3}{2\gamma_{ij}} \tag{S4}
\]

2 CONTACT PROBABILITY

Based on the above chromatin polymer model, we have discussed generic \( n \)-body contact probability in Ref. (21). The pairwise \((n = 2)\) contact probability is given by

\[
p_{ij}^G(r_c) = \left(1 + \frac{3}{2\gamma_{ij}r_c^2}\right)^{-3/2} - 2\Delta \sqrt{\gamma_{ij}} e^{-r_c^2}
\]

\[
p_{ij}^S(r_c) = \text{erf} \left( \frac{\gamma_{ij}^{1/2} r_c}{r_c} \right) - 2\Delta \sqrt{\frac{\gamma_{ij}}{\pi}} e^{-r_c^2} \tag{S5}
\]

with the special function \( \text{erf}(x) = 2\pi^{-1/2} \int_0^x dt e^{-t^2} \). The superscript denotes the Gaussian or rectangular profile of \( F_{\text{Hi-C}}(r) \), which describes the distance-dependent efficiency of the cross-linking agent in Hi-C experiments. More specifically, we have assumed

\[
F_{\text{Hi-C}}^G(r) = e^{-\frac{r_c^2}{2\Delta^2}}
\]

\[
F_{\text{Hi-C}}^S(r) = \begin{cases} 1, & r < r_c \\ 0, & \text{otherwise} \end{cases} \tag{S6}
\]
to obtain Eqs. S5 where \( \Delta \) denotes a characteristic capture radius.

3 CO-SESEGREGATION PROBABILITY

To derive the co-segregation probability of chromatin segment pairs as the GAM experiment (14), we begin with the probability that the \( i \)-th segment is located at \( z_i = \alpha \) from the COM of the chain,

\[
p_{z_i=\alpha} = \langle \delta(z_i - \alpha) \rangle = \int \mathcal{D}z \delta(z_i - \alpha) e^{-\frac{1}{2}z^2/L_z} \int \mathcal{D}z e^{-\frac{1}{2}z^2/L_z} \propto \frac{1}{\sqrt{2\pi\sigma_{ii}}} e^{-\frac{1}{2}z_i^2/\sigma_{ii}}.
\]

Similarly, the probability that the \( i \)-th and \( j \)-th segments are simultaneously found at \( z_i = \alpha \) and \( z_j = \beta \), respectively, is given by

\[
p_{z_i=\alpha, z_j=\beta} = \langle \delta(z_i - \alpha)\delta(z_j - \beta) \rangle \propto \frac{1}{2\pi\sqrt{M_{ij}}} e^{-\frac{1}{2}z_i^2/(\sigma_{ii} + \Delta^2)} e^{-\frac{1}{2}z_j^2/(\sigma_{jj} + \Delta^2)}.
\]

with \( M_{ij} = \sigma_{ii}\sigma_{jj} - \sigma_{ij}^2 \).

Gaussian sectioning. Next, we assume that the horizontal slice \( S^\Delta_z \) at \( h = z \) with thickness \( \Delta \) is sectioned with a Gaussian probability

\[
F_G^{\text{GAM}}(\alpha) = e^{-\frac{1}{2}(\alpha - z)^2/\Delta^2}
\]

where the superscript again notates the Gaussian profile of \( F_{\text{GAM}} \). For simplicity, we have omitted the superscript “G” in the main text.

Combining Eqs. S7-S9, it is straightforward to formulate the probability of the \( i \)-th segment being sectioned by \( S^\Delta_z \) as

\[
c_i^{0,G}(z, \Delta) = \int_{-\infty}^{+\infty} d\alpha F_G^{\text{GAM}}(\alpha) \propto p_{z_i=\alpha}
= \frac{\Delta}{\sqrt{\sigma_{ii} + \Delta^2}} e^{-\frac{1}{2}z_i^2/(\sigma_{ii} + \Delta^2)},
\]

and the co-segregation probability as

\[
c_{ij}^{0,G}(z, \Delta) = \int d\alpha \int d\beta F_G^{\text{GAM}}(\alpha) F_G^{\text{GAM}}(\beta) \propto p_{z_i=\alpha, z_j=\beta}
= \frac{\Delta^2}{\sqrt{\Omega_{ij}}} e^{-\frac{1}{2}z_i^2/(\sigma_{ii} + \Delta^2)} e^{-\frac{1}{2}z_j^2/(\sigma_{jj} + \Delta^2)}
\]

