

RESEARCH ARTICLE

Extracting multi-way chromatin contacts from Hi-C data

Lei Liu^{1*}, Bokai Zhang¹, Changbong Hyeon^{2*}**1** Key Laboratory of Optical Field Manipulation of Zhejiang Province, Department of Physics, Zhejiang Sci-Tech University, Hangzhou, China, **2** School of Computational Sciences, Korea Institute for Advanced Study, Seoul, Korea* leiliu@zstu.edu.cn (LL); hyeoncb@kias.re.kr (CH)

OPEN ACCESS

Citation: Liu L, Zhang B, Hyeon C (2021) Extracting multi-way chromatin contacts from Hi-C data. *PLoS Comput Biol* 17(12): e1009669. <https://doi.org/10.1371/journal.pcbi.1009669>**Editor:** Ferhat Ay, La Jolla Institute for Allergy and Immunology, UNITED STATES**Received:** June 14, 2021**Accepted:** November 19, 2021**Published:** December 6, 2021**Copyright:** © 2021 Liu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.**Data Availability Statement:** The software package and associated documentation are available online (<https://github.com/leiliu2015/HLM-Nbody>).**Funding:** This work was supported in part by National Natural Science Foundation of China (12104404 to L.L.), ZSTU intramural grant (20062226-Y to L.L.) at Zhejiang Sci-Tech University, and KIAS Individual Grant (CG035003 to C.H.) at Korea Institute for Advanced Study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

There is a growing realization that multi-way chromatin contacts formed in chromosome structures are fundamental units of gene regulation. However, due to the paucity and complexity of such contacts, it is challenging to detect and identify them using experiments. Based on an assumption that chromosome structures can be mapped onto a network of Gaussian polymer, here we derive analytic expressions for n -body contact probabilities ($n > 2$) among chromatin loci based on pairwise genomic contact frequencies available in Hi-C, and show that multi-way contact probability maps can in principle be extracted from Hi-C. The three-body (triplet) contact probabilities, calculated from our theory, are in good correlation with those from measurements including Tri-C, MC-4C and SPRITE. Maps of multi-way chromatin contacts calculated from our analytic expressions can not only complement experimental measurements, but also can offer better understanding of the related issues, such as cell-line dependent assemblies of multiple genes and enhancers to chromatin hubs, competition between long-range and short-range multi-way contacts, and condensates of multiple CTCF anchors.

Author summary

The importance of DNA looping is often mentioned as the initiation step of gene expression. However, there are growing evidences that ‘chromatin hubs’ comprised of multiple genes and enhancers play vital roles in gene expressions and regulations. Currently a number of experimental techniques to detect and identify multi-way chromosome interactions are available; yet detection of such multi-body interactions is statistically challenging. This study proposes a method to predict multi-way chromatin contacts from pairwise contact frequencies available in Hi-C dataset. Since chromosomes are made of polymer chains, the pairwise contact probabilities are not entirely independent from each other, but certain types of correlations are present reflecting the underlying chromosome structure. We extract these correlations hidden in Hi-C dataset by leveraging theoretical argument based on polymer physics.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Recent advances in experimental techniques [1–6] offer unprecedented glimpses into the chromosome structures inside cell nuclei, in the form of pairwise distances and contact frequencies between genomic loci. Gene expressions are, however, realized when multiple genomic loci, e.g., promoters and enhancers separated over large genomic distances, are brought together to form a regulatory element [7–10], which underscores the importance of resolving chromatin interactions beyond pairwise two-body contacts. In particular, the transcriptional regulation is fine-tuned via a complex network of cooperative and competitive interactions among nearby genes mediated by transcription factors [11, 12].

A plethora of experimental methods have recently been developed to detect multi-way contacts in chromosome [13], which include super-resolution chromatin tracing [14], and both ligation-based (3way-4C [15], COLA [16], Tri-C [17], MC-4C [18, 19], Pore-C [20] and single-cell techniques [21–24]) and ligation-free methods (GAM [25] and SPRITE [26]). Detection of multi-way chromatin contacts using these experimental methods, however, are statistically limited due to the paucity of such contacts. The probability of detecting a particular n -body contacts from a genomic region of interest consisting of N statistical segments is approximately $p_n \sim \binom{N}{n}^{-1} \langle c \rangle^n V$ ($n \geq 2$), where $\langle c \rangle \sim N/V$ is the effective concentration of the segments in the volume V which can be approximated using the Flory radius [27] ($R_F \sim N^\nu$) to $V \sim R_F^3 \sim N^{3\nu}$ with $\nu(\geq 1/3)$ being the Flory exponent [28, 29]. Since $\binom{N}{n} \sim N^n/n!$ for $N \gg n$, it is expected that the n -body contact probability scales as $p_n \sim n!/N^{3\nu(n-1)}$, which makes experimental detection of n -body contacts with larger n statistically more demanding. To get around the detection problem, one usually either conducts a genome-wide study at low resolution (a small N), or performs a high-resolution experiment by focusing only on the contacts formed at a few prescribed sites [16, 30].

Since the polymer physics idea was first used to explore the physical characteristics of chromosomes as long polymer chains confined in a small nuclear space [28, 31–38], computational strategies of incorporating genomic constraints from experimental measurements and epigenomic information into polymer-based modeling of 3D chromosome structure have recently gained much traction [39–54]. Many studies, which generate an ensemble of 3D chromosome structures, highlight the heterogeneous and probabilistic nature of chromosome structure [45, 47, 50, 51, 54]. Once an ensemble of structural models of chromosomes are obtained from computational approaches, it is straightforward to count the multi-way chromatin contacts directly from them and to quantify the corresponding contact probabilities. Specifically, recent studies, which have generated 3D chromosome structures based on the strings and binders switch (SBS) model, have shown that the probabilities of triplet chromatin contacts for HoxD and α -globin regions calculated from the structures are in good agreement with those from 3way-4C and Tri-C experiments, respectively [55, 56]. CHROMatin mIXture (CHROMATIX) model was used to address the functional relevance of many-body contacts among genomic elements enriched at transcriptionally active loci [30].

In this study, we calculate n -body contact probabilities from Hi-C data by using analytic expressions derived from the formalism of Heterogeneous Loop Model (HLM) [50, 51, 57]. Although the original aim of HLM was to reconstruct 3D chromosome structures from Hi-C [50, 51], it is still possible to calculate the map of n -body contact probabilities without explicitly counting those contacts from the generated chromosome structures. Through comparisons between the multi-way chromatin contacts derived from HLM and those from three separate measurements from Tri-C, MC-4C, and SPRITE, we will show that the n -body contact probability maps are in good agreement with experimental measurements (Results). We also explore the relation between pairwise and higher-order contacts and discuss how to avoid the most

evident false-positiveness in the analysis of multi-way chromatin contacts experiments ([Discussions](#)). Finally, the numerical details of training HLM based on Hi-C and the mathematical derivations for contact probabilities are provided in the [Methods](#) section.

Results

Polymer theory of *n*-body chromatin contacts

The chromatin fiber in a genomic region of interest is modeled as a coarse-grained polymer chain composed of *N* monomers (or sites), each representing a chromatin segment of a pre-scribed genomic length. We assume that the chromatin effective energy landscape can be described by a summation of harmonic restraints on the spatial distances between all monomer pairs,

$$U_{\mathcal{K}}(\mathbf{r}) = \sum_{i=0}^{N-1} \sum_{j=i+1}^{N-1} \frac{k_{ij}}{2} (\vec{r}_i - \vec{r}_j)^2 = \frac{3}{2} \mathbf{r}^T \mathbf{K} \mathbf{r}, \tag{1}$$

where $\mathbf{r} = (\vec{r}_0, \vec{r}_1, \vec{r}_2 \dots, \vec{r}_{N-1})$ specifies the 3D structure of the polymer chain, and \mathcal{K} is a *N*-by-*N* symmetric stiffness matrix of elements k_{ij} . The so-called Kirchhoff matrix is defined by $\mathbf{K} \equiv \mathcal{D} - \mathcal{K}$, where \mathcal{D} is a diagonal matrix with elements $\mathcal{D}_{ii} = \sum_j k_{ij}$. We assume that the probability of the chromatin to adopt a particular structure is given by

$$P(\mathbf{r}) = \frac{1}{\Xi} e^{-U_{\mathcal{K}}(\mathbf{r})/(k_B T)}, \tag{2}$$

where $k_B T$ is our energy unit (the Boltzmann constant times the temperature) and $\Xi = (\det(\mathbf{K})/(2\pi)^N)^{3/2}$ is a normalization constant such that the integration of [Eq 2](#) over all possible structures equals to 1.

The primary assumption of HLM ([Eq 1](#)) results in a probability distribution of the physical distance between any (*i, j*) monomers, r_{ij} [[50, 58](#)]

$$P(r_{ij}) = 4\pi \left(\frac{\gamma_{ij}}{\pi}\right)^{3/2} r_{ij}^2 e^{-\gamma_{ij} r_{ij}^2}, \tag{3}$$

where γ_{ij} is a function of \mathbf{K} -matrix with a positive value. More specifically, $\gamma_{ij} = (\sigma_{ii} + \sigma_{jj} - 2\sigma_{ij})^{-1}/2$ for $i > 0$ and $\gamma_{ij} = \sigma_{jj}^{-1}/2$ for $i = 0$, in which σ_{ij} denotes the (*i, j*)-th element of the inverse matrix of \mathbf{K} . To assess our assumption, we compare [Eq 3](#) with the fluorescence *in situ* hybridization (FISH) data recently reported by Takei *et al* [[24](#)]. By using DNA seqFISH+ method, they measured the 3D coordinates of 2,460 loci spaced approximately 1 Mb apart across the whole genome, together with additional 60 consecutive loci at 25 kb resolution on each chromosome, in 446 mouse embryonic stem cells ([Fig 1A](#)).

