Biophysical Journal
Article

Biophysical Society

Dissecting the cosegregation probability from genome architecture mapping

Lei Liu,[1,*] Xinmeng Cao,[1] Bokai Zhang,[2] and Changbong Hyeon[3,*]
[1]Key Laboratory of Optical Field Manipulation of Zhejiang Province, Department of Physics, Zhejiang Sci-Tech University, Hangzhou, China; [2]School of Physical Science and Technology, Southwest University, Chongqing, China; and [3]School of Computational Sciences, Korea Institute for Advanced Study, Seoul, Korea

ABSTRACT Genome architecture mapping (GAM) is a recently developed methodology that offers the cosegregation probability of two genomic segments from an ensemble of thinly sliced nuclear profiles, enabling us to probe and decipher three-dimensional chromatin organization. The cosegregation probability from GAM binned at 1 Mb, which thus probes the length scale associated with the genomic separation greater than 1 Mb, is, however, not identical to the contact probability obtained from Hi-C, and its correlation with interlocus distance measured with fluorescence in situ hybridization 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 cosegregation 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 three-dimensional genome structures from fluorescence in situ hybridization, suggest that to attain strong correlation with the interlocus distance, a properly normalized version of cosegregation 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 cosegregation probability captured from genome architecture mapping (GAM) analysis. Our polymer model, which offers analytical expressions for the cosegregation probability as well as for the contact probability and interlocus distance, enables quantitative comparison between the data from GAM, Hi-C, and fluorescence in situ hybridization. Although the plain cosegregation probabilities from GAM are not well correlated with interlocus distances measured from fluorescence in situ hybridization, properly normalized versions of the probability calculated from a large number of nuclear profiles can still reasonably represent the interlocus distance. Our study offers instructions of how to take full advantage of GAM analysis in deciphering three-dimensional 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. 1 *A*). Loci that are distant in space are expected to have smaller chance to be cosectioned in the same NP. It has been presumed that the cosectioning/cosegregation probability between two genomic sites (c_{ij}), i and j, is inversely related to their mean spatial distance ($\overline{r}_{ij}$) (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 cosegregation 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)

Submitted May 7, 2022, and accepted for publication September 16, 2022.
*Correspondence: leiliu@zstu.edu.cn or hyeoncb@kias.re.kr
Editor: Toshio Tsukiyama.
https://doi.org/10.1016/j.bpj.2022.09.018

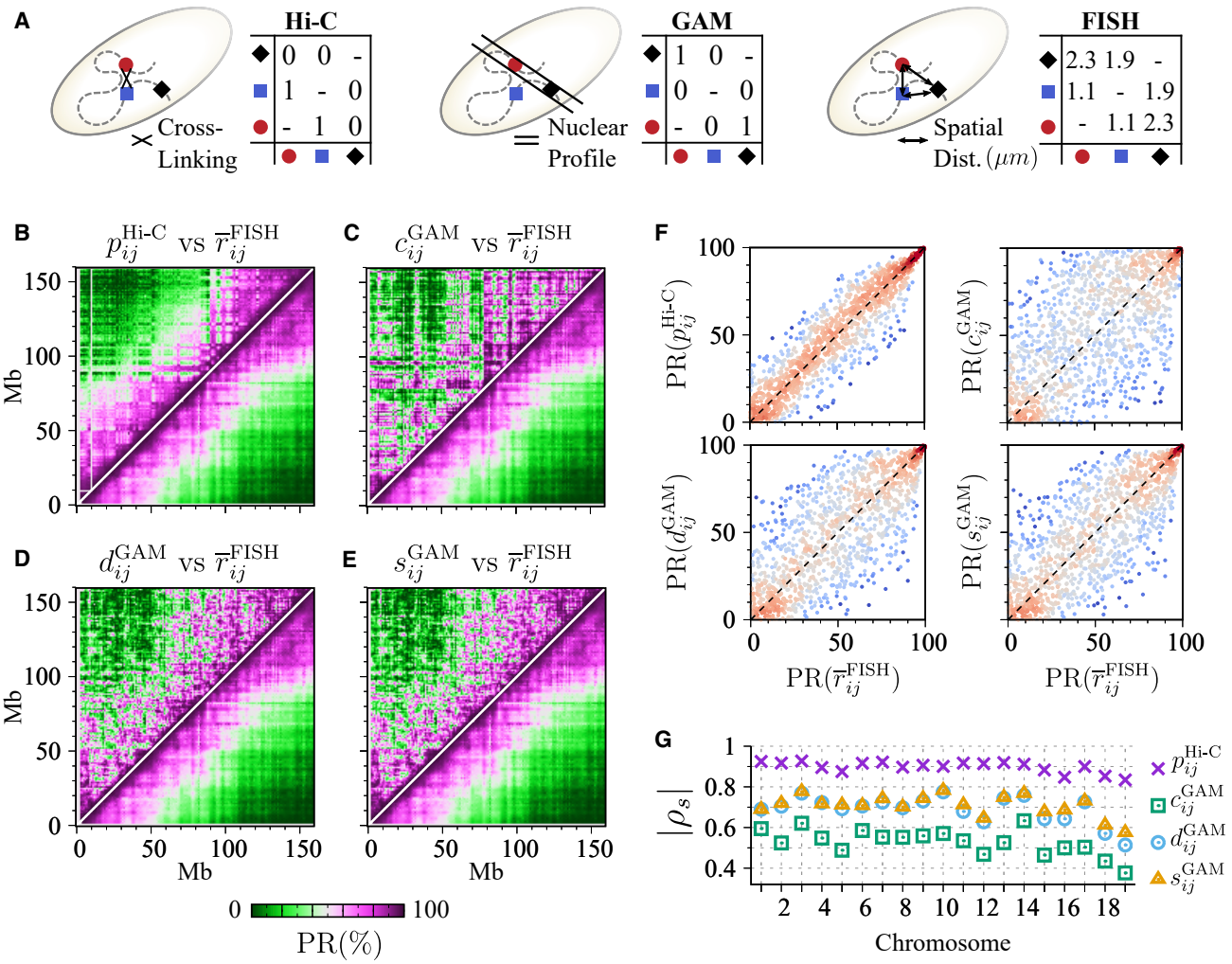

FIGURE 1 Comparison between GAM, Hi-C, and FISH. (*A*) Schematics illustrating the three methods. From an ensemble of cells, Hi-C measures the cross-linking frequencies between two genomic loci, GAM measures the cosegregation frequencies in a nuclear profile, and FISH measures the interlocus spatial distances. (*B–D*) GAM data are compared with FISH that probed the chromosomes of mouse ESCs using the 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 interlocus distance ($\overline{r}_{ij}^{\mathrm{FISH}}$) on the chromosome 3 binned at 1 Mb (*bottom right corner* of each panel) (24) versus (*B*) PR of Hi-C contact probability (5) ($p_{ij}^{\mathrm{Hi-C}}$), (*C*) PR of cosegregation probability from GAM (14) (c_{ij}^{GAM}), (*D*) PR of normalized linkage disequilibrium (NLD) (14) (d_{ij}^{GAM}), and (*E*) PR of normalized point-wise mutual information (NPMI) (27) (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 $p_{ij}^{\mathrm{Hi-C}}$, c_{ij}^{GAM}, d_{ij}^{GAM}, and s_{ij}^{GAM} against $\overline{r}_{ij}^{\mathrm{FISH}}$ for the 19 autosomes of mouse ESCs.

and lamin and nuclear body association (22–26), can be measured by leveraging cryo fluorescence in situ hybridization (FISH)-combined GAM.

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 cosegregation frequencies (or probabilities, c_{ij}) to identify topologically associated domains (TADs) that are interacting with each other. In reference to cryoFISH images, they showed that the interacting TADs have higher contact probabilities (p_{ij}) and smaller spatial distances ($\overline{r}_{ij}$) than noninteracting 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 $\overline{r}_{ij}$ 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 the inverse relationship between $\overline{r}_{ij}$ and c_{ij} is still remains as an open question. To be specific, when the cosegregation 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 $\overline{r}_{ij} < \overline{r}_{kl}$?

To address the above question, we calculate $\overline{r}_{ij}$ by using 3D imaging data of mouse embryonic stem cells (ESCs) (24) and compare them with c_{ij} obtained from GAM (24)

by means of their percentile rank (PR), i.e., PR($\overline{r}_{ij}^{\mathrm{FISH}}$) versus PR($c_{ij}^{\mathrm{GAM}}$), where the superscripts "FISH," "GAM," and "Hi-C" are added to specify the experimental data source for clarity. For the case of $\overline{r}_{ij}$, we consider that PR is higher when $\overline{r}_{ij}$ is smaller. As shown in Fig. 1 *B*, PR($\overline{r}_{ij}^{\mathrm{FISH}}$) depicted in a matrix form for 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 cosegregation probability (c_{ij}) and its variants—the normalized linkage disequilibrium (NLD; d_{ij}) (14) and the normalized pointwise mutual information (NPMI; s_{ij}) (27) (see materials and methods)—such similarity between the patterns is significantly weaker (Fig. 1 *D–E*). The scatter plot of PR(p_{ij}) against PR($\overline{r}_{ij}$) also displays a stronger correlation than those of PR(c_{ij}), PR(d_{ij}), and PR(s_{ij}) (Fig. 1 *F*). The Spearman's rank correlation of p_{ij} with $\overline{r}_{ij}$ is significantly greater than that of GAM-based measures (c_{ij}, d_{ij}, s_{ij}) for all autosomes of mouse ESCs (Fig. 1 *G*).