where

\[
\Psi_{ij} = \sigma_{ii} + \sigma_{jj} - 2\sigma_{ij} + 2\Delta^2
\]

\[
\Omega_{ij} = (\sigma_{ii} + \Delta^2)(\sigma_{jj} + \Delta^2) - \sigma_{ij}^2.
\]
• **NPs collected from uniform slicing.** To mimic the procedure of GAM (14), we assume that the NPs are collected uniformly from \( z \in [-R, R] \) \((R \geq 0)\). Then the segregation probability is given by

\[
c_i^G(R, \Delta) = \frac{1}{2R} \int_{-R}^{R} dz \ c_i^{oG}(z, \Delta) = \sqrt{\frac{\pi}{2}} R \ \text{erf} \left( \frac{R}{\sqrt{2(\sigma_{ii} + \Delta^2)}} \right),
\]

(S13)

and the co-segregation probability has the expression of

\[
c_{ij}^G(R, \Delta) = \frac{1}{2R} \int_{-R}^{R} dz \ c_{ij}^{oG}(z, \Delta) = \sqrt{\frac{\pi}{2}} R \ \text{erf} \left( \frac{\Psi_{ij}}{2\Omega_{ij}} \right) R.
\]

(S14)

• **NPs collected from Gaussian slicing.** Alternatively, we assume the NPs are obtained from a Gaussian probability of \( p(z) \propto e^{-\frac{1}{2}z^2/R^2} \) (i.e., a higher chance to be sliced at the center of cell nucleus than the apexes). This gives rise to the segregation probability as

\[
c_i^{G*}(R, \Delta) = \frac{1}{\sqrt{2\pi R}} \int_{-\infty}^{\Delta} dz \ e^{\frac{1}{2}z^2/R^2} e^{-\frac{z^2}{2R^2}} = \frac{\Delta}{\sqrt{\sigma_{ii} + \Delta^2 + R^2}}.
\]

(S15)

and the co-segregation probability as

\[
c_{ij}^{G*}(R, \Delta) = \frac{1}{\sqrt{2\pi R}} \int_{-\infty}^{\Delta} dz \ e^{\frac{1}{2}z^2/R^2} e^{-\frac{z^2}{2R^2}} = \frac{\Delta^2}{\sqrt{\Psi_{ij}R^2 + \Omega_{ij}}}.
\]

(S16)

As shown in Fig. S2, \( c_i^{G*} \) and its correlation with \( \bar{r}_{ij} \) calculated using Eq. S16 are similar to those calculated with Eq. S14 (see Fig. 2D and E in the main text).

**Rectangular sectioning.** Sections with sharp edges in GAM can be modeled more realistically by employing a rectangular profile. The above results (Eqs. S10-S14) can be reformulated by replacing \( F_{GAM}^S \) in Eq. S9 with

\[
F_{GAM}^S(\alpha) = \begin{cases} 1, & |\alpha - z| < \Delta \\ 0, & \text{otherwise} \end{cases}.
\]

(S17)

The segregation probability of the \( i \)-th segment from a slice made at \( h = z \) has a form of

\[
c_{i}^{oS}(z, \Delta) = \int_{-\infty}^{\infty} d\alpha \ F_{GAM}^S(\alpha) \times p_{z_i = \alpha} = \frac{1}{2} \left( \text{erf} \left( \frac{z + \Delta}{\sqrt{2\sigma_{ii}}} \right) - \text{erf} \left( \frac{z - \Delta}{\sqrt{2\sigma_{ii}}} \right) \right)
\]

(S18)

When the slices are collected from the range \( |z| \leq R \), the segregation probability becomes

\[
c_i^S(R, \Delta) = \frac{1}{2R} \int_{-R}^{R} dz c_i^{oS}(z, \Delta) = \frac{B}{2R} \left( \frac{\Delta + R}{B} \text{erf} \left( \frac{\Delta + R}{\sqrt{2\sigma_{ii}}} \right) - \frac{\Delta - R}{B} \text{erf} \left( \frac{\Delta - R}{\sqrt{2\sigma_{ii}}} \right) - \sqrt{\frac{2\sigma_{ii}}{\pi}} \left( e^{\frac{i\Delta}{\sigma_{ii}}} - 1 \right) \right).
\]