At 1 Mb resolution ([Fig 1C](#)), the distribution of the physical distances between four loci pairs on chromosome 3 (distinguished by markers of different colors) can be well fitted by [Eq 3](#). When we replot the experimental data with a scaled distance ($\sqrt{\gamma_{ij}} r_{ij}$), all rescaled data are distributed around a master curve

$$\frac{P(\sqrt{\gamma_{ij}} r_{ij})}{\sqrt{\gamma_{ij}}} = y = f(x) = \frac{4x^2}{\sqrt{\pi}} e^{-x^2}, \tag{4}$$

which is shown as the black solid line in the inset of [Fig 1C](#). Taking all 151 imaged loci on chr3 as a test case, we analyzed the distance distributions of all intra-chromosome loci pairs (i.e., $151 \times 150/2 = 11,325$ pairs), calculated γ for each pair, and plotted the rescaled data. As shown

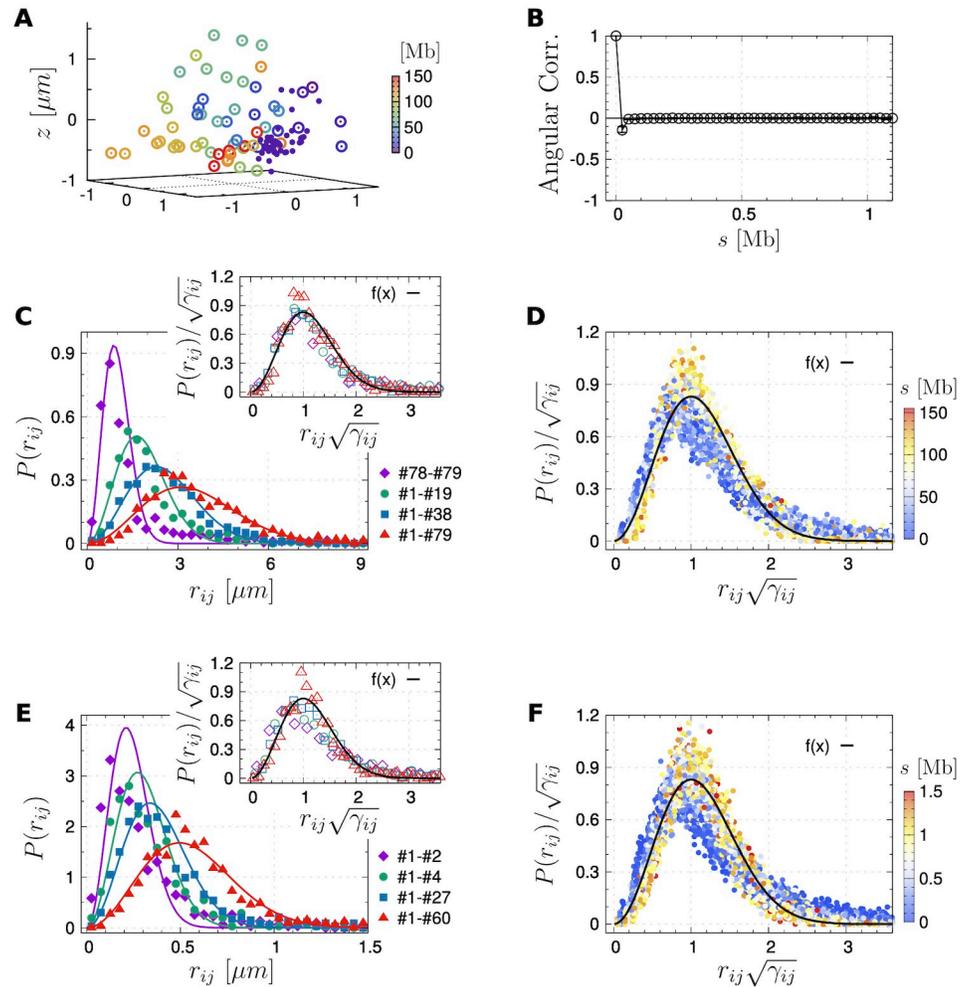


Fig 1. Chain organization of chr3 of mouse ES cells measured in the DNA seqFISH+ experiment [24]. (A) 3D positions of 151 imaged loci spaced about 1 Mb apart (large empty dots) and 60 consecutive loci with an equal space of 25 kb (small solid dots) in one of 446 sample cells. (B) Angular correlation of the chromatin segments at 25 kb resolution. (C) Distribution of the physical distance of four loci pairs at 1 Mb resolution (markers) with their fittings to our theory (lines, Eq 3), which are rescaled by using the fitting parameter (γ_{ij}) and plotted in the inset. (D) Rescaled pairwise distance distributions of all loci pairs, which are colored by their genomic lengths. (E, F) Similar as (C, D) but at 25 kb resolution.

<https://doi.org/10.1371/journal.pcbi.1009669.g001>

in Fig 1D, regardless of the genomic distance between these loci (labeled by different colors of the dots), data collected over the whole chromosome lie around the master curve, lending support to the validity of Eq 3. This also holds when analyzing the distance distribution of loci at a finer resolution of 25 kb (Fig 1E and 1F).

Next, the cross-linking probability function $F(r)$, describes how likely two chromatin fragments are captured by cross-linking agents if they are spaced by a distance of r . Based on $F(r)$, one can count the contact frequency for any monomer pair in a given structural ensemble, as well as the higher-order contact frequency among any group of n monomers ($2 < n \leq N$). The higher-order n -body contact is formed when n monomers (genomic segments) are simultaneously in spatial proximity and within a capture radius of cross-linking agent. The n -body contact probability, e.g., 3-body ($n = 3$, triplet) contact probability between i, j , and k monomer, p_{ijk} , can thus be calculated by integrating the probability of a particular chromatin

structure (Eq 2) multiplied by the chance to form simultaneous cross-linking among the n monomers in that structure over all possible chromatin structures. Thanks to the special form of the energy function (Eq 1), HLM yields analytic formulae. Specifically, when the effective capture radius of the cross-linking agent is denoted by r_c , we obtain the following results:

- The pairwise contact probability p_{ij} can be written as

$$p_{ij} = \left(1 + \frac{3}{2\gamma_{ij}r_c^2}\right)^{-3/2}, \tag{5}$$

with γ_{ij} defined in Eq 3.

- The triplet contact probability p_{ijk} is

$$p_{ijk} = \left[1 + \frac{6(a + b + c)}{r_c^2} + \frac{27z}{r_c^4}\right]^{-3/2}, \tag{6}$$

where $a = \sigma_{ii} + \sigma_{jj} - 2\sigma_{ij}$, $b = \sigma_{ij} + \sigma_{jk} - \sigma_{ik} - \sigma_{jj}$, $c = \sigma_{jj} + \sigma_{kk} - 2\sigma_{jk}$, and $z = ac - b^2$.

- The 4-body contact probability p_{ijkl} is

$$p_{ijkl} = [\det(\mathbf{W}_4)\det(\mathbf{W}_4^{-1} + \mathbf{\Lambda}_4)]^{-3/2} \tag{7}$$

where \mathbf{W}_4 is a 3×3 matrix of the form $\begin{pmatrix} a & b & f \\ b & c & d \\ f & d & e \end{pmatrix}$ with $d = \sigma_{jk} + \sigma_{kl} - \sigma_{jl} - \sigma_{kk}$, $e = \sigma_{kk} + \sigma_{ll} - 2\sigma_{kl}$, $f = \sigma_{ik} + \sigma_{jl} - \sigma_{il} - \sigma_{jk}$, and $\mathbf{\Lambda}_4 = \frac{3}{r_c^2} \begin{pmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{pmatrix}$. The parameters a, b, c are identical to those in Eq 6.

Readers are referred to the [Methods](#) section for detailed derivations. As briefly described above, these results are equivalent to reconstructing an ensemble of 3D chromatin structures followed by counting explicitly n -body contact frequencies in the ensemble.

The stiffness matrix \mathcal{K} has to be determined before calculating the n -body contact probability and predict specific multi-way chromatin interactions. As illustrated in [Fig 2](#), \mathcal{K} -matrix can be determined from $\mathcal{K}^* = \arg \min_{\mathcal{K}} L(\mathcal{K})$, i.e., by minimizing $L(\mathcal{K})$ —a cost function that quantifies the difference between the \mathcal{K} -dependent pairwise contact probabilities of the model ($p_{ij}(\mathcal{K})$) and those from pairwise contacts experiment ($p_{ij}^{\text{Hi-C}}$) (see the [Methods](#) section for the details). We next compare our predictions of many-body chromatin contacts with those from measurements found in the literature. Detailed information on which genomic region are modeled, the cell types, the model resolutions and the experiment sources are provided in Table A in [S1 Appendix](#).

Comparison with Tri-C

Our first case study for our HLM-based theoretical predictions (see [Eq 6](#)) concerns the mouse α -globin locus, an extensively studied model system [61], whose three-body contacts are available from Tri-C measurement [17]. The globin genes are flanked by several CTCF-binding sites (the 3rd row in [Fig 3A](#)) in both erythroid and embryonic stem (ES) cells. The Capture-C heatmaps of the two cell types ([Fig 3B](#)). The top and bottom panels are for the ES and erythroid

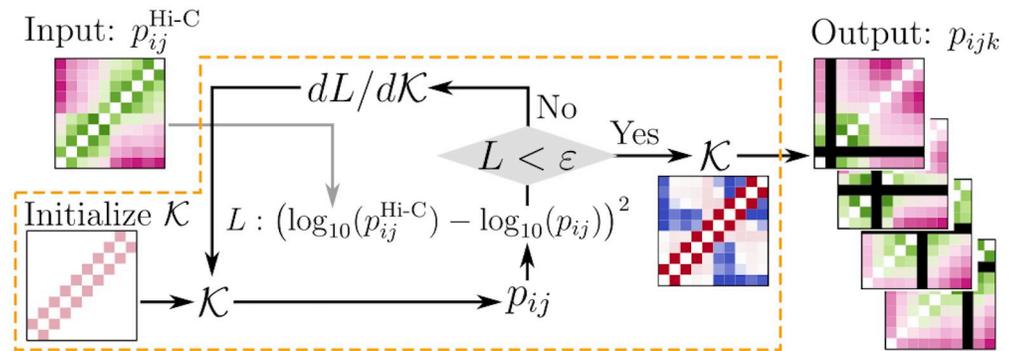


Fig 2. Flow chart of HLM that is enclosed by the dashed orange box. It takes two-body contacts from Hi-C as input, and calculates n -body ($n > 2$) contact probability. The stiffness matrix \mathcal{K} is updated until the algorithm finds \mathcal{K}^* that minimizes the cost function $L = L(\mathcal{K})$ which quantifies the difference between pairwise contact probabilities of Hi-C data and those determined from \mathcal{K} matrix.

<https://doi.org/10.1371/journal.pcbi.1009669.g002>

cells, respectively) show that the erythroid cell display more frequent contacts (stronger interactions) between the α -globin genes (Hba-a1/2) and the five upstream enhancer elements (R1, R2, R3, R4 and Rm).

The \mathcal{K} -matrices of HLM, obtained separately for both cell lines based on Hi-C data, are not only used to calculate p_{ij} that captures the variation of local pairwise contacts upon ES-to-erythroid cell differentiation (Fig 3B), but also to predict triplet contacts (p_{ijk}) at different genomic locus in a cell-type specific manner (Fig 3C and 3D). We calculated the two maps of triplet contact probability (i) at the genomic locus of the strongest enhancer R2 (Fig 3C) and (ii) at the upstream CTCF-binding site HS-39 (Fig 3D). The Pearson correlations (PC) between the HLM-predicted triplet contacts and those from Tri-C data are generally high (PC \approx 0.75 – 0.89). See the caption of Fig 3 for details). In particular, for the triplet contact probability map calculated at the HS-39 viewpoint, the correlations are comparable to those reported by a simulation study using the SBS model (PC \approx 0.8) [56]. We further calculated three correlation coefficients specifically designed to compare contact matrices, namely the distance-corrected Pearson correlation [46], stratum-adjusted correlation [62] and stratified Pearson correlation (S1 Fig). As summarized in Table B in S1 Appendix, overall, HLM matches better with Tri-C experiment than SBS model.