In light of the general agreement between Hi-C and FISH (3,24,25,28–30), the significantly weaker correlation of cosegregation frequencies from GAM with the mean interlocus 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 cosegregation 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 data set. Our study, which enables quantitative comparison between GAM, Hi-C, and 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 data sets

The genome-wide GAM data was downloaded from the GEO repository (14) (GEO: 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 cosegregation probability, c_i and c_{ij}, respectively.

We used the Hi-C data set of mouse ESCs measured by Dixon et al. (5). KR-normalized intrachromosome contact probability at 1 Mb resolution, p_{ij}, was calculated by using the Juicer toolbox (31).

The DNA seqFISH+ data set was fetched from the 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 (32). 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 intrachromosome loci pair (i,j) binned at 1 Mb, we then calculated their spatial distances for each allele in all cells, which yields the mean interlocus distance $\overline{r}_{ij}$.

The DamID of lamin B1 protein in mouse ESCs was downloaded from GEO: GSE17051 (33), where the base-2 logarithm of the fold enrichment of the interactions between chromatin loci and nuclear lamin (q_{LB1}) was available. A chromatin segment has a positive (negative) value of q_{LB1} if it has higher (lower) chance to associate with lamin than the genome-wide average level.

Two variants of cosegregation probability

Of several possible methods for normalizing GAM data (34), we consider the two most popular ones, NLD and NPMI.

(1) To account for the observation that different loci have different chances to be cryosectioned, the NLD, which was originally proposed in population genetics to calculate the nonrandom association of two alleles at different loci (35,36), has been employed for the analysis of GAM (14). The NLD is defined as

$$d_{ij} = D_{ij} \Big/ D_{ij}^{\max}, \tag{1}$$

where

$$D_{ij} = c_{ij} - c_i c_j \tag{2}$$

compares the cosegregation probability of i and j-th loci (c_{ij}) with the probability of statistical independence, normalized by the theoretical maximum

$$D_{ij}^{\max} = \begin{cases} \min\left(c_i c_j,\ (1-c_i)(1-c_j)\right), & D_{ij} < 0 \\ \min\left(c_i(1-c_j),\ c_j(1-c_i)\right), & D_{ij} > 0. \end{cases} \tag{3}$$

(2) Alternatively, the NPMI has been used for the GAM analysis as well (27):

$$s_{ij} = -\frac{\log\left(\frac{c_{ij}}{c_i c_j}\right)}{\log\left(c_{ij}\right)}, \tag{4}$$

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 (37).

RESULTS AND DISCUSSION

Theoretical analyses

Heterogeneous loop model

We use the heterogeneous loop model (HLM) (21,38–44) to derive analytical expressions for the contact and cosegregation probabilities as well as mean interlocus 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,

$$\begin{aligned} U_{\mathcal{K}}(r) &= \sum_{i=0}^{N-1}\sum_{j=i+1}^{N-1}\frac{k_{ij}}{2}\left|\vec{r}_i - \vec{r}_j\right|^2, \\ &= \frac{1}{2}r^T\cdot\mathcal{L}\cdot r \end{aligned} \tag{5}$$

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 $\mathcal{L}$ is defined as $\mathcal{L} = \mathcal{D} - \mathcal{K}$, where $\mathcal{K}$ is a stiffness matrix of element k_{ij} and $\mathcal{D}$ represents a diagonal matrix with $\mathcal{D}_{ii} = \sum_j k_{ij}$. The probability of the chromatin to adopt a particular structure is then written as

$$P(r) \propto e^{-U_{\mathcal{K}}(r)/k_{\mathrm{B}}T}, \tag{6}$$

where $k_{\mathrm{B}}T$ is the energy unit of the model with k_B the Boltzmann constant and T the absolute temperature. The model parameters ($\mathcal{K}$) 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 a chromosome can be directly derived based on Eqs. 5 and 6 along with Hi-C data. HLM and its variants have been exploited to study experimental measurements (21–44). The contact probability calculated from the HLM is in excellent agreement with the measurements (21,40,41). Specifically, despite the cell-to-cell variability of 3D genome over population (44,45), the intrachromosomal interlocus spatial distance distributions predicted by HLM can still be validated against those measured from DNA seqFISH+ imaging (see Fig. 1 in (21)).

Although it does not affect the discussion in this work, the HLM has limitations when it gets to a resolution greater than $\mathcal{O}(10^2)$ basepairs, 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 $\mathcal{K}$ into a covariance matrix Σ whose matrix element is denoted as $\sigma_{ij} \equiv (\Sigma)_{ij}$, one finds the mean spatial distance between the i- and j-th segments averaged over all possible chromosome configurations as (detailed derivations are given in the supporting material)

$$\overline{r}_{ij} = \frac{2}{\sqrt{\pi\gamma_{ij}}}, \tag{7}$$

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

$$p_{ij}(r_c) = \left(1 + \frac{3}{2\gamma_{ij}r_c^2}\right)^{-3/2}, \tag{8}$$

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 ($\overline{r}_{ij}$ and p_{ij}) are related as $p_{ij} = (1 + 3\pi\overline{r}_{ij}^2/8r_c^2)^{-3/2}$, and hence they are in perfect correlation, giving rise to the Spearman's rank correlation coefficient $\rho_s(p_{ij}, \overline{r}_{ij}) = -1$. The consistency between Hi-C (p_{ij}) and FISH ($\overline{r}_{ij}$) shown in Fig. 1 *B* and *G* has been demonstrated by a number of experimental studies (28–30,46).

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 interlocus distance based on Eqs. 8 and 7, 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 $\overline{r}_{ij}$ from FISH can be well predicted with the correlation coefficient $\rho_s \simeq 0.9$, 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.

Cosegregation probability

The probability of the i-th genomic segment being in a horizontal slice (S_z^{Δ}) with a thickness Δ sectioned at a height z relative to the center of mass (COM) of the chain is derived as follows (see supporting material)

$$c_i^o(z, \Delta) = \frac{\Delta}{\sqrt{\sigma_{ii} + \Delta^2}} e^{-\frac{1}{2}z^2 \Big/ \left(\sigma_{ii} + \Delta^2\right)}, \tag{9}$$

and the cosegregation probability is formulated as

$$c_{ij}^o(z, \Delta) = \frac{\Delta^2}{\sqrt{\Omega_{ij}}} e^{-\frac{z^2}{2\left(\Omega_{ij}/\Psi_{ij}\right)}}, \tag{10}$$

where

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

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

Eqs. 7–10 are the general theoretical expressions derived in the framework of HLM with σ_{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; supporting material 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_z^Δ. The numerical results of (co)segregation probability obtained from explicit 3D structures (Fig. 2 *A*) are in perfect agreement with our analytic expressions in Eqs. 9 and 10 (Fig. 2 *B* and *C*).

Eqs. 9 and 10 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^o and c_{ij}^o increase with the thickness of the slice, but their variation with z is nontrivial. The radius of the polymer ensemble depicted in Fig. 2 *A* is $R \approx 5$. Thus, in an 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. 2 *B*). As shown in Fig. 2 *A*, 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 nonuniform. The ranking order of c_{ij}^o demonstrates qualitative difference with z as well (Fig. 2 *C*).