(S19)
where \( B = e^{-\frac{(\Delta + R)^2}{2\sigma^2}} \). The corresponding co-segregation probability,

\[
c_{ij}^{oS}(z, \Delta) = \int \int d\alpha d\beta \int_{z-\Delta}^{z+\Delta} da \ e^{-\frac{a^2}{2\sigma^2}} \left( \text{erf} \left( \frac{\alpha - \sigma_{ij} z + \Delta}{\sqrt{2M_{ij} \sigma_{ij}}} \right) - \text{erf} \left( \frac{\alpha - \sigma_{ij} z - \Delta}{\sqrt{2M_{ij} \sigma_{ij}}} \right) \right),
\]

and \( c_{ij}^{S}(R, \Delta) = \frac{1}{2R} \int_{-R}^{+R} dz \ c_{ij}^{oS}(z, \Delta) \) shown in Fig. S3 are obtained with numerical integration. All the results using rectangular sectioning are similar to those calculated using the Gaussian sectioning and are presented in Fig. S3.

### 4 3D STRUCTURES OF POLYMER CHAIN

In HLM, the normal coordinate vector of a polymer chain \( X_\alpha = Q^T r_\alpha \) satisfies

\[
\langle X_{p,\alpha} X_{q,\beta} \rangle = \frac{k_BT}{\lambda_p} \delta_{pq} \delta_{\alpha\beta}.
\]  

(S21)

where \( Q \) and \( \lambda \) are defined in Eq. S2, \( k_B T \) is our energy unit, \( \alpha \) and \( \beta \) represent \( x, y \) and \( z \), and \( p, q = 1, 2, 3, \ldots, N-1 \). Based on this relation, a 3D conformation of the polymer chain can be generated in two steps (43). First, we obtain the normal coordinates by using

\[
X_{p,\alpha} = \begin{cases} 
0, & p = 0 \\
\xi_\alpha / \sqrt{\lambda_p}, & p > 0
\end{cases}
\]

(S22)

where the random variable \( \xi_\alpha \) obeys the normal distribution with \( \langle \xi_\alpha \rangle = 0 \) and \( \langle \xi_\alpha^2 \rangle = 1 \). Next, the normal coordinates are converted to the Cartesian coordinates of polymer segments by

\[
r_\alpha = Q X_\alpha.
\]

(S23)

Repeating this protocol \( n \) times yields a structural ensemble of the polymer chain containing \( n \) samples.
Figure S1: Comparing mouse ESC autosomes modeled at 1-Mb resolution with HLM against experimental data. (A) Percentile rank (PR) of the contact probability of chromosome 3 from HLM ($p_{ij}^{HLM}$, top left conner) versus that from Hi-C (5) ($p_{ij}^{Hi-C}$, bottom right conner). (B) PR of the mean inter-locus spatial distance from HLM ($\bar{r}_{ij}^{HLM}$, top left conner) versus that from FISH (24) ($\bar{r}_{ij}^{FISH}$, bottom right conner). (C) The Spearman’s rank correlation coefficient for all autosomes.

Figure S2: Co-segregation probability of a Gaussian polymer chain which are calculated by assuming the slices position following a Gaussian profile (see Eqs. S15 and S16). (A) and (B) are similar to Fig. 2D-E in the main text, respectively.
Figure S3: Co-segregation probability of a Gaussian polymer chain which are calculated by assuming the rectangular sectioning profile (see Eqs. S17-S20). (A)-(D) are similar to Fig. 2B-E in the main text, respectively.
Figure S4: In silico GAM results of chromosome 12. (A) Percentile ranks of $c_{ij}$ sectioned by slices with a width of $\Delta = 0.5 \mu m$ and a height of $z = 0$ (top-left) and $7 \mu m$ (bottom-right). The Spearman's correlation coefficient between the two is $\rho_s = 0.75$. (B) The correlations ($\rho_s$) between the mean physical distance and co-segregation probability, (C) NLD, and (D) NPMI as a function of the positional uncertainty ($R$) and width ($\Delta$) of slices for varying $n$. The symbol “×” denotes the values of ($R, \Delta$) that maximizes $|\rho_s|$. 

Figure S5: In silico GAM results of chromosome 18. Please refer to the caption of Fig. S4 for explanations. The Spearman’s correlation between the top-left and bottom-right panels in (A) is 0.69.
Figure S6: The Spearman’s correlation between the mean physical distance ($\bar{r}_{ij}$) and Hi-C contact probability ($p_{ij}$), GAM co-segregation probability ($c_{ij}$), and its two variants ($d_{ij}$ and $s_{ij}$) as a function of the genomic distance between the ($i$, $j$) sites in different chromosomes. The mean and standard deviation averaged over all the autosomes are shown in Fig. 5A in the main text.