The comparison between triplet contact probabilities in ES cells over all sites of interest (SOIs) and those in erythroid cells underscores the change in higher-order contacts upon cell differentiation (Fig 3E). Upon the activation of α -globin gene (ES \rightarrow erythroid cell), R2 simultaneously interacts with the gene promoters and its nearby enhancers especially R1 more frequently, which suggests the formation of a regulatory hub. HS-39, HS-38 and distal downstream CTCF binding sites are not involved in this hub, but form diffuse interactions in between. These findings from the HLM-predicted triplet contacts are consistent with the Tri-C analyses conducted at R2 and HS-39 [17] (see S2 Fig).

We made additional comparisons at the globin gene promoter Hba-a1 and the downstream CTCF boundary element HS+44, whose triplet contact probabilities have not yet been measured (Fig 3F). At the viewpoint Hba-a1, the promoter interacts with upstream enhancers in a cooperative manner, which supports the above analyses. By contrast, in the presence of contact with a downstream boundary element, Hba-a1 becomes less likely to interact with another downstream boundary site in erythroid cells.

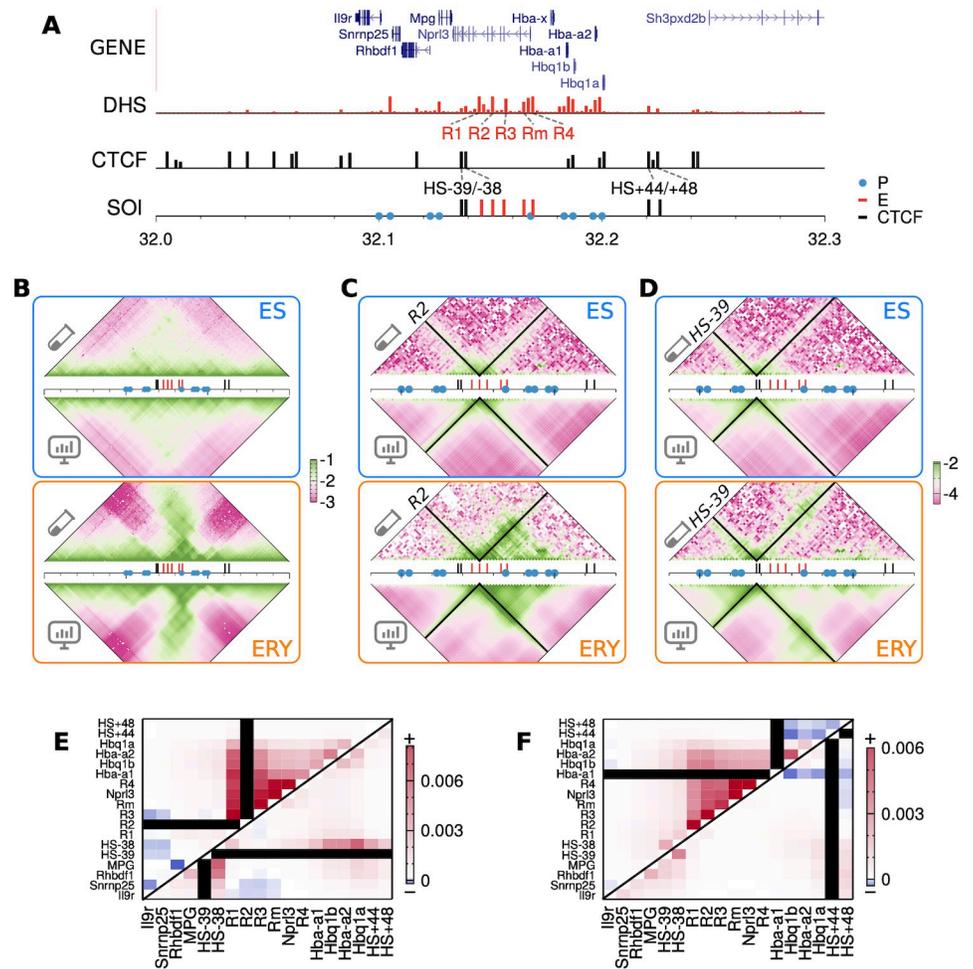


Fig 3. Three-body contacts predicted with HLM versus those from Tri-C. (A) Gene notation, DNase hypersensitive sites, CTCF-binding sites [59, 60] and genomic regions of interest (sites of interest, SOI) at mouse α -globin locus on chr11. (B) $\log_{10}(p_{ij})$ from Capture-C (top triangle) and HLM (bottom triangle) at 2-kb resolution. (C, D) $\log_{10}(p_{ijk})$ from Tri-C (top) and HLM (bottom) at $k = R2$ and $k = HS-39$, whose positions are labeled with black strips. The Pearson correlations between Tri-C and HLM are 0.80 (ES) and 0.75 (erythroid) for $k = R2$, and 0.85 (erythroid) for $k = HS-39$. (E) Cell-line difference of the predicted three-body contacts, $p_{ijk}^{\text{Erythroid}} - p_{ijk}^{\text{ES}}$, among the SOIs with respect to the viewpoints R2, HS-39 and (F) Hba-a1, HS+44.

<https://doi.org/10.1371/journal.pcbi.1009669.g003>

Comparison with multi-contact 4C sequencing (MC-4C)

Next, we study the formation of higher-order contacts among CTCF binding sites in the cohesin release factor (WAPL) lacking cells [63]. According to the loop extrusion model [64–66], the ring-shaped protein complex, cohesin, binds to DNA and progressively enlarges chromatin loop until the dynamics is hindered by two convergently orientated CTCF-bound sites. Knockout of the WAPL promotes the chromatin loop extension [63].

We carried out HLM calculation for a 1.28-Mb genomic region on chromosome 8, which contains 13 CTCF-binding sites, for both wild-type (WT) and WAPL-deficient (Δ WAPL) human chronic myeloid leukemia (HAP1) cells, whose triplet contacts were measured with MC-4C [18]. Although the Hi-C dataset (Fig 4A, top) is noisier than that used in the first case study (Fig 3B, top), the HLM-predicted triplet contacts at the CTCF binding sites E (Fig 4B)

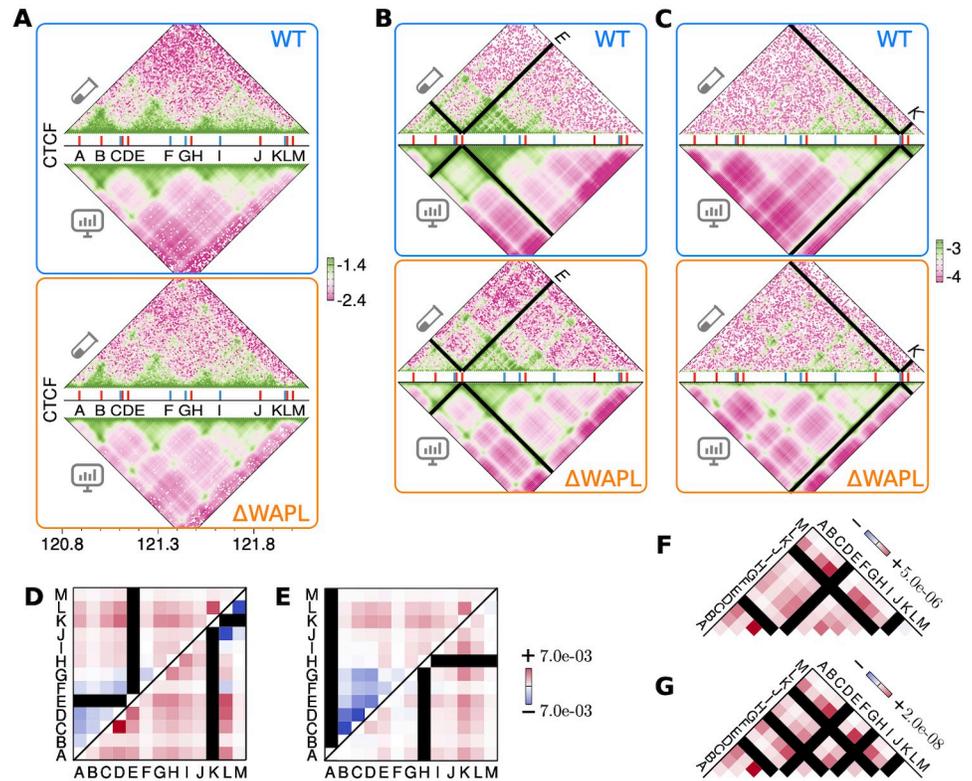


Fig 4. Comparing many-body contact predictions with MC-4C. (A) Comparing $\log_{10}(p_{ij})$ from Hi-C and HLM at 10-kb resolution of a 1.28 Mb region on chr8, where there are thirteen CTCF-binding sites labeled from A to M. The genomic position of forward- and backward-oriented CTCF sites are marked with red and blue sticks, respectively. (B, C) Comparing $\log_{10}(p_{ijk})$ from MC-4C and HLM at a viewpoint of site E and K, respectively. From the viewpoint of E (K), the values of PC between the model and the MC-4C experiment are 0.83 (0.81) and 0.71 (0.73) for the WT and Δ WAPL cells, respectively. (D) Cell-line difference of the predicted three-body contacts, namely $p_{ijk}^{\Delta WAPL} - p_{ijk}^{WT}$, among the CTCF-binding sites from the viewpoint of A, E and (E) H, K, respectively. (F, G) Cell-line difference of four- and five-body contacts from double- and triple-anchored viewpoints, whose positions are marked with the black strips. The numbers label the maximum amplitudes of changes.

<https://doi.org/10.1371/journal.pcbi.1009669.g004>

and K (Fig 4C) show patterns similar to those from the MC-4C measurement [18]. To account for the enrichment of long-range triplet contacts comprised of distal CTCF binding elements (e.g., the triplet A, H and K) in the WAPL-lacking cells, Allahyar *et al.* [18] put forward a “traffic jam” of CTCF roadblock-trapped cohesin complexes.