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

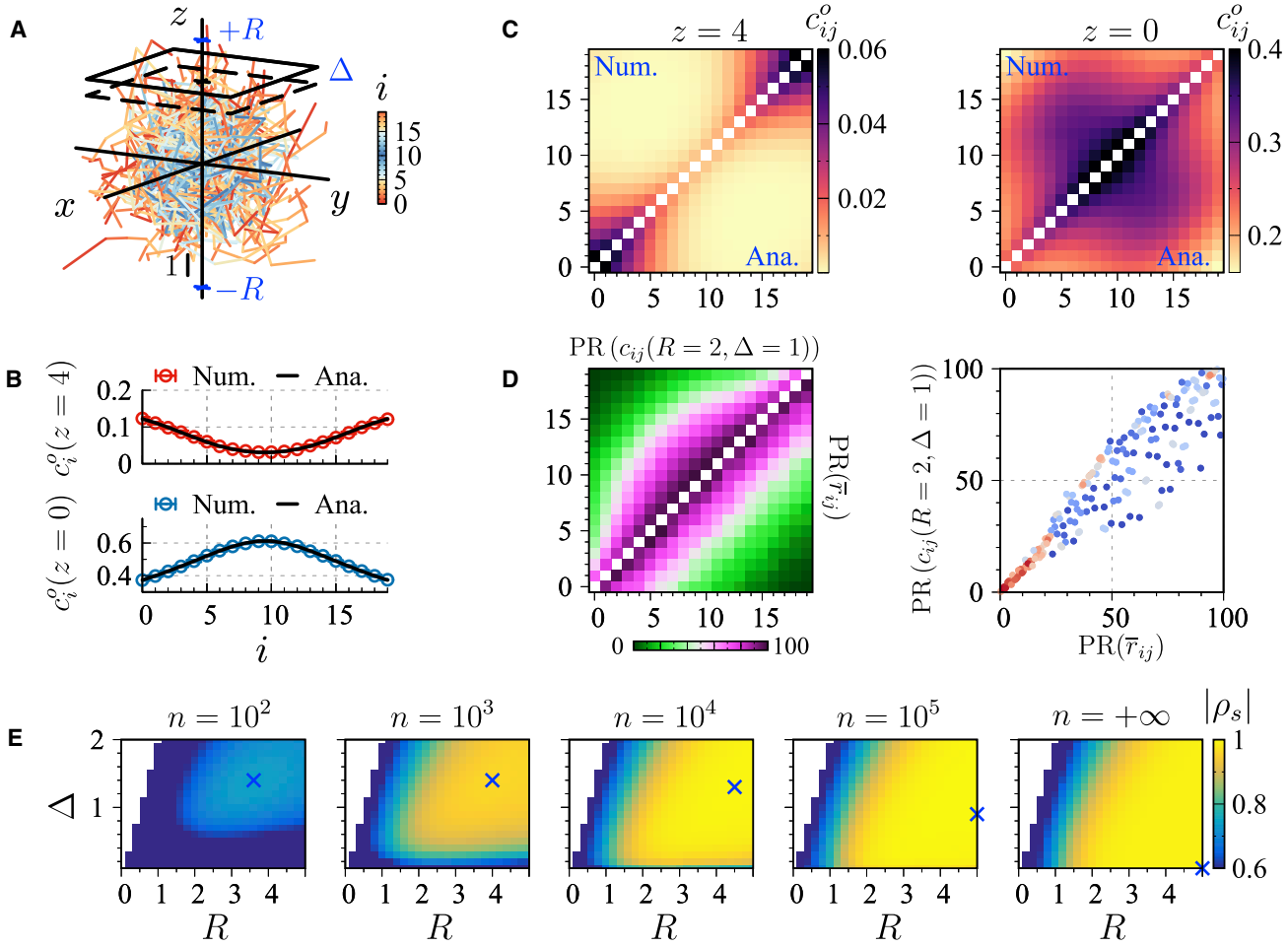

FIGURE 2 Cosegregation of monomers in a Gaussian phantom chain. (*A*) Depicted are 100 chains randomly selected from an ensemble, each composed of 20 monomers. GAM data would be collected from an ensemble of slices with a thickness Δ and at a height z ($|z| < R$). The terminal and the central parts of the chain are colored in red and blue, respectively. (*B*) Segregation and (*C*) cosegregation probabilities, c_i^o and c_{ij}^o, calculated at different values of z with $\Delta = 1$. (*D*) (*Left panel*) PR of the cosegregation probability (*top left*) compared with the PR of the mean interlocus distance (*bottom right*). (*Right panel*) Scatter plot of $(\mathrm{PR}(c_{ij}),\ \mathrm{PR}(\bar{r}_{ij}))$. The density of the data point in each pixel is color coded from blue to red. (*E*) Spearman's rank correlation coefficient (ρ_s) between c_{ij} and $\bar{r}_{ij}$ as a function of R and Δ, where the cosegregation probabilities were calculated by using n slices. The $|\rho_s|$ with $n \to \infty$ is obtained by using Eqs. 7 and 12. The condition of (R^*, Δ^*) that gives rise to the strongest correlation $|\rho_s|$ for different n is marked with the symbol $\times$.

$$c_i(R,\,\Delta) \;=\; \frac{1}{2R}\int_{-R}^{R} c_i^o(z,\,\Delta)\mathrm{d}z \\ =\; \sqrt{\frac{\pi}{2}}\;\frac{\Delta}{R}\mathrm{erf}\left(\frac{R}{\sqrt{2(\sigma_{ii}+\Delta^2)}}\right), \tag{11}$$

with $\mathrm{erf}(x)\equiv\frac{2}{\sqrt{\pi}}\int_0^x dte^{-t^2}$. Similarly, the mean cosegregation probability is obtained as

$$c_{ij}(R,\,\Delta) \;=\; \sqrt{\frac{\pi}{2}}\;\frac{\Delta^2}{R}\Psi_{ij}^{-1/2}\mathrm{erf}\left(\sqrt{\frac{\Psi_{ij}}{2\Omega_{ij}}}R\right). \tag{12}$$

Eq. 12 suggests that c_{ij} is decided by the range of the nuclei being sliced, namely $[-R,\,+R]$. Unlike Ψ_{ij}, which is a monotonic function of the mean distance, $\Psi_{ij} = (2\gamma_{ij})^{-1}+2\Delta^2 = \pi\bar{r}_{ij}^2/8+2\Delta^2$, Ω_{ij} changes nonmonotonically 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. As illustrated in Fig. 2 *D*, $\mathrm{PR}(c_{ij})$ clearly differs from $\mathrm{PR}(\bar{r}_{ij})$.

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 $|\rho_s|$ as a function of R and Δ in Fig. 2 *E*. As n increases from $n = 10^3$ to $n\to\infty$, the value (R^*,Δ^*) that maximizes $|\rho_s|$ shifts toward large R and small Δ.

The segregation (Eq. 11) and cosegregation probabilities (Eq. 12) were derived by assuming that the sectioning was made with Gaussian probability at height z and that the NPs were collected uniformly over the nuclear volume. 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 supporting material).

Cosegregation probability of HLM-based chromosome model

The behaviors of cosegregation 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. 3 *A*), produces an ensemble of structures characterized with three distinct domains (Fig. 3 *B*). 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_i^o and c_{ij}^o on z; see Fig. 3 *D* and *E*), the nonuniform radial distribution of chromosome (Fig. 3 *C–E*), and the improved correlation between c_{ij} and $\bar{r}_{ij}$ at greater n (Fig. 3 *F* and *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. 5), 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 data set of the DNA seqFISH+ experiment, translated the COM of each structure to the origin, and rotated the structure to a random orientation. Horizontal slices with a thickness Δ were obtained from the structures by randomly sampling the position along the z axis in the range of $z\in[-R,\,+R]$ (Fig. 4 *A*). 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. 4 *A*, the chromatin segments inside the cell nucleus have a nonuniform radial distribution. Second, we confirm that cosegregation probabilities (c_{ij}^o) calculated at the different positions of the nuclear slice ($z = 0$ and 7 μm) are qualitatively different from each other (Fig. 4 *B*). More pronounced dependence of c_{ij}^o on z is found for other chromosomes (Figs. S4 *A* and S5 *A*). Third, the Spearman's rank correlation of c_{ij}^{FISH} with $\bar{r}_{ij}^{\mathrm{FISH}}$ follows the same trend with R, Δ, and n (Fig. 4 *C*) in accord with our theoretical analyses using the Gaussian polymer chain (Fig. 2 *E*) and HLM for chromosome 5 of GM12878 cell (Fig. 3 *G*). For $n\gg 1$, the correlation of three cosegregation probabilities with $\bar{r}_{ij}$ is maximized at a large R and at a small Δ. However, it is noteworthy that even with $n = 100,000$, the best correlation between c_{ij}^{FISH} and $\bar{r}_{ij}^{\mathrm{FISH}}$ is about $|\rho_s|\simeq 0.63$, which is significantly weaker than the correlation of p_{ij} of Hi-C with $\bar{r}_{ij}^{\mathrm{FISH}}$ ($|\rho_s|\simeq 0.9$; see Fig. 1 *G*). 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 $|\rho_s|$ for NLD and NPMI to exceed 0.9 (Fig. 4 *D* and *E*). Similar conclusions are drawn from chromosomes 12 (Fig. S4) and 18 (Fig. S5).

Stratified comparison of GAM with FISH

Thus far, all intrachromosome segment pairs were taken into account to calculate the ρ_s. To see how the fidelity of GAM data in representing the chromosome architecture changes at different genomic scales, we recalculated ρ_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 (47,48). Despite large variations among different chromosomes (Fig. 5 *A* and S6), at short genomic separation ($|i-j|<40$ Mb), Hi-C contact probability (p_{ij}) is significantly better correlated with $\bar{r}_{ij}$ than