Examining the WAPL knockout-induced change in triplet contacts among CTCF binding sites, we plot the changes of p_{ijk} ($\Delta p_{ijk} = p_{ijk}^{\Delta WAPL} - p_{ijk}^{WT}$) at four viewpoints, $k = E, K$ and A, H in Fig 4D and 4E, respectively. While the site A is colocalized more frequently with the distant downstream sites K and L in Δ WAPL cell, its interaction with the nearby loci (sites B-G) is reduced significantly (e.g., p_{ABC} is reduced by more than 60% from 1.1×10^{-3} to 4.2×10^{-4}). Similar patterns of “long-range enrichment and short-range depletion of triplet contacts” are identified for those from other upstream and downstream viewpoints (e.g., the sites E and K). Although the MC-4C measurement [18] did not highlight the short-range depletion of triplet contacts, our finding is consistent with the contact frequencies measured with respect to the sites E and K (S3(A) Fig). In fact, the two Hi-C datasets, one for WT and the other for Δ WAPL cell, show that the pairwise contact frequencies of Δ WAPL cells in the domains flanked by short-range CTCF binding sites are smaller than those of the WT (S3(B) Fig), which directly

confirm the short-range depletion of contacts. The observation of long-range enrichment and short-range depletion of triplet contacts does not always hold; the central sites H and I are the two exceptions that do not demonstrate depletions of short-range triplet contacts.

The aggregation of cohesin molecules has been identified in Δ WAPL cells with super-resolution imaging [18]; however, it has not been known whether or not a similar type of aggregation occurs in CTCF-bound chromatin sites. To explore the possibility of condensation of CTCF-bound chromatin sites, we calculate higher-order chromatin contacts in both conditions of WT and Δ WAPL cells. The changes in 4-body and 5-body contacts among CTCF-bound sites from WT to Δ WAPL cells from double- and triple-anchored view points are all positive, namely, $p_{ijkl}^{\Delta WAPL} - p_{ijkl}^{WT} > 0$ (Fig 4F) and $p_{ijklm}^{\Delta WAPL} - p_{ijklm}^{WT} > 0$ (Fig 4G), which lends supports to the picture surmising the clustering of domain boundary in Δ WAPL cells. Although their probabilities are small, up to 4-fold increases of 4-body and 5-body contacts are identified among the SOIs (see S4 Fig).

Comparison with SPRITE

Lastly, we model a 0.45 Mb genomic region of GM12878 cell (Fig 5A and 5B), which contains transcriptionally active genes BCL2 and multiple super-enhancers, annotated in Ref. [30] by Perez-Rathke *et al.*, and compared the predicted triplet contacts with the results from SPRITE

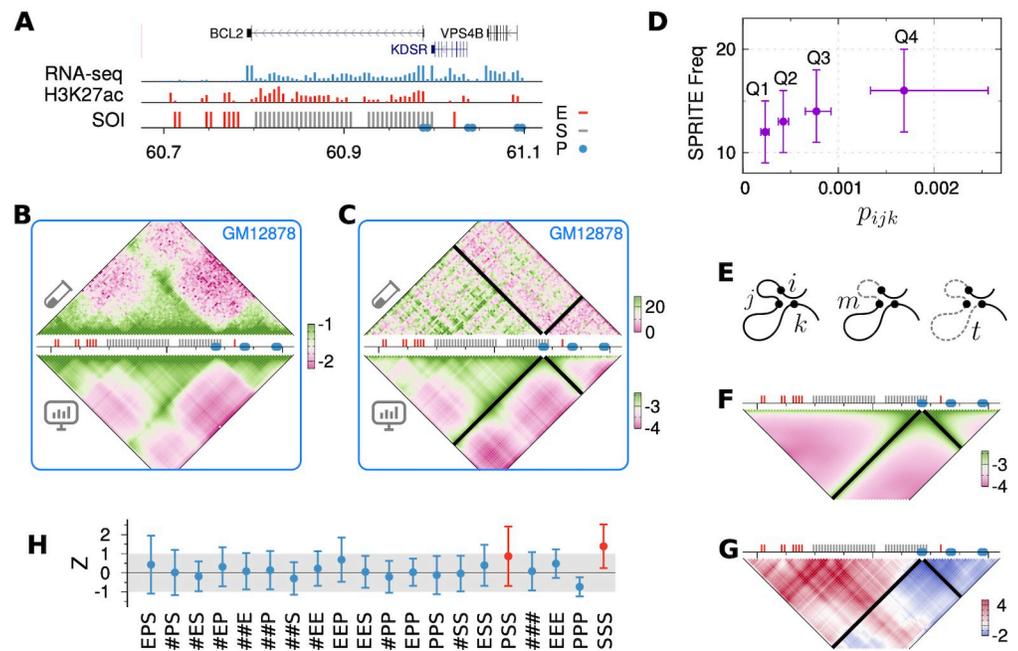


Fig 5. Comparing triplet contact predictions with SPRITE. (A) Genes, RNA-seq and H3K27ac ChIP-seq signals [60, 67] in a 0.45 Mb region on human chr18 with the positions of three types of SOIs (Enhancers, Promoters and Super-enhancers). (B) $\log_{10}(p_{ij})$ from Hi-C compared with that from HLM at 5-kb resolution. (C) The triplet contact frequency from SPRITE is compared with $\log_{10}(p_{ijk})$ from HLM anchored at an active promoter site. (D) SPRITE frequency of triplets which are divided into four quantiles based on ascending order of p_{ijk} . (E) For three sites, $i < j < k$, along a polymer chain, the minor and major sections, denoted by m and t , respectively, are depicted with dashed lines. (F) The expected three-body contact probability p_{ijk} , and (G) the specificity Z-score $Z_{ijk} \equiv (p_{ijk} - \bar{p}_{ijk}) / \sigma_{ijk}$. (H) The mean specificity Z-score of triplet contacts with respect to all possible combinations of annotations (E: Enhancers, P: Promoters, S: Super-enhancers, #: sites without any annotation). The error bars are the standard deviations.

<https://doi.org/10.1371/journal.pcbi.1009669.g005>

[26]. The human BCL2 gene encodes a membrane protein which blocks the apoptotic death in B-lymphoblastoid cells.

We expect that the similarity between the model and the experiment is not so high as the foregoing two case studies, because the data from SPRITE, designed for a genome-wide study, is sparse with 5-kb resolution, and the chromatin fragments cocaptured by SPRITE are not necessarily in spatial proximity since SPRITE does not distinguish between direct and indirect cross-linkings. Indeed, a weak correlation is found in the triplet contacts anchored at the active BCL2 gene promoter (PC = 0.18; see Fig 5C). To validate our models against the measurements, we sort the triplets in an ascending order of their predicted contact probabilities, and divide them into four quantiles. Then the frequency of the quantilized triplets in SPRITE increases with the predicted probability, p_{ijk} (Fig 5D).

To identify the specificity in a single cell type, we defined an expected triplet contact probability \bar{p}_{ijk} ($i < j < k$) as the mean contact probability of all triplets in the “sections” of interest along the chain (Fig 5E). The minor and major sections are determined as $m = \min(|j - i|, |k - j|)$ and $t = |k - i|$, respectively. \bar{p}_{ijk} depends only on m and t (Fig 5F), which reflects the chain connectivity and the global compactness of chromatin. The ratio of the difference between the observed and expected probabilities scaled by the standard deviation is then defined as the *specificity Z-score*, $Z_{ijk} (\equiv (p_{ijk} - \bar{p}_{ijk})/\sigma_{ijk})$, for each triplet (Fig 5G). The triplet contacts are deemed specific if $Z_{ijk} > 1$.

Z_{ijk} averaged over all possible combinations of functional annotations in the region of interest are shown in Fig 5H. The concurrent triplet contacts between two super-enhancers with a promoter (PSS) or with a third super-enhancer (SSS) are specifically enriched at this loci. This finding comports well with the Perez-Rathke *et al*'s finding that multiple super-enhancers are involved in higher-order chromatin contacts more frequently than other elements [30]. We confirm similar results from modeling at another active locus (S5 Fig).

Discussions

Relation between pairwise and triplet contacts

For a given polymer chain, is there any simple relation to associate two-body contacts with three-body contacts? Through both experimental and theoretical studies, there has been much effort to address this question and related issues in the context of concurrent chromatin interactions [14, 15, 18, 19, 68].

Here, we address this issue in a principled way by considering a Gaussian polymer chain without any specific interaction as a reference system. In the heatmaps of n -body contact probability of the Gaussian polymer chain consisting of N monomers, the enrichments of many-body contacts are observed near the diagonal part of $N \times N$ matrix as well as around the viewpoints (Fig 6A–6C). For any triplet, ijk ($i < j < k$), the contact probability reads (see Eq 26 for the derivation)

$$p_{ijk} = p(m, t) = \left[1 + \frac{6t}{\gamma} + \frac{27m(t-m)}{\gamma^2} \right]^{-3/2}, \quad (8)$$

where $m = \min(|j - i|, |k - j|)$ is the minor section, $t = |k - i|$ is the major section of the triplet (Fig 5E), and γ is the stiffness constant which restrains the consecutive sites. For $t \gg m$, it follows that $p_{ijk} \sim t^{-3/2}$, the scaling exponent of which is identical to that of the two-body contact probability [28, 29]. For a fixed t , $p(m, t)$ changes non-monotonically as a function of m , having its minimum and maximum at $m = t/2$ and $m = 1$, respectively (Fig 6B). According to Eq 8

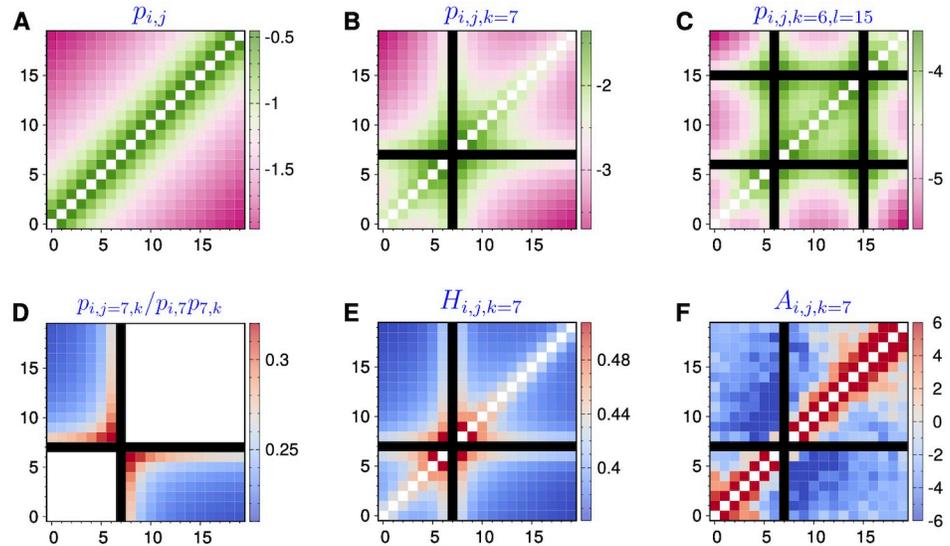


Fig 6. Many-body contacts in a Gaussian polymer chain consisting of 20 sites. (A) Two-body contact probability $\log_{10}(p_{ij})$. (B) Three-body contact probability $\log_{10}(p_{ijk})$ anchored at the 7-th site ($k = 7$). (C) Four-body contact probability $\log_{10}(p_{ijkl})$ double anchored at the 6-th and 15-th sites ($k = 6, l = 15$). (D-F) Boost factor $p_{ijk}/p_{ij}p_{jk}$, hub-score H_{ijk} , and association z-score A_{ijk} from the viewpoint of the 7-th site.

<https://doi.org/10.1371/journal.pcbi.1009669.g006>

the specificity Z-score defined in the subsection **Comparison with SPRITE** satisfies $Z_{ijk} = 0$ for all triplets in a Gaussian chain, as anticipated.

An enhancement of triplet contact compared to two independent pairwise loops was discussed by Polovnikov *et al.* [68], who used the delta function for the cross-linking probability between genomic sites, such that sites were considered in contact only if they overlapped with each other. It was found that $p_{ijk}/p_{ij}p_{jk} \geq 1$ with a maximum at $m = t/2$. This finding, however, is not universally held if one considers other types of cross-linking probability. When Gaussian and Heaviside step functions are used as the cross-linking probabilities, triplet contact probabilities are always suppressed in comparison with the product of two binary contact probabilities, i.e., $p_{ijk}/p_{ij}p_{jk} \leq 1$, and the largest suppression occurs at $m = t/2$ (Fig 6D and S6(D) Fig).

No clear-cut relation exists between pairwise (p_{ij}) and triplet contact probabilities (p_{ijk}), which however was the underlying presumption of the data processing in several experiments. For example, a hub score was calculated to identify synergistic interaction hubs in the analysis of 3way-4C data [15], which was defined by $H_{ijk} = 3p_{ijk}/(p_{ij}p_{jk} + p_{ij}p_{ik} + p_{jk}p_{ik})$ approximately, whereas our Gaussian chain model clarifies that the hub score is not a constant (Fig 6E). In fact, as discussed in the experiment [15], enrichment of the score is always found at the triplets close to any viewpoint (e.g., Figs 3 and 4 in Ref. [15]).

To detect cooperative (competitive) contacts, an association Z-score, defined as $A_{ijk} = (C_{ijk}^+ - \mathbb{E}(C_{ijk}^-))/\sigma(C_{ijk}^-)$, were computed in the analyses of MC-4C measurements [18, 19]. C_{ijk}^+ is the number of reads cocapturing the genetic sites i, j and k (the viewpoint). $\mathbb{E}(C_{ijk}^-)$ and $\sigma(C_{ijk}^-)$ are the mean and standard deviation of C_{ijk}^- , the number of reads capturing the sites j and k but not i . A positive (negative) A_{ijk} , which implies that the chance of site j being in association with k and any other third site is (dis) favored when site i is interacting with k , was interpreted as a measure of cooperative (competitive) triplet contacts. However, the calculation based on Gaussian polymer chain shows that adjacent monomer pairs along the chain display

a greater association z-scores A_{ijk} (Fig 6F). This issue of false-positiveness in the short-range cooperativity among the multiple sites has been noticed in Ref. [18].

A recent super-resolution imaging study on chromatin fibers in IMR90 cells has explored whether or not the association between two genetic loci facilitates the contact with a third site [14]. For triplets ijk satisfying $i < j < k$, two conditional probabilities $p_{jk|r_{ij} \leq r_c}$ and $p_{jk|r_{ij} > r_c}$ ($p_{jk|r_{ij} \leq r_c}$ is the contact probability between the k -th and j -th sites under the condition that i -th and j -th sites are in contact, and $p_{jk|r_{ij} > r_c}$ is similarly defined) were compared with the unconditioned contact probability p_{jk} . It was found that regardless of cell types and other chemical treatments, a relation $p_{jk|r_{ij} > r_c} < p_{jk} < p_{jk|r_{ij} \leq r_c}$ was always satisfied among three probabilities, for most triplets of genomic loci [14]. Our theoretical framework allows us to validate this relation for any triplet in a polymer chain (see S7 Fig). The generic characteristics of multi-body contacts of Gaussian chain mentioned above do not depend on the specific form of cross-linking probability (S6 Fig).

As a further test, we counted the average, maximum, and minimum pairwise contact probabilities of the triplets anchored at R2 in the α -globin locus of mouse erythroid cells (S8(B)–S8(D) Fig). They have dissimilar patterns with respect to the results from Tri-C experiment (S8(A) Fig), which is also reflected by their overall lower Pearson correlations than HLM calculated at two different viewpoints in two different cell types (S8E Fig). The relation between pairwise and multi-way contacts is more complicated than intuition.

Miscellaneous

The prediction of HLM is not always consistent with experiments. For example, when scrutinizing the WAPL knockout-induced change of triplet contacts at viewpoint of the CTCF binding site E, an enrichment involving site G can be found in MC-4C (see the left panel of S3 Fig) which is not in agreement with our model prediction (Fig 4D). Many factors can contribute to this discordance, such as the quality of the Hi-C dataset at the locus of interest. Despite much effort devoted in processing raw Hi-C data, bias which remains in the Hi-C contact matrix after normalization will propagate to the model and subsequent predictions on higher-order contacts.

Another apparent missing component of our approach is the excluded volume interaction, which plays an indisputable role in determining polymer behavior. However, as shown in Fig 1, chromatin segment pairs have non-zero probabilities at small physical distances, in contrast to the expectation that segments are not allowed to overlap with each other. We explain the absence of excluded volume by using a polymer melt, which contains 125 polymer chains each composed of 641 monomers (S9 Fig). The system was simulated considering only chain connectivity and Weeks-Chandler-Anderson type excluded volume interactions [69]. The latter manifests itself as the zero-valued probability of the bond length ($r_{i,i+1}$) and intra-chain pairwise monomer distance (r_{ij}) when $r < 1$ in units of the monomer diameter (the circular black dots in S9(C) and S9(D) Fig). However, when the melt is coarse-grained such that we model 8 or 64 consecutive segments as a coarse-grained center, the coarse-grained segments overlap with each other. Together with the results shown in Fig 1B, it is justified to ignore the excluded volume and bending penalty beyond certain scales.

As a further test of the general applicability of HLM for the prediction of n -body contacts, we modeled the β -globin locus of mouse ES cells (S10 Fig) and the Pcdh α locus of mouse neural progenitor cells (S11 Fig) by using the Hi-C data reported by Bonev *et al* [70]. The pairwise and predicted triplet contact probabilities at multiple viewpoints are well correlated with the experiments (see detailed PC coefficient in the figure caption and Table A in S1 Appendix).

Conclusions

To recapitulate, we have derived analytic expressions of n -body contact probabilities (for any $n > 2$) based on a recent chromatin polymer model (HLM) [50] and developed a method to predict multi-way chromatin contacts from Hi-C data.

First, the predicted triplet chromatin contacts are in reasonable correlation with the results of two independent measurements (i) Tri-C [17] and (ii) MC-4C [18], also in accordance with the genome-wide study using (iii) SPRITE [26]. Besides confirming the experimental findings, the suggested method was used to explore the multi-way chromatin contacts at any viewpoint of genomic locus of interest, which allowed us to discover some of key features not previously underscored in each measurement. (i) For the mouse α -globin locus, a cell-line dependent interaction pattern for the promoter is found when the viewpoint is anchored at the gene promoter. (ii) Although the previous study highlighted only the enrichment of long-range triplet contacts among CTCF binding sites in the Δ WAPL cells, we find that depletion of short-range contacts occurs for some triplets to compensate the enrichment. Our calculations also lend support to the aggregates of CTCF-bound boundary elements by explicitly showing the enrichments of four- and five-way chromatin contacts. (iii) Lastly, our analysis captures the enrichment of triplet contacts involving super-enhancers at transcriptionally active loci.

With an increasing contact order n , the probability of forming multi-way contacts decreases by orders of magnitude (see how the range of scale bars changes with increasing n in Fig 6A–6C). Our theoretical approach can be used to circumvent this statistical limitation inherent to experimental detection of multi-way chromatin contacts. All the computer codes discussed here are provided in <https://github.com/leiliu2015/HLM-Nbody>, so that multi-way contacts can be calculated from an input Hi-C dataset. The methodology developed here will be of great help to elucidate the regulatory roles played by complex chromatin topology.

Methods

Numerical details for determining the stiffness matrix \mathcal{K}^*

Our previous numerical procedure [50, 51] to determine \mathcal{K}^* -matrix has been improved in this paper by adapting those in the recent modeling studies of chromatin by two other groups [71–73]. The best solution, \mathcal{K}^* , which is consistent with a given Hi-C data, was determined with Hi-C data by optimizing the cost function, $\mathcal{K}^* = \arg \min_{\mathcal{K}} L(\mathcal{K})$ (see Fig 2). Different forms of the cost function can be conceived:

$$L_0(\mathcal{K}) = \sum_{i < j} (p_{ij}(\mathcal{K}) - p_{ij}^{\text{Hi-C}})^2 \quad (9a)$$

$$L_1(\mathcal{K}) = \sum_{i < j} (\log_{10}(p_{ij}(\mathcal{K})) - \log_{10}(p_{ij}^{\text{Hi-C}}))^2 \quad (9b)$$

$$L_2(\mathcal{K}) = 1 - \text{PC}(\log_{10}(p_{ij}(\mathcal{K})), \log_{10}(p_{ij}^{\text{Hi-C}})) \quad (9c)$$

$$L_3(\mathcal{K}) = \frac{2}{N(N-1)} \sum_{i < j} \left(\frac{\langle r_{ij}^2(\mathcal{K}) \rangle}{\langle r_{ij}^2 \rangle_{\text{Hi-C}}} - 1 \right)^2 \quad (9d)$$

where $p_{ij}^{\text{Hi-C}}$ is the pairwise contact probability observed in Hi-C experiments, $\text{PC}(x, y)$ stands for the Pearson correlation coefficient between x and y , and $\langle r_{ij}^2 \rangle_{\text{Hi-C}} = 3/(2\gamma_{ij}^{\text{Hi-C}})$ can be derived from $p_{ij}^{\text{Hi-C}}$ based on either Eq 12 for Gaussian contact probability or Eq. S1 for contact probability in the form of Heaviside step function.

To avoid non-physical negative eigenvalues of the Kirchhoff matrix, the interaction strengths were previously required to be non-negative ($k_{ij} \geq 0$) during the minimization [50, 71]. Shinkai *et al.* have overcome the issue of negative k_{ij} by alternately updating the backbone and non-backbone interaction strengths [72]. Even though there are some pairs with $k_{ij} < 0$, as long as the resulting \mathcal{K} -matrix is positive-definite [72], the potential of mean force between the i - and j -th monomer is proportional to $-\ln P(r_{ij})$, which has a physically meaningful minimum at $r_{ij} = 1/\sqrt{\gamma_{ij}}$.