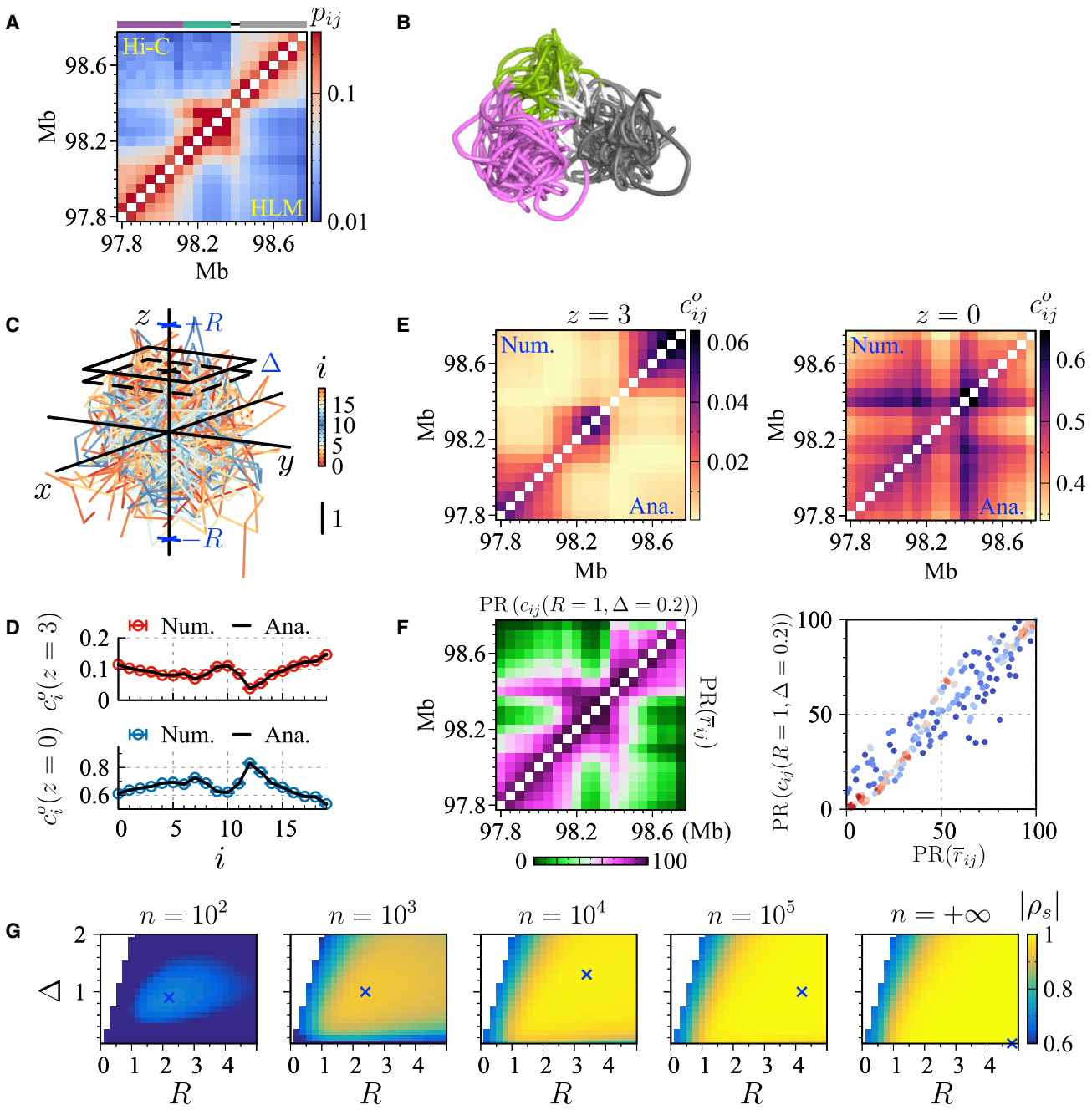

FIGURE 3 Cosegregation probability from the HLM-based structural ensemble at a resolution of 50 kb for a 1-Mb genomic region on chromosome 5 (chr5:97800000–98800000) of GM12878 cells. (*A*) The contact probability (p_{ij}) measured by Rao et al. using Hi-C (6) (*top left*) is compared with that predicted 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 (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. Cosegregation was determined from an ensemble of slices of a thickness Δ and a stochastic height z, satisfying $|z| < R$. (*D*) Segregation and (*E*) cosegregation probabilities, c_i^o and c_{ij}^o, calculated at different values of z with $\Delta = 1$. (*F*) (*Left panel*) PR of the cosegregation probability (*top left*) compared with that of the mean interlocus distance (*bottom right*). (*Right panel*) Scatter plot of $(\mathrm{PR}(c_{ij}), \mathrm{PR}(\bar{r}_{ij}))$. (*G*) Spearman's rank correlation coefficient (ρ_s) of c_{ij} against $\bar{r}_{ij}$ as a function of R and Δ. The symbol $\times$ marks the condition of (R^*, Δ^*) that gives rise to the strongest correlation at different n.

the GAM-based cosegregation probability and its variants (c_{ij}, d_{ij}, s_{ij}). Only for the extremely long-range segment pairs ($|i-j| > 75$ Mb) does the GAM data set marginally outperform the p_{ij} from Hi-C. Given that $p_{ij}^{\mathrm{Hi-C}} (< 0.001)$ is typically orders of magnitude smaller than the corresponding $c_{ij}^{\mathrm{GAM}} (> 0.1)$ for long-range pairs, the bad performance of Hi-C might be an outcome of its relatively higher noise-to-signal ratio (NSR) (Eq. 14).

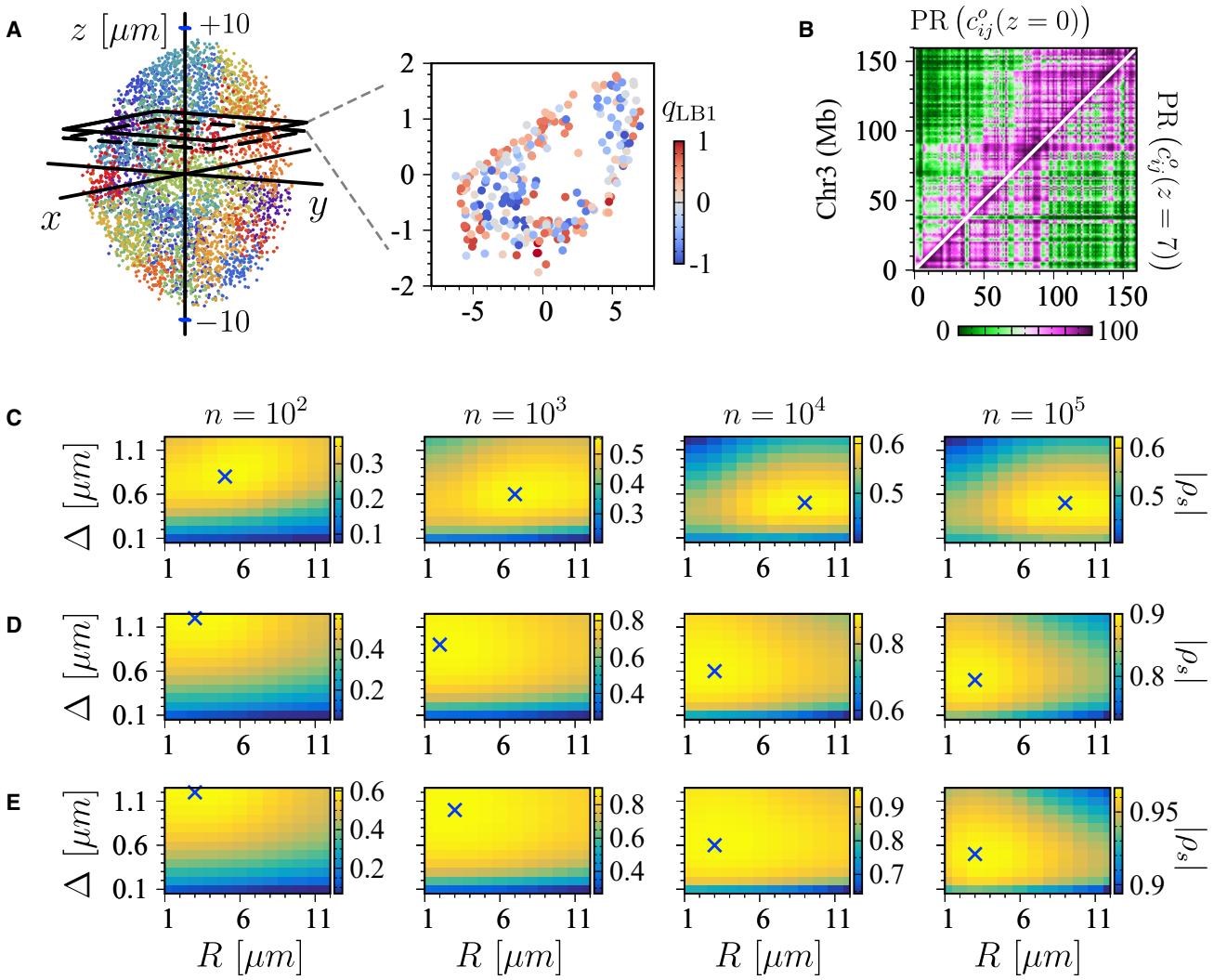

FIGURE 4 In silico GAM on the DNA seqFISH+ data set (24). (*A*) A typical 3D image of the genome of mouse ESC where different chromosomes are distinguished by different colors. Shown on the right is a slice, in which the sectioned loci are color coded by their lamin association propensities (q_{LB1}). (*B*) PR of c_{ij} sectioned by slices with $\Delta = 0.5$ μm and at $z = 0$ and 7 μm, shown in the top left and bottom right panels, respectively. The Spearman's rank correlation coefficient between the two c^o_{ij}s is $\rho_s = 0.83$. The Spearman's rank correlations ($|\rho_s|$) of $\bar{r}^{FISH}_{ij}$ against (*C*) c^{FISH}_{ij}, (*D*) d^{FISH}_{ij} (NLD), and (*E*) s^{FISH}_{ij} (NPMI) as a function of the slice range of nuclei (R) and slice width (Δ) with varying sample size (n), where the $\times$ symbol marks the value of (R,Δ) that maximizes $|\rho_s|$. The heatmaps in (*C*)–(*E*) quantify the variations of correlation over the full range of their respective data.

Lastly, for two intrachromosomal 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 interlocus 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 ($\bar{r}_{ij}$ versus $\bar{r}_{kl}$). When the p-value is smaller, it is more likely that the two pairs display different distances (Fig. 5 *B*); if $r_{ij} < r_{kl}$, then $p_{ij} > p_{kl}$ is highly likely. However, according to our explicit calculation summarized in Fig. 5 *C*, the above statement regarding the relation between interlocus distance and contact probability does not nicely translate into the GAM-based cosegregation probabilities. In Fig. 5 *C*, 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 cosegregation probabilities (c_{ij}, d_{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.,

$$f_x = \frac{n\left[\left(x_{ij}/x_{kl} > 1\right) \text{ and } \left(P < P^*\right)\right]}{n[P < P^*]}, \tag{13}$$