We compared four different methods to optimize the cost functions: the first method minimizes the cost by random sampling (RS) of the parameter space [72]; the second one updates \mathcal{K} via steepest gradient descent (GD) [71]; the third and fourth methods, RMSprop [74] and ADAM [75], calculate both the gradient and its second moment to accelerate the model training [76].

As a case study, the decay of the cost function during the model training of a 2.4-Mb region on chr8 in mouse embryonic stem cells is shown in S12(A) Fig. The ADAM optimizer outperforms other methods significantly. Quality of the final model was calibrated by using the stratum adjusted correlation (SCC) [62] and PC between $\log_{10}(p_{ij})$ from Hi-C and that from the model. As shown in the accompanying Table B in S1 Appendix, while L_1 and L_3 are both good candidates, L_2 seems to be the best choice. However, the dependence of contact probability on the subchain size s , defined as $P(s) = \sum_{i=0}^{i=N-s-1} p_{i,i+s}/(N-s)$, shows that the optimal model based on minimizing L_2 underestimates the overall contacts (S12(B) Fig), although the pattern of $\log_{10}(p_{ij})$ from the model is highly correlated with that from Hi-C. Similar conclusions can be drawn from training a model of a 10-Mb region on chr5 in GM12878 cells (S13 Fig), which our previous work used as a test case [50]. The models trained with new schemes clearly achieve better quality (S13(D) Fig).

Taken together, in this study we calculated the \mathcal{K} -matrix based on Hi-C data by minimizing L_1 with F_0 , or minimizing L_3 with F_1 using ADAM optimizer, where F_0 and F_1 are two possible functional forms of cross-linking probability explained below.

Derivations of the multi-body contact probability

Despite many potential biases in Hi-C [13], the frequency of two chromatin fragments being cross-linked in millions of cells is ideally determined by the probability density of their spatial distance $P(r)$, and the efficiency of the cross-linking agent. The latter contribution can be included by the r -dependent cross-linking probability of fragments $F(r)$. One can consider using a Gaussian function, $F_0(r) = e^{-3r^2/2r_c^2}$, for cross-linking probability, which makes many-body contacts easier to handle mathematically. Alternatively, the Heaviside step function $F_1(r) = \Theta(r_c - r)$ is conceived as well, as a natural choice of the cross-linking probability, such that fragment pairs are cross-linked if the spatial distance r is smaller than an effective capture radius of the cross-linking agent (r_c).

The contact probabilities based on the two cross-linking probabilities, $F_0(r)$ and $F_1(r)$ are denoted by “ $p^{(0)}$ ” and “ $p^{(1)}$ ”, respectively. Unless stated otherwise, the results shown in the main text were calculated with $F_0(r)$, and “ p ” was used as a shorthand notation of “ $p^{(0)}$ ” for brevity.

Here we present the derivation of n -body contact probability based on the Gaussian cross-linking probability, $F_0(r)$. The derivation based on the Heaviside step function, $F_1(r)$, is provided in S1 Appendix.

Generic many-body contact probability. Based on Eqs 1 and 2, if we set $\vec{r}_0 = \mathbf{0}$, the many-body contact probability among n (≥ 2) monomers in a given set $\mathcal{A} = (i, j, \dots)$ can be

directly calculated as

$$\begin{aligned}
 p_n^{(0)} &= \frac{1}{\Xi} \int D\mathbf{r} P(\mathbf{r}) \prod_{(i,j) \in \mathcal{M}} F_0(\|\vec{r}_i - \vec{r}_j\|) \\
 &= \left[\frac{1}{\Xi_x} \int D\mathbf{x} e^{-\frac{1}{2}\mathbf{x}^T \mathbf{K}_0 \mathbf{x}} \prod_{(i,j) \in \mathcal{M}} e^{-\frac{3}{2r_c^2}(x_i - x_j)^2} \right]^3 \\
 &= \left[\frac{1}{\Xi_x} \int D\mathbf{x} e^{-\frac{1}{2}\mathbf{x}^T (\mathbf{K}_0 + \Delta_0) \mathbf{x}} \right]^3 \\
 &= \left[\frac{\det(\mathbf{K}_0 + \Delta_0)}{\det(\mathbf{K}_0)} \right]^{-3/2},
 \end{aligned} \tag{10}$$

where $D\mathbf{r} \equiv \prod_i d\mathbf{r}_i$, $P(\mathbf{r})$ is the probability density of the chain configuration (\mathbf{r}) and \mathcal{M} is the enumeration of all the pair-wise combinations of the elements in \mathcal{A} . Δ is a matrix of elements Δ_{uv} whose values are given by

$$\Delta_{uv} = \frac{3}{r_c^2} \times \begin{cases} n - 1, & u = v \text{ and } u \in \mathcal{A} \\ -1, & u \neq v \text{ and } u, v \in \mathcal{A}, \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

in which $u, v \in (0, 1, 2, \dots, N - 1)$. The \mathbf{K} and Δ with subscript “0” in Eq 10 signifies the matrices whose 0-th row and column are removed from the original ones.

Pairwise contact probability. To avoid computing determinants of large matrices in the general solution (Eqs 10 and 11), p_n can alternatively be calculated based on n -body correlation function. For example, considering the probability density of the pairwise distance between the i -th and j -th monomer in three-dimensional space (Eq 3), the pairwise contact probability between the i -th and j -th monomers can be formulated as [71, 72]

$$\begin{aligned}
 p_{ij}^{(0)} &= \int_0^\infty P(r_{ij}) F_0(r_{ij}) dr_{ij} = \int_0^\infty P(r_{ij}) e^{-3r_{ij}^2/2r_c^2} dr_{ij} \\
 &= \left(1 + \frac{3}{2\gamma_{ij} r_c^2} \right)^{-3/2} = \left(1 + \frac{\langle r_{ij}^2 \rangle}{r_c^2} \right)^{-3/2},
 \end{aligned} \tag{12}$$

where Eq 3 and $\langle r_{ij}^2 \rangle = 3/(2\gamma_{ij})$ is used in the second row.

Three-body contact probability. For three-body contact probability p_{ijk} , we first consider the probability density of the distances between the i -th and j -th monomers, and between the j -th and k -th monomers projected on one dimension

$$\begin{aligned}
 P(x_{ij}, x_{jk}) &= \langle \delta[x_{ij} - (x_i - x_j)] \delta[x_{jk} - (x_j - x_k)] \rangle \\
 &= \int d\mathbf{x} \delta[x_{ij} - (x_i - x_j)] \delta[x_{jk} - (x_j - x_k)] P(\mathbf{x}) \\
 &\propto \int d\mathbf{k} e^{ik_1 x_{ij}} e^{ik_2 x_{jk}} \int d\mathbf{x} e^{-ik_1(x_i - x_j)} e^{-ik_2(x_j - x_k)} P(\mathbf{x}) \\
 &\propto \int d\mathbf{k} e^{ik_1 x_{ij}} e^{ik_2 x_{jk}} e^{-\frac{1}{2}\mathbf{k}^T \mathbf{W}_3 \mathbf{k}} \\
 &= \frac{1}{2\pi\sqrt{z}} e^{-\frac{1}{2z}(cx_{ij}^2 - 2bx_{ij}x_{jk} + ax_{jk}^2)},
 \end{aligned} \tag{13}$$

where $\mathbf{k} = (k_1, k_2)$, $\mathbf{W}_3 = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ is a matrix with elements

$$\begin{aligned} a &= \sigma_{ii} + \sigma_{jj} - 2\sigma_{ij} \\ b &= \sigma_{ij} + \sigma_{jk} - \sigma_{ik} - \sigma_{jj} \\ c &= \sigma_{jj} + \sigma_{kk} - 2\sigma_{jk}, \end{aligned} \tag{14}$$

and

$$z = \det(\mathbf{W}_3) = ac - b^2. \tag{15}$$

Remember that σ is the element of \mathbf{K}^{-1} , and $\sigma_{ij} = 0$ if $i \times j = 0$. Then, the three-body contact probability in 3D is given by

$$\begin{aligned} p_{ijk}^{(0)} &= \left[\int \int dx_{ij} dx_{jk} P(x_{ij}, x_{jk}) F_0(x_{ij}) F_0(x_{jk}) F_0(x_{ij} + x_{jk}) \right]^3 \\ &= \left[1 + \frac{6(a + b + c)}{r_c^2} + \frac{27z}{r_c^4} \right]^{-3/2}. \end{aligned} \tag{16}$$

Four-body contact probability. Four-body contact probability p_{ijkl} can be derived in the same manner. By defining $\mathbf{u}_x = (x_{ij}, x_{jk}, x_{kl})^T$, the one dimensional probability density of three pairwise distances, $P(x_{ij}, x_{jk}, x_{kl})$, is given by

$$\begin{aligned} &\langle \delta[x_{ij} - (x_i - x_j)] \delta[x_{jk} - (x_j - x_k)] \delta[x_{kl} - (x_k - x_l)] \rangle \\ &\propto \int d\mathbf{k} e^{ik_1 x_{ij}} e^{ik_2 x_{jk}} e^{ik_3 x_{kl}} e^{-\frac{1}{2} \mathbf{k}^T \mathbf{W}_4 \mathbf{k}} \\ &= \frac{1}{(2\pi)^{3/2} \sqrt{\det(\mathbf{W}_4)}} e^{-\frac{1}{2} \mathbf{u}_x^T (\mathbf{W}_4)^{-1} \mathbf{u}_x}, \end{aligned} \tag{17}$$

where now $\mathbf{k} = (k_1, k_2, k_3)$. \mathbf{W}_4 has a matrix form of

$$\mathbf{W}_4 = \begin{pmatrix} a & b & f \\ b & c & d \\ f & d & e \end{pmatrix}, \tag{18}$$

with elements in Eq 14 and

$$\begin{aligned} d &= \sigma_{jk} + \sigma_{kl} - \sigma_{jl} - \sigma_{kk} \\ e &= \sigma_{kk} + \sigma_{ll} - 2\sigma_{kl} \\ f &= \sigma_{ik} + \sigma_{jl} - \sigma_{il} - \sigma_{jk}. \end{aligned} \tag{19}$$