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

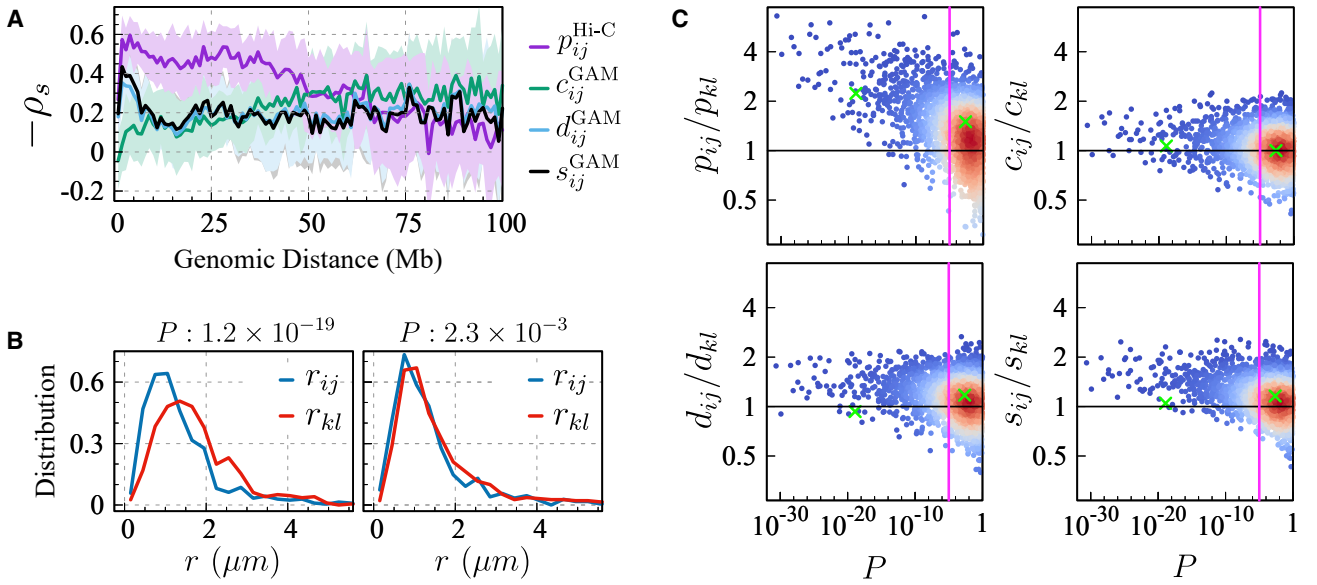

FIGURE 5 Stratified comparison of GAM (14) and Hi-C (5) against FISH data (24). (*A*) Autosome-averaged Spearman's rank correlation coefficient ($|\rho_s|$) between $\bar{r}_{ij}^{\text{FISH}}$ and Hi-C contact probability (p_{ij}), GAM cosegregation probability (c_{ij}), and its two variants (d_{ij}, s_{ij}) as a function of the genomic separation between the two sites $(i,\ j)$. The mean and the standard deviation of ρ_s averaged over all autosomes are plotted with solid lines and shades, respectively. (*B*) Comparison between the distance distributions of one intrachromosomal 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 (×). The vertical line depicts $p = 1 \times 10^{-5}$.

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 cosegregation probability, the corresponding noise-to-signal ratio (NSR) is given by

$$\text{NSR}_{\text{GAM}} = \frac{\sigma_{c_{ij}}}{c_{ij}} = \sqrt{\frac{1 - c_{ij}}{c_{ij} n}}, \tag{14}$$

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 cosegregation probabilities, which display power-law decays ($p_{ij},\ c_{ij} \sim |i - j|^{-\alpha}$) (49), measurements are less reliable (or large NSR) for long-range segment pairs (Eq. 14). Furthermore, since the variations of the cosegregation probability with respect to the variations of Δ and R satisfy $\partial c_{ij}/\partial\Delta < 0$ and $\partial c_{ij}/\partial R > 0$ for $n \gg 1$ (Eq. 12), the NSR_{GAM} is reduced for $\delta\Delta < 0$ and $\delta R > 0$, i.e.,

$$\delta(\text{NSR}_{\text{GAM}}) = -\frac{(\partial c_{ij}/\partial\Delta)\delta\Delta + (\partial c_{ij}/\partial R)\delta R}{(2nc_{ij}^2 \text{NSR}_{\text{GAM}})} < 0. \tag{15}$$

The reduced NSR_{GAM} at larger R and smaller Δ accounts for better correlation with $\bar{r}_{ij}$, which manifests itself as the enhanced $|\rho_s|$ for $n \gg 1$ under the same condition (Figs. 2 *E* and 3 *G*).

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

$$\text{NSR}_{\text{FISH}} = \frac{\sigma_{r_{ij}}}{\bar{r}_{ij}} = \sqrt{\frac{3\pi - 8}{8n}}. \tag{16}$$

A 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 cosegregation 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 cosegregation probability changes nonmonotonically 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 cosegregation probability depends on the slice range of nuclei and the thickness and number of nuclear sections. The correlation between the cosegregation 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 at short genomic range (Fig. 5). As a result, GAM data are not always straightforward to interpret. Thus, the findings made with GAM analysis, such as interchromosome contact, higher-order interaction, and associations with lamin and nucleolus, are usually supplemented with other measurements (9,14). Our in silico GAM experiment shows that if

(co)segregation probabilities from NPs with slice thickness of $\sim$0.5 μm are the only available data, then as long as the data are collected from a sufficient number ($n \gtrsim 10^3$) of NPs, NPMI (s_{ij}) is, albeit not perfect, slightly better suited than NLD (d_{ij}) for faithful interpretation of the 3D chromatin organization.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2022.09.018.

AUTHOR CONTRIBUTIONS

L.L. and C.H. designed the research. L.L. and X.C. carried out simulations. L.L., X.C., B.Z., and C.H. analyzed the data and wrote the article.

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, and 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.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Szabo, Q., D. Jost, …, 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, …, 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, …, 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, …, 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, …, 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, …, 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, …, O. J. Rando. 2020. Ultrastructural details of mammalian chromosome architecture. *Mol. Cell.* 78:554–565.e7, e7.
8. Beagrie, R. A., C. J. Thieme, …, S. Bianco. 2021. Multiplex-GAM: genome-wide identification of chromatin contacts yields insights not captured by Hi-C. Preprint at bioRxiv. https://doi.org/10.1101/2020.07.31.230284.
9. Winick-Ng, W., A. Kukalev, …, A. Pombo. 2021. Cell-type specialization is encoded by specific chromatin topologies. *Nature.* 599:684–691.
10. Ou, H. D., S. Phan, …, C. C. O'Shea. 2017. ChromEMT: visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science.* 357:eaag0025.
11. Dekker, J., A. S. Belmont; …, 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., A. Scialdone, …, 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. Tanay. 2016. Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature.* 540:296–300.
17. Darrow, E. M., M. H. Huntley, …, E. L. Aiden. 2016. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc. Natl. Acad. Sci. USA.* 113:E4504–E4512.
18. Oudelaar, A. M., J. O. J. Davies, …, 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, …, 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, …, 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, …, 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, …, L. Cai. 2021. Integrated spatial genomics reveals global architecture of single nuclei. *Nature.* 590:344–350.
25. Takei, Y., S. Zheng, …, L. Cai. 2021. Single-cell nuclear architecture across cell types in the mouse brain. *Science.* 374:586–594.
26. Kamat, K., Y. Qi, …, B. Zhang. 2021. Genome compartmentalization with nuclear landmarks: random yet precise. Preprint at bioRxiv. https://doi.org/10.1101/2021.11.12.468401.
27. Fiorillo, L., F. Musella, …, 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, …, X. Zhuang. 2018. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science.* 362:eaau1783.
29. Cardozo Gizzi, A. M., D. I. Cattoni, …, M. Nollmann. 2019. Microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms. *Mol. Cell.* 74:212–222.e5.
30. Liu, M., Y. Lu, …, S. Wang. 2020. Multiplexed imaging of nucleome architectures in single cells of mammalian tissue. *Nat. Commun.* 11:2907.
31. Durand, N. C., M. S. Shamim, …, E. L. Aiden. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* 3:95–98.
32. Kuhn, R. M., D. Haussler, and W. J. Kent. 2013. The UCSC genome browser and associated tools. *Briefings Bioinf.* 14:144–161.

33. Peric-Hupkes, D., W. Meuleman, …, B. van Steensel. 2010. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell.* 38:603–613.

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

35. Lewontin, R. C. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics.* 49:49–67.

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

37. Razak, F. A. 2014. The derivation of mutual information and covariance function using centered random variables. *In* AIP Conf. Proc.. American Institute of Physics, pp. 883–889.

38. Bohn, M., D. W. Heermann, and R. van Driel. 2007. Random loop model for long polymers. *Phys. Rev. E - Stat. Nonlinear Soft Matter Phys.* 76:051805.

39. 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.

40. 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.

41. 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.

42. Shinkai, S., M. Nakagawa, …, S. Onami. 2020. PHi-C: deciphering Hi-C data into polymer dynamics. *NAR Genom. Bioinform.* 2:lqaa020.

43. Shinkai, S., T. Sugawara, …, S. Onami. 2020. Microrheology for Hi-C data reveals the spectrum of the dynamic 3D genome organization. *Biophys. J.* 118:2220–2228.

44. Shi, G., and D. Thirumalai. 2021. From Hi-C contact map to three-dimensional organization of interphase human chromosomes. *Phys. Rev. X.* 11:011051.

45. Shi, G., and D. Thirumalai. 2019. Conformational heterogeneity in human interphase chromosome organization reconciles the FISH and Hi-C paradox. *Nat. Commun.* 10:3894–3910.

46. Jia, B. B., A. Jussila, …, B. Ren. 2022. A spatial genome aligner for multiplexed DNA-FISH. Preprint at bioRxiv. https://doi.org/10.1101/2022.03.25.485845.

47. Bianco, S., D. G. Lupiáñez, …, M. Nicodemi. 2018. Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat. Genet.* 50:662–667.