Then, from Eq 17, one gets

$$\begin{aligned}
 p_{ijkl}^{(0)} &= \left[\int \int \int dx_{ij} dx_{jk} dx_{kl} P(x_{ij}, x_{jk}, x_{kl}) F_0(x_{ij}) F_0(x_{jk}) F_0(x_{kl}) \right. \\
 &\quad \left. \times F_0(x_{ij} + x_{jk}) F_0(x_{jk} + x_{kl}) F_0(x_{ij} + x_{jk} + x_{kl}) \right]^3 \\
 &= \left[\frac{1}{(2\pi)^{3/2} \sqrt{\det(\mathbf{W}_4)}} \int d\mathbf{u}_x e^{-\frac{1}{2}\mathbf{u}_x^T (\mathbf{W}_4)^{-1} \mathbf{u}_x} e^{-\frac{1}{2}\mathbf{u}_x^T \Lambda_4 \mathbf{u}_x} \right]^3 \\
 &= [\det(\mathbf{W}_4) \det((\mathbf{W}_4)^{-1} + \Lambda_4)]^{-3/2},
 \end{aligned} \tag{20}$$

where

$$\Lambda_4 = \frac{3}{r_c^2} \begin{pmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{pmatrix}. \tag{21}$$

Conditional pairwise contact probability. With super-resolution tracing, Bintu *et al.* examined whether the association between two chromatin loci facilitates or prevents the contact with a third locus [14]. More specifically, given three monomers of indices $i < j < k$, two conditional probabilities $p(r_{jk} \leq r_c | r_{ij} \leq r_c)$ and $p(r_{jk} \leq r_c | r_{ij} > r_c)$ were compared with the marginal one $p(r_{jk} \leq r_c)$ (see Fig 5 in Ref. [14]). They found “cooperativity”, namely $p(r_{jk} \leq r_c | r_{ij} > r_c) < p(r_{jk} \leq r_c) < p(r_{jk} \leq r_c | r_{ij} \leq r_c)$, among more than 80% triplets of CTCF-bound and generic loci, as well as little effect of cohesin depletion induced by auxin treatment.

Assuming $F_0(r)$, we next aim at deriving the conditional contact probabilities of the j - and k -th monomers given that the j -th monomer is in contact with the i -th monomer ($p_{j|i}^{(0)}$), or given that the j -th monomer is in contact with the i -th monomer ($p_{jk|i}^{(0)}$), with an ordering of indices $i < j < k$. The first conditional probability can be calculated based on the relation that $p_{jk|i}^{(0)} = p_{ij,jk}^{(0)} / p_{ij}^{(0)}$, in which

$$\begin{aligned}
 p_{ij,jk}^{(0)} &= \frac{1}{\Xi} \int d\mathbf{r} P(\mathbf{r}) F_0(\|\vec{r}_i - \vec{r}_j\|) F_0(\|\vec{r}_j - \vec{r}_k\|) \\
 &= \left[\frac{\det(\mathbf{K}_0 + \mathbf{\Omega})}{\det(\mathbf{K}_0)} \right]^{-3/2}
 \end{aligned} \tag{22}$$

with non-zero elements $\mathbf{\Omega}_{ii} = \mathbf{\Omega}_{kk} = \mathbf{\Omega}_{jj}/2 = -\mathbf{\Omega}_{ij} = -\mathbf{\Omega}_{jk} = 3/r_c^2$, or equivalently

$$\begin{aligned}
 p_{ij,jk}^{(0)} &= \left[\int \int dx_{ij} dx_{jk} P(x_{ij}, x_{jk}) F_0(x_{ij}) F_0(x_{jk}) \right]^3 \\
 &= \left[1 + \frac{3(a+c)}{r_c^2} + \frac{9z}{r_c^4} \right]^{-3/2},
 \end{aligned} \tag{23}$$

where we have applied Eqs 13 and 14. By using the relation that $p_{jk}^{(0)} = p_{jk|i}^{(0)} p_{ij}^{(0)} + p_{jk\setminus ij}^{(0)} (1 - p_{ij}^{(0)})$, the second conditional pairwise contact probability can be determined as

$$p_{jk\setminus ij}^{(0)} = \frac{p_{jk}^{(0)} - p_{jk|i}^{(0)} p_{ij}^{(0)}}{1 - p_{ij}^{(0)}} = \frac{p_{jk}^{(0)} - p_{ij,jk}^{(0)}}{1 - p_{ij}^{(0)}}, \tag{24}$$

where the marginal contact probabilities $p_{ij}^{(0)}$ and $p_{jk}^{(0)}$ are given by Eq 12.

Together with $p_{ij}^{(0)} = (1 + 3a/r_c^2)^{-3/2}$ and $p_{jk}^{(0)} = (1 + 3c/r_c^2)^{-3/2}$, it is straightforward to prove that for any triplet

$$p_{jk \setminus ij}^{(0)} \leq p_{jk}^{(0)} \leq p_{jk \setminus ij}^{(0)} \tag{25}$$

The equality signs hold if and only if $b = 0$ (see Eq 14), which corresponds to the case that x_{ij} and x_{jk} are completely uncorrelated.

Gaussian polymer chain. At last we consider the simplest model, a Gaussian polymer chain, in which $k_{ij} = k_0$ if $|i - j| = 1$ and $k_{ij} = 0$ otherwise. It follows that $\sigma_{ij} = \min(i, j)/k_0$. With Eq 16, it straightforwardly leads to

$$p_{ijk}^{(0)} = p(s_1, s_2) = \left[1 + \frac{6(s_1 + s_2)}{k_0} + \frac{27s_1s_2}{k_0^2} \right]^{-3/2}, \tag{26}$$

where $s_1 = j - i$ and $s_2 = k - j$ for any triplet with $i < j < k$.

There is a power-law decay of $p_{ijk}^{(0)}$ with s_1 as $p_{ijk}^{(0)} \sim s_1^{-3/2}$ if $s_1 \gg s_2$. It is also easy to prove that the quantity, $p_{ijk}/p_{ij}p_{jk}$, discussed in Fig 6D is a nontrivial function of s_1 and s_2 . If $s_1 + s_2 = \text{constant}$, it has its minimum at $s_1 = s_2$, and maximum at $s_1 = 1$ or $s_2 = 1$.

In addition, as a result of $b = 0$ in Eq 14, we notice that $p_{ijk}^{(0)} = p_{jk \setminus ij}^{(0)} = p_{ij}^{(0)}$ for a Gaussian chain (see Eqs 22–25).

Supporting information

S1 Appendix. Derivation of n-body contact probability based on the cross-linking probability modeled with the Heaviside step function. Table A. Genomic regions simulated in this work. Table B. Pearson correlation (PC), stratum-adjusted correlation (SCC8) and distance-corrected Pearson correlation (DCPC9) of the contact probabilities predicted by SBS and HLM compared with Capture-C1 and Tri-C2 experiments. Table C. Stratum adjusted correlation (SCC) and PC coefficients between Hi-C and HLM model. (PDF)

S1 Fig. Stratified Pearson correlations of the HLM-predicted contact probabilities. Stratified PC at α -globin locus of mouse (A) ES and (B) erythroid cells compared with Capture-C (2-body) and Tri-C (3-body) experiments. (TIF)

S2 Fig. Changes of Tri-C triplet contact frequencies in the α -globin region of mouse erythroid cells with respect to ES cells. The analysis was done at the viewpoints of R2 (top) and HS-39 (bottom). Following the statistical analysis in the experiment [17], we use the symbol * to mark all triplet interactions with significant changes ($P < 0.01$). The erythroid cell-specific regulatory hub and diffuse interactions among CTCF boundary sites, which are highlighted by dashed triangles, are both captured by Tri-C and our theory (Fig 3E). (TIF)

S3 Fig. Analysis of the changes in triplet contacts ($p_{ijk}^{\text{AWAPL}} - p_{ijk}^{\text{WT}}$) and in pairwise contacts ($p_{ij}^{\text{AWAPL}} - p_{ij}^{\text{WT}}$) among CTCF binding sites in WAPL lacking cells. (A) The triplet contacts from MC-4C data [18] with respect to the viewpoints of E (top) and K (bottom). The triplet contacts predicted by HLM are shown in Fig 4D. (B) The enrichment of pairwise contacts between long-range CTCF binding sites (off-diagonal elements in red corresponding to $(p_{ij}^{\text{AWAPL}} - p_{ij}^{\text{WT}}) \geq 0.05$) is counteracted by the depletion of contacts in the domains flanked by short-range CTCF binding sites (matrix elements along the diagonal block in blue

corresponding to $(p_{ij}^{\text{AWAPL}} - p_{ij}^{\text{WT}}) \lesssim -0.05$). The data of pairwise contacts are obtained from Hi-C data [63].

(TIF)

S4 Fig. WAPL depletion-induced fold changes of n -body contact probability. Fold changes of (A) four-body contacts double-anchored at sites E, K and (B) five-body contacts triple-anchored at sites E, H and K. The absolute change of contact probabilities is shown in Fig 4F and 4G, respectively.

(TIF)

S5 Fig. Comparison between HLM and SPRITE. Comparison between HLM and SPRITE similarly to those of Fig 5 in a 0.52 Mb region on human chr12. (C) From the viewpoint of the promoter of active ETV6 gene which encodes an ETS family transcription factor.

(TIF)

S6 Fig. Many-body contacts of a Gaussian polymer chain calculated with step function.

Same as Fig 6 but using $F_1(r)$ (Heaviside step function) as the cross-linking probability.

(TIF)

S7 Fig. Conditional pairwise contact probability with either a Gaussian (left column) or Heaviside step (right column) cross-linking probability. (A) Heatmaps of $\log_{10}(p_{ij})$ in a 1.23-Mb region on chr21 of IMR90 cells from Hi-C and $\log_{10}(p_{ij})$ from HLM. (B) Comparison between the unconditioned contact probability p_{jk} , the conditional contact probability, $p_{jk|ij}$, and $p_{jk\setminus ij}$ calculated for CTCF-site triplets. (C) p_{jk} , $p_{jk|ij}$, and $p_{jk\setminus ij}$ calculated for all triplets ijk of $i < j < k$, which are sorted in an ascending order of p_{jk} .

(TIF)

S8 Fig. Triplet contacts predicted by simple rules with reference to Fig 3C and 3D. (A)

Three-body contact matrix at α -globin locus of mouse erythroid cells measured by Tri-C experiment or predicted by using three simple rules (B-D). (E) Pearson correlations between the results from 4 methods and the experiment.

(TIF)

S9 Fig. Excluded volume in a polymer melt at different levels of coarse graining. (A) A typical configuration of a dense polymer melt and (B) one polymer chain in the melt at three levels of coarse graining. The beads in (B) are colored differently along the chain, with a diameter of the most probable bond length at the corresponding scales. (C) Probability of the bond length and (D) intra-chain pairwise monomer distance.

(TIF)

S10 Fig. HLM of mouse β -globin locus on chr7 at resolution of 8kb. (A) Pairwise contact probability from Hi-C [70] compared with HLM, which has a PC of 0.996. (B) Triplet contact probabilities from Tri-C compared with HLM, which are anchored at 3'HS1 and HS2 with PCs of 0.62 and 0.88, respectively.