48. Yang, T., F. Zhang, …, Q. Li. 2017. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* 27:1939–1949.

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

Biophysical Journal, Volume *121*

Supplemental information

Dissecting the cosegregation probability from genome architecture mapping

Lei Liu, Xinmeng Cao, Bokai Zhang, and Changbong Hyeon

Supplementary Information: Dissecting the co-segregation probability from genome architecture mapping

Lei Liu,[1] Xinmeng Cao,[1] Bokai Zhang,[2] and Changbong Hyeon[3]
[1)] *Key Laboratory of Optical Field Manipulation of Zhejiang Province, Department of Physics, Zhejiang Sci-Tech University, Hangzhou 310018, China*
[2)] *School of Physical Science and Technology, Southwest University, Chongqing 400715, China*
[3)] *Korea Institute for Advanced Study, Seoul 02455, Korea*

(Dated: 15 September 2022)

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[1–5], 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

$$U(\boldsymbol{r}) = \sum_{i=0}^{N-1}\sum_{j=i+1}^{N-1}\frac{k_{ij}}{2}|\vec{r}_i - \vec{r}_j|^2 = \sum_{\alpha}\frac{1}{2}\boldsymbol{r}_\alpha^{\mathrm{T}} L \boldsymbol{r}_\alpha \qquad \text{(S1)}$$

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

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

$$\Sigma = Q\ \mathrm{diag}\left(0, \lambda_1^{-1}, \lambda_2^{-1}, \cdots, \lambda_{N-1}^{-1}\right) Q^{\mathrm{T}}, \qquad \text{(S2)}$$

where the i-th column of the orthorgonal matrix Q and λ_i are the corresponding i-th eigenvector and eigenvalue of $\mathcal{L}$, respectively. By using the notation of

$$\gamma_{ij} = \frac{1}{2}\left(\sigma_{ii} + \sigma_{jj} - 2\sigma_{ij}\right)^{-1} \qquad \text{(S3)}$$

where σ_{ij} is the (i,j)-th element of Σ, 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

$$\begin{aligned}\bar{r}_{ij} &= \frac{2}{\sqrt{\gamma_{ij}\pi}}\\ \bar{r^2_{ij}} &= \frac{3}{2\gamma_{ij}}.\end{aligned} \qquad \text{(S4)}$$

2. CONTACT PROBABILITY

Based on the above chromatin polymer model, we have discussed generic n-body contact probability in Ref.[6]. The pairwise ($n = 2$) contact probability is given by

$$\begin{aligned}p_{ij}^{\mathrm{G}}(r_c) &= \left(1 + \frac{3}{2\gamma_{ij}r_c^2}\right)^{-3/2}\\ p_{ij}^{\mathrm{S}}(r_c) &= \mathrm{erf}\left(\gamma_{ij}^{1/2} r_c\right) - 2\Delta\sqrt{\frac{\gamma_{ij}}{\pi}}e^{-\gamma_{ij}r_c^2}\end{aligned} \qquad \text{(S5)}$$

with the special function $\mathrm{erf}(x) = 2\pi^{-1/2}\int_0^x dt e^{-t^2}$. 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

$$\begin{aligned}F_{\text{Hi-C}}^{\mathrm{G}}(r) &= e^{-\frac{3r^2}{2\Delta^2}}\\ F_{\text{Hi-C}}^{\mathrm{S}}(r) &= \begin{cases}1, & r < r_c\\ 0, & \text{otherwise}\end{cases}\end{aligned} \qquad \text{(S6)}$$

to obtain Eqs. S5 where Δ denotes a characteristic capture radius.

3. CO-SESEGREGATION PROBABILITY

To derive the co-segregation probability of chromatin segment pairs as the GAM experiment[7], we begin with the probability that the i-th segment is located at $z_i = \alpha$ from the COM of the chain,

$$\begin{aligned}p_{z_i=\alpha} = \langle\delta(z_i - \alpha)\rangle &= \frac{\int \mathcal{D}\boldsymbol{z}\,\delta(z_i - \alpha)e^{-\frac{1}{2}\boldsymbol{z}^{\mathrm{T}}L\boldsymbol{z}}}{\int \mathcal{D}\boldsymbol{z}e^{-\frac{1}{2}\boldsymbol{z}^{\mathrm{T}}L\boldsymbol{z}}}\\ &\propto \int_{-\infty}^{+\infty}dq\int_{-\infty}^{+\infty}\mathcal{D}\boldsymbol{z}e^{-iq(z_i-\alpha)}e^{-\frac{1}{2}\boldsymbol{z}^{\mathrm{T}}L\boldsymbol{z}}\\ &= \frac{1}{\sqrt{2\pi\sigma_{ii}}}e^{-\frac{1}{2}\alpha^2/\sigma_{ii}}.\end{aligned} \qquad \text{(S7)}$$

Similarly, the probability that the i-th and j-th segments are simultaneously found at $z_i = \alpha$ and $z_j = \beta$, respec-

tively, is given by

$$\begin{aligned} p_{z_i=\alpha,z_j=\beta} &= \langle \delta(z_i-\alpha)\delta(z_j-\beta)\rangle \\ &\propto \int_{-\infty}^{+\infty} dq \int_{-\infty}^{+\infty} dk \int_{-\infty}^{+\infty} \mathcal{D}\boldsymbol{z} e^{-iq(z_i-\alpha)} e^{-ik(z_j-\beta)} e^{-\frac{1}{2}\boldsymbol{z}^{\mathrm{T}} L \boldsymbol{z}} \\ &= \frac{1}{2\pi\sqrt{M_{ij}}} e^{-\frac{1}{2M_{ij}}\left(\sigma_{ii}\alpha^2+\sigma_{jj}\beta^2-2\sigma_{ij}\alpha\beta\right)} \end{aligned} \tag{S8}$$

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

Gaussian sectioning. Next, we assume that the horizontal slice (S_z^Δ) at $h = z$ with thickness Δ is sectioned with a Gaussian probability

$$F_{\mathrm{GAM}}^{\mathrm{G}}(\alpha) = e^{-\frac{1}{2}(\alpha-z)^2/\Delta^2} \tag{S9}$$

where the superscript again notates the Gaussian profile of F_{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_z^Δ as

$$\begin{aligned} c_i^{\mathrm{o,G}}(z,\Delta) &= \int_{-\infty}^{+\infty} d\alpha \; F_{\mathrm{GAM}}^{\mathrm{G}}(\alpha) \times p_{z_i=\alpha} \\ &= \frac{\Delta}{\sqrt{\sigma_{ii}+\Delta^2}} e^{-\frac{1}{2}z^2/\left(\sigma_{ii}+\Delta^2\right)}, \end{aligned} \tag{S10}$$

and the co-segregation probability as

$$\begin{aligned} c_{ij}^{\mathrm{o,G}}(z,\Delta) &= \int d\alpha \int d\beta \; F_{\mathrm{GAM}}^{\mathrm{G}}(\alpha) F_{\mathrm{GAM}}^{\mathrm{G}}(\beta) \times p_{z_i=\alpha,z_j=\beta} \\ &= \frac{\Delta^2}{\sqrt{\Omega_{ij}}} e^{-\frac{1}{2}z^2\frac{\Psi_{ij}}{\Omega_{ij}}}, \end{aligned} \tag{S11}$$

where

$$\begin{aligned} \Psi_{ij} &= \sigma_{ii}+\sigma_{jj}-2\sigma_{ij}+2\Delta^2 \\ \Omega_{ij} &= (\sigma_{ii}+\Delta^2)(\sigma_{jj}+\Delta^2)-\sigma_{ij}^2. \end{aligned} \tag{S12}$$

- **NPs collected from uniform slicing.** To mimic the procedure of GAM[7], we assume that the NPs are collected uniformly from $z \in [-R, R]$ ($R \geq 0$). Then the segregation probability is given by

$$\begin{aligned} c_i^{\mathrm{G}}(R,\Delta) &= \frac{1}{2R}\int_{-R}^{+R} dz \; c_i^{\mathrm{o,G}}(z,\Delta) \\ &= \sqrt{\frac{\pi}{2}}\frac{\Delta}{R}\mathrm{erf}\left(\frac{R}{\sqrt{2(\sigma_{ii}+\Delta^2)}}\right), \end{aligned} \tag{S13}$$

and the co-segregation probability has the expression of

$$\begin{aligned} c_{ij}^{\mathrm{G}}(R,\Delta) &= \frac{1}{2R}\int_{-R}^{+R} dz \; c_{ij}^{\mathrm{o,G}}(z,\Delta) \\ &= \sqrt{\frac{\pi}{2}}\frac{\Delta^2}{R}\Psi_{ij}^{-1/2}\mathrm{erf}\left(\sqrt{\frac{\Psi_{ij}}{2\Omega_{ij}}}R\right). \end{aligned} \tag{S14}$$

- **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

$$\begin{aligned} c_i^{\mathrm{G}^*}(R,\Delta) &= \frac{1}{\sqrt{2\pi}R}\int_{-\infty}^{+\infty} dz \; c_i^{\mathrm{o,G}}(z,\Delta) e^{-\frac{z^2}{2R^2}} \\ &= \frac{\Delta}{\sqrt{\sigma_{ii}+\Delta^2+R^2}}, \end{aligned} \tag{S15}$$

and the co-segregation probability as

$$\begin{aligned} c_{ij}^{\mathrm{G}^*}(R,\Delta) &= \frac{1}{\sqrt{2\pi}R}\int_{-\infty}^{+\infty} dz \; c_{ij}^{\mathrm{o,G}}(z,\Delta) e^{-\frac{z^2}{2R^2}} \\ &= \frac{\Delta^2}{\sqrt{\Psi_{ij}R^2+\Omega_{ij}}}. \end{aligned} \tag{S16}$$

As shown in Fig. S2, c_{ij}^{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 $F_{\mathrm{GAM}}^{\mathrm{G}}$ in Eq. S9 with

$$F_{\mathrm{GAM}}^{\mathrm{S}}(\alpha) = \begin{cases} 1, & |\alpha - z| < \Delta \\ 0, & \text{otherwise} \end{cases}. \tag{S17}$$

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

$$\begin{aligned} c_i^{\mathrm{o,S}}(z,\Delta) &= \int_{-\infty}^{+\infty} d\alpha \; F_{\mathrm{GAM}}^{\mathrm{S}}(\alpha) \times p_{z_i=\alpha} \\ &= \frac{1}{2}\left(\mathrm{erf}\left(\frac{z+\Delta}{\sqrt{2\sigma_{ii}}}\right) - \mathrm{erf}\left(\frac{z-\Delta}{\sqrt{2\sigma_{ii}}}\right)\right) \end{aligned} \tag{S18}$$

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

$$c_i^{\mathrm{S}}(R,\Delta) = \frac{1}{2R}\int_{-R}^{+R} dz c_i^{\mathrm{o,S}}(z,\Delta)$$
$$= \frac{B}{2R}\left(\frac{\Delta+R}{B}\mathrm{erf}\left(\frac{\Delta+R}{\sqrt{2\sigma_{ii}}}\right) - \frac{\Delta-R}{B}\mathrm{erf}\left(\frac{\Delta-R}{\sqrt{2\sigma_{ii}}}\right) - \sqrt{\frac{2\sigma_{ii}}{\pi}}\left(e^{\frac{2R\Delta}{\sigma_{ii}}}-1\right)\right) \quad \text{(S19)}$$

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

$$c_{ij}^{\mathrm{o,S}}(z,\Delta) = \int d\alpha \int d\beta\ F_{\mathrm{GAM}}^{\mathrm{S}}(\alpha)F_{\mathrm{GAM}}^{\mathrm{S}}(\beta) \times p_{z_i=\alpha,z_j=\beta}$$
$$= \frac{1}{2\sqrt{2\pi\sigma_{ii}}}\int_{z-\Delta}^{z+\Delta} d\alpha e^{-\frac{\alpha^2}{2\sigma_{ii}}}\left(\mathrm{erf}\left[\sqrt{\frac{\sigma_{ii}}{2M_{ij}}}\left(-\frac{\sigma_{ij}}{\sigma_{ii}}\alpha + z + \Delta\right)\right] - \mathrm{erf}\left[\sqrt{\frac{\sigma_{ii}}{2M_{ij}}}\left(-\frac{\sigma_{ij}}{\sigma_{ii}}\alpha + z - \Delta\right)\right]\right), \quad \text{(S20)}$$

and $c_{ij}^{\mathrm{S}}(R,\Delta) = \frac{1}{2R}\int_{-R}^{+R} dz\ c_{ij}^{\mathrm{o,S}}(z,\Delta)$ 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 $\boldsymbol{X}_\alpha = Q^{\mathrm{T}}\boldsymbol{r}_\alpha$ satisfies

$$\langle X_{p,\alpha}X_{q,\beta}\rangle = \frac{k_{\mathrm{B}}T}{\lambda_p}\delta_{pq}\delta_{\alpha\beta}, \quad \text{(S21)}$$

where Q and λ are defined in Eq. S2, $k_{\mathrm{B}}T$ is our energy unit, α and β represent x, y and z, and $p, q = 1, 2, 3, \cdots, N-1$. Based on this relation, a 3D conformation of the polymer chain can be generated in two steps[3]. First, we obtain the normal coordinates by using

$$X_{p,\alpha} = \begin{cases} 0, & p = 0 \\ \xi_\alpha/\sqrt{\lambda_p}, & p > 0 \end{cases}, \quad \text{(S22)}$$

where the random variable ξ_α 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

$$\boldsymbol{r}_\alpha = Q\boldsymbol{X}_\alpha. \quad \text{(S23)}$$

Repeating this protocol n times yields a structural ensemble of the polymer chain containing n samples.

[1] M. Bohn, D. W. Heermann, and R. van Driel, "Random loop model for long polymers," Phys. Rev. E **76**, 051805 (2007).

[2] G. Le Treut, F. Képès, and H. Orland, "A polymer model for the quantitative reconstruction of chromosome architecture from HiC and GAM data," Biophys. J. **115**, 2286–2294 (2018).

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

[4] L. Liu, M. H. Kim, and C. Hyeon, "Heterogeneous loop model to infer 3D chromosome structures from Hi-C," Biophys. J. **117**, 613–625 (2019).

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

[6] L. Liu, B. Zhang, and C. Hyeon, "Extracting multi-way chromatin contacts from Hi-C data," PLoS Comput. Biol. **17**, e1009669 (2021).

[7] R. A. Beagrie, 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, "Complex multi-enhancer contacts captured by genome architecture mapping," Nature **543**, 519–524 (2017).

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

[9] Y. Takei, 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, "Integrated spatial genomics reveals global architecture of single nuclei," Nature **590**, 344–350 (2021).

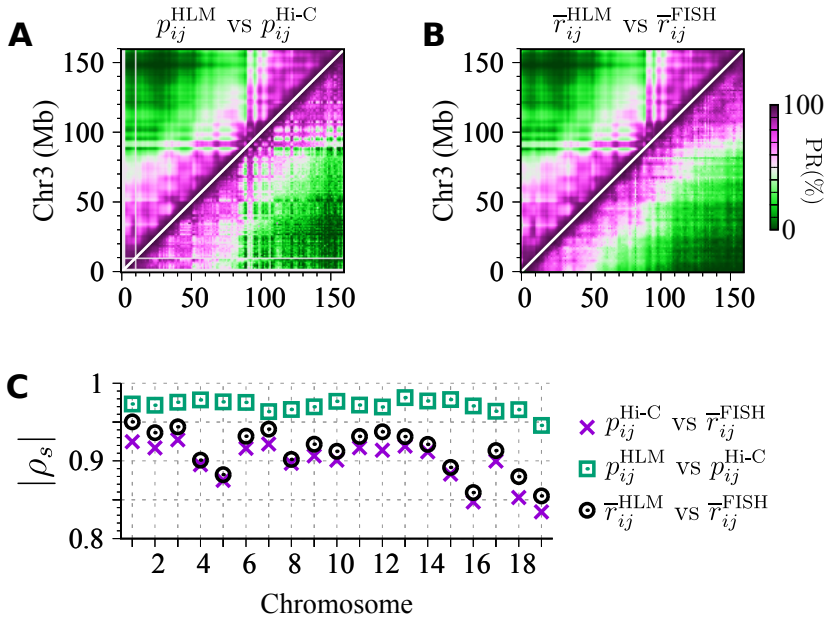

FIG. S1. Comparing mouse ESC autosomes modeled at 1-Mb resolution with HLM against experimental data. (A) Percentile rank (PR) of the contact probability of chromosome 3 from HLM (p_{ij}^{HLM}, top left conner) versus that from Hi-C[8] ($p_{ij}^{\mathrm{Hi\text{-}C}}$, bottom right conner). (B) PR of the mean inter-locus spatial distance from HLM ($\overline{r}_{ij}^{\mathrm{HLM}}$, top left conner) versus that from FISH[9] ($\overline{r}_{ij}^{\mathrm{FISH}}$, bottom right conner). (C) The Spearman's rank correlation coefficient for all autosomes.

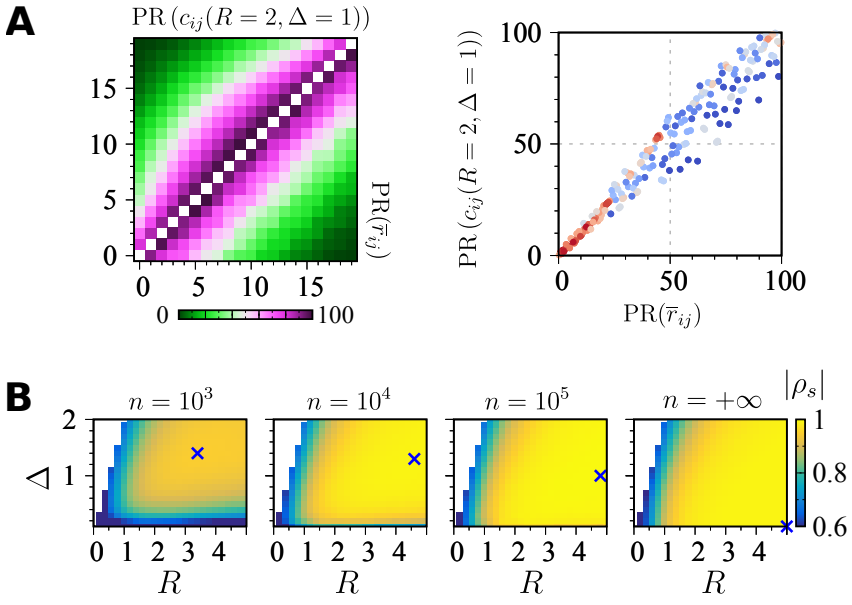

FIG. S2. Co-segregation probability of a Gaussian polymer chain which are calculated by assuming the slices position following a Gaussian profile (see Eqs. S15 and S16). **(A)** and **(B)** are similar to Fig. 2D-E in the main text, respectively.

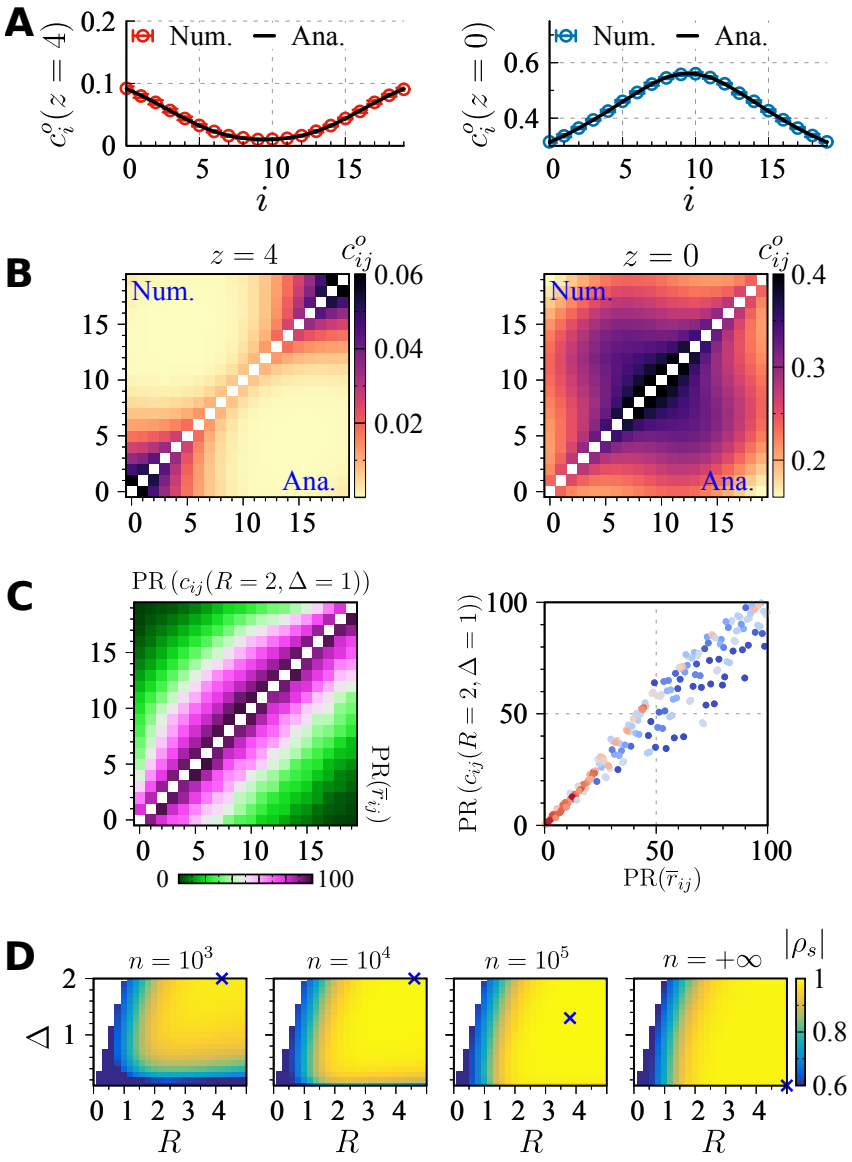

FIG. S3. Co-segregation probability of a Gaussian polymer chain which are calculated by assuming the rectangular sectioning profile (see Eqs. S17-S20). **(A)-(D)** are similar to Fig. 2B-E in the main text, respectively.

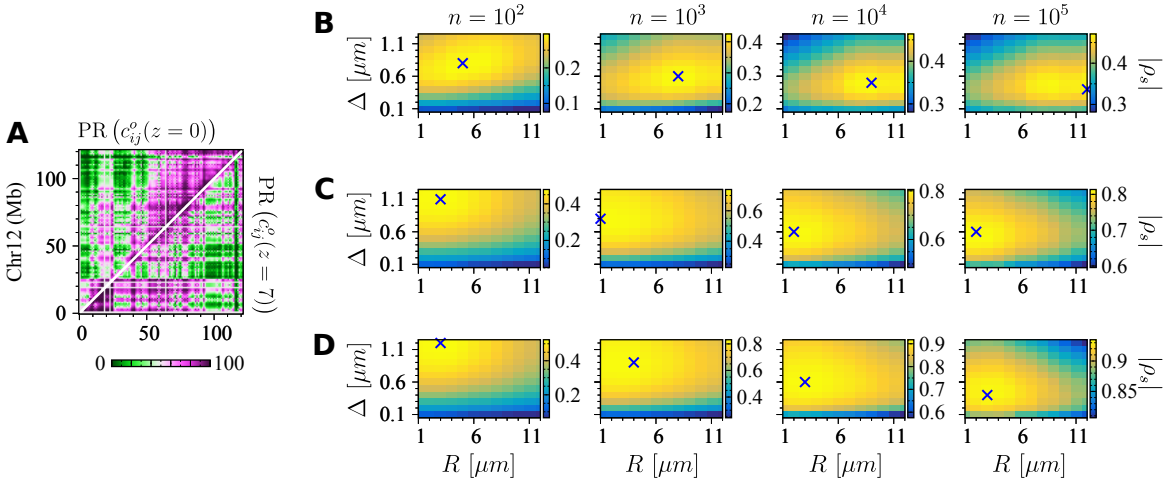

FIG. S4. In silico GAM results of chromosome 12. **(A)** Percentile ranks of c_{ij} sectioned by slices with a width of $\Delta = 0.5$ μm and a height of $z = 0$ (top-left) and 7 μm (bottom-right). The Spearman's correlation coefficient between the two is $\rho_s = 0.75$. **(B)** The correlations (ρ_s) between the mean physical distance and co-segregation probability, **(C)** NLD, and **(D)** NPMI as a function of the positional uncertainty (R) and width (Δ) of slices for varying n. The symbol "$\times$" denotes the values of (R, Δ) that maximizes $|\rho_s|$.

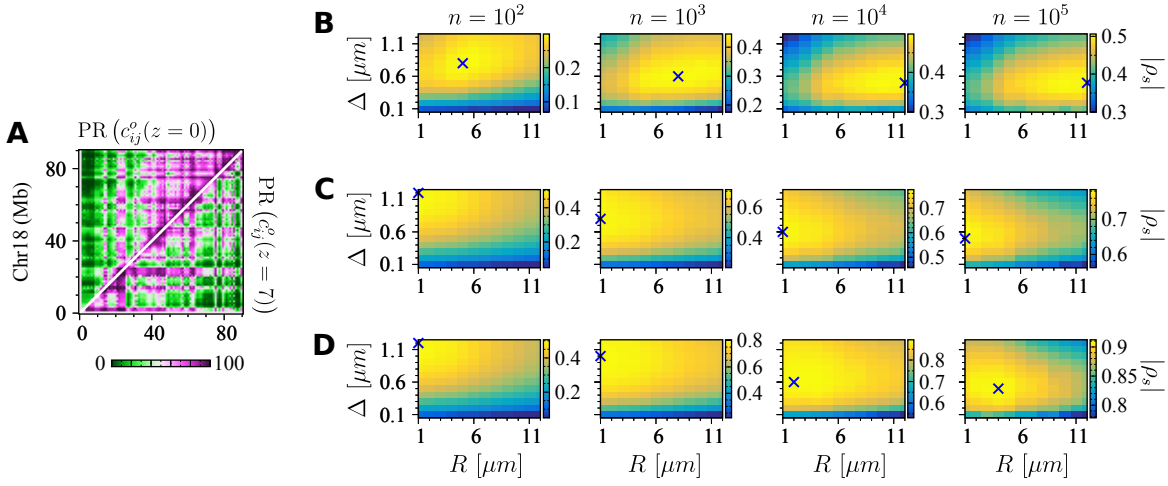

FIG. S5. In silico GAM results of chromosome 18. Please refer to the caption of Fig. S4 for explanations. The Spearman's correlation between the top-left and bottom-right panels in **(A)** is 0.69.

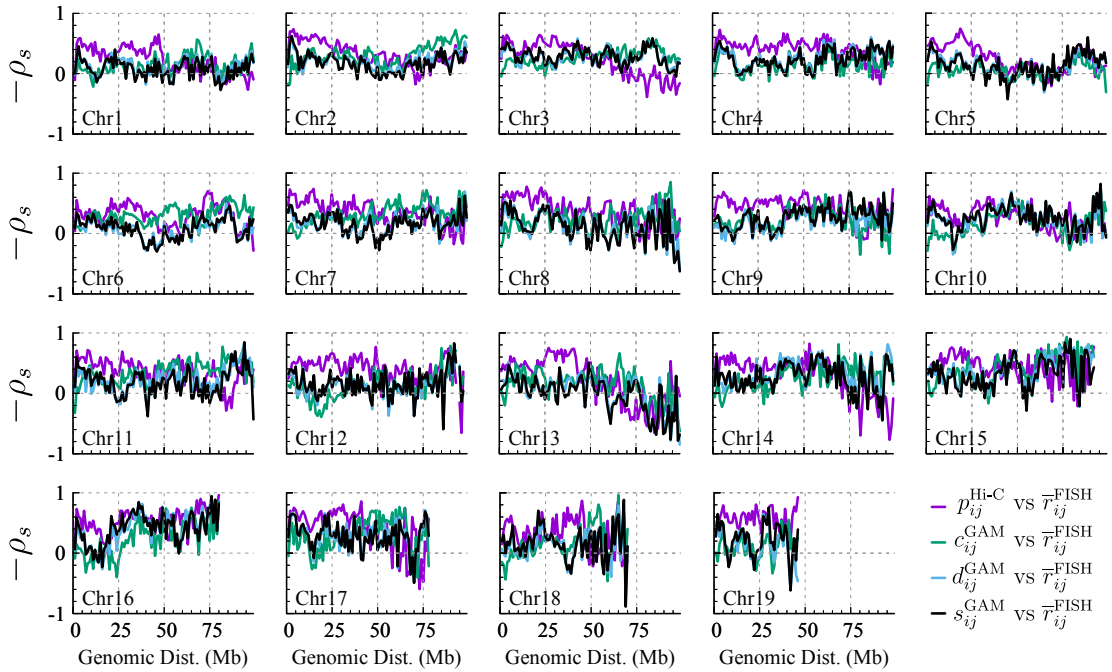

FIG. S6. The Spearman's correlation between the mean physical distance ($\bar{r}_{ij}$) and Hi-C contact probability (p_{ij}), GAM co-segregation probability (c_{ij}), and its two variants (d_{ij} and s_{ij}) as a function of the genomic distance between the (i, j) sites in different chromosomes. The mean and standard deviation averaged over all the autosomes are shown in Fig. 5A in the main text.