(TIF)

S11 Fig. HLM of mouse Pcdh α locus on chr18 at 5 kb resolution (see also the caption of S10 Fig). Compared with the MC-4C dataset, (A) the pairwise contact probability has a PC of 0.98, and (B) the triplet contact probabilities have PCs of 0.84, 0.76, 0.76, 0.88, and 0.70 at the viewpoint of Pcdh α 1, Pcdh α 11, Pcdh α c1, HS7, and HS5-1, respectively.

(TIF)

S12 Fig. Comparison between the model trainings with different choices of cost functions and optimizers. We trained a polymer model of a 2.4-Mb genomic region on chr8 in mouse ES cells at 25-kb resolution [70]. **(A)** The trajectories of various cost functions L in a log-log scale, minimized by using one of the four methods (RS, GD, RMSprop, and ADAM) with different cross-linking probabilities F_α ($\alpha = 0, 1$). **(B)** Comparing $P(s)$ from Hi-C and from three models, which were all trained with ADAM using F_0 , but with different forms of the cost functions. **(C)** Comparison of $\log_{10}(p_{ij})$ from Hi-C (top) with that from the model trained by minimizing L_1 with ADAM (bottom).
(TIF)

S13 Fig. Comparison between the models trained for a 10-Mb genomic region at 50-kb resolution in GM12878 cells [77] with different choices of cost functions and optimizers. **(A-C)** Same as the caption of S12 Fig. **(D)** Pearson correlations between Hi-C and HLM in our previous work [50] (the black line), and new models trained in this work (the colored lines) as a function of genomic separation, s .
(TIF)

Acknowledgments

We thank Dr. Yonghyun Song for careful reading of the manuscript and helpful comments. We thank the Center for Advanced Computation in KIAS for providing computing resources.

Author Contributions

Conceptualization: Lei Liu, Changbong Hyeon.

Data curation: Lei Liu.

Formal analysis: Lei Liu, Changbong Hyeon.

Funding acquisition: Lei Liu, Changbong Hyeon.

Investigation: Lei Liu, Bokai Zhang, Changbong Hyeon.

Methodology: Lei Liu, Changbong Hyeon.

Project administration: Changbong Hyeon.

Resources: Changbong Hyeon.

Software: Lei Liu.

Supervision: Changbong Hyeon.

Validation: Lei Liu, Bokai Zhang, Changbong Hyeon.

Visualization: Lei Liu.

Writing – original draft: Lei Liu, Changbong Hyeon.

Writing – review & editing: Lei Liu, Bokai Zhang, Changbong Hyeon.

References

1. Wang S, Su JH, Beliveau BJ, Bintu B, Moffitt JR, Wu Ct, et al. Spatial organization of chromatin domains and compartments in single chromosomes. *Science*. 2016; 353(6299):598–602. <https://doi.org/10.1126/science.aaf8084> PMID: 27445307
2. Gu B, Swigut T, Spencley A, Bauer MR, Chung M, Meyer T, et al. Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science*. 2018; 359(6379):1050–1055. <https://doi.org/10.1126/science.aao3136> PMID: 29371426

3. Cardozo Gizzi AM, Cattoni D I, Fiche JB, Espinola SM, Gurgo J, Messina O, et al. Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Mol Cell*. 2019; 74(1):212–222. <https://doi.org/10.1016/j.molcel.2019.01.011> PMID: 30795893
4. Ohno M, Ando T, Priest DG, Kumar V, Yoshida Y, Taniguchi Y. Sub-nucleosomal Genome Structure Reveals Distinct Nucleosome Folding Motifs. *Cell*. 2019; 176(3):520–534. <https://doi.org/10.1016/j.cell.2018.12.014> PMID: 30661750
5. Hsieh THS, Cattoglio C, Slobodyanyuk E, Hansen AS, Rando OJ, Tjian R, et al. Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol Cell*. 2020; 78(3):539–553. <https://doi.org/10.1016/j.molcel.2020.03.002> PMID: 32213323
6. Krietenstein N, Abraham S, Venev SV, Abdennur N, Gibcus J, Hsieh THS, et al. Ultrastructural Details of Mammalian Chromosome Architecture. *Mol Cell*. 2020; 78(3):554–565. <https://doi.org/10.1016/j.molcel.2020.03.003> PMID: 32213324
7. Srinivas K, Sabari BR, Coffey EL, Klein IA, Bojia A, Zamudio AV, et al. Enhancer features that drive formation of transcriptional condensates. *Molecular cell*. 2019; 75(3):549–561. <https://doi.org/10.1016/j.molcel.2019.07.009>
8. Oudelaar AM, Harrold CL, Hanssen LLP, Telenius JM, Higgs DR, Hughes JR. A revised model for promoter competition based on multi-way chromatin interactions at the α -globin locus. *Nat Commun*. 2019; 10:5412. <https://doi.org/10.1038/s41467-019-13404-x> PMID: 31776347
9. Beagrie RA, Thieme CJ, Annunziatella C, Baugher C, Zhang Y, Schueler M, et al. Multiplex-GAM: genome-wide identification of chromatin contacts yields insights not captured by Hi-C. *bioRxiv*. 2021;.
10. Lim B, Levine MS. Enhancer-promoter communication: hubs or loops? *Curr Opin Gene Develop*. 2021; 67:5–9. PMID: 33202367
11. Miller JA, Widom J. Collaborative competition mechanism for gene activation in vivo. *Molecular and cellular biology*. 2003; 23(5):1623–1632. <https://doi.org/10.1128/MCB.23.5.1623-1632.2003> PMID: 12588982
12. Hermesen R, Tans S, Ten Wolde PR. Transcriptional regulation by competing transcription factor modules. *PLoS Comput Biol*. 2006; 2(12):e164. <https://doi.org/10.1371/journal.pcbi.0020164> PMID: 17140283
13. Kempfer R, Pombo A. Methods for mapping 3D chromosome architecture. *Nat Rev Genet*. 2020; 21(4):207–226. <https://doi.org/10.1038/s41576-019-0195-2> PMID: 31848476
14. Bintu B, Mateo LJ, Su JH, Sinnott-Armstrong NA, Parker M, Kinrot S, et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*. 2018; 362(6413): eaau1783. <https://doi.org/10.1126/science.aau1783> PMID: 30361340
15. Olivares-Chauvet P, Mukamel Z, Lifshitz A, Schwartzman O, Elkayam NO, Lubling Y, et al. Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature*. 2016; 540:296–300. <https://doi.org/10.1038/nature20158> PMID: 27919068
16. Darrow EM, Huntley MH, Dudchenko O, Stamenova EK, Durand NC, Sun Z, et al. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc Natl Acad Sci U S A*. 2016; 113(31):E4504–E4512. <https://doi.org/10.1073/pnas.1609643113> PMID: 27432957
17. Oudelaar AM, Davies JOJ, Hanssen LLP, Telenius JM, Schwessinger R, Liu Y, et al. Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nat Genet*. 2018; 50:1744–1751. <https://doi.org/10.1038/s41588-018-0253-2> PMID: 30374068
18. Allahyar A, Vermeulen C, Bouwman BAM, Krijger PHL, Versteegen MJAM, Geeven G, et al. Enhancer hubs and loop collisions identified from single-allele topologies. *Nat Genet*. 2018; 50(8):1151–1160. <https://doi.org/10.1038/s41588-018-0161-5> PMID: 29988121
19. Vermeulen C, Allahyar A, Bouwman BAM, Krijger PHL, Versteegen MJAM, Geeven G, et al. Multi-contact 4C: long-molecule sequencing of complex proximity ligation products to uncover local cooperative and competitive chromatin topologies. *Nat Protoc*. 2020; 15:364–397. <https://doi.org/10.1038/s41596-019-0242-7> PMID: 31932773
20. Ulahannan N, Pendleton M, Deshpande A, Schwenk S, Behr JM, Dai X, et al. Nanopore sequencing of DNA concatemers reveals higher-order features of chromatin structure. *bioRxiv*. 2019;.
21. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*. 2013; 502(7469):59–64. <https://doi.org/10.1038/nature12593> PMID: 24067610
22. Tan L, Xing D, Chang CH, Li H, Xie XS. Three-dimensional genome structures of single diploid human cells. *Science*. 2018; 361(6405):924–928. <https://doi.org/10.1126/science.aat5641> PMID: 30166492

23. Tan L, Xing D, Daley N, Xie XS. Three-dimensional genome structures of single sensory neurons in mouse visual and olfactory systems. *Nat Struct Mol Biol.* 2019; 26(4):297–307. <https://doi.org/10.1038/s41594-019-0205-2> PMID: 30936528
24. Takei Y, Yun J, Zheng S, Ollikainen N, Pierson N, White J, et al. Integrated spatial genomics reveals global architecture of single nuclei. *Nature.* 2021; 590(7845):344–350. <https://doi.org/10.1038/s41586-020-03126-2> PMID: 33505024
25. Beagrie RA, Antonioand Schueler Scialdone M, Kraemer DCA, Chotalia M, Xie SQ, Barbieri M, et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature.* 2017; 543:519–524. <https://doi.org/10.1038/nature21411> PMID: 28273065
26. Quinodoz SA, Ollikainen N, Tabak B, Palla A, Schmidt JM, Detmar E, et al. Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell.* 2018; 174(3):744–757.e24. <https://doi.org/10.1016/j.cell.2018.05.024> PMID: 29887377
27. de Gennes PG. *Scaling Concepts in Polymer Physics.* Ithaca and London: Cornell University Press; 1979.
28. Halverson JD, Smrek J, Kremer K, Grosberg AY. From a melt of rings to chromosome territories: the role of topological constraints in genome folding. *Rep Prog Phys.* 2014; 77:022601. <https://doi.org/10.1088/0034-4885/77/2/022601> PMID: 24472896
29. Liu L, Hyeon C. Contact statistics highlight distinct organizing principles of proteins and RNA. *Biophys J.* 2016; 110(11):2320–2327. <https://doi.org/10.1016/j.bpj.2016.04.020> PMID: 27276250
30. Perez-Rathke A, Sun Q, Wang B, Boeva V, Shao Z, Liang J. CHROMATIX: computing the functional landscape of many-body chromatin interactions in transcriptionally active loci from deconvolved single cells. *Genome Biol.* 2020; 21:13. <https://doi.org/10.1186/s13059-019-1904-z> PMID: 31948478
31. Grosberg AY, Nechaev SK, Shakhnovich EI. The role of topological constraints in the kinetics of collapse of macromolecules. *J Phys.* 1988; 49(12):2095–2100. <https://doi.org/10.1051/jphys:0198800490120209500>
32. Grosberg A, Rabin Y, Havlin S, Neer A. Crumpled globule model of the three-dimensional structure of DNA. *Europhys Lett.* 1993; 23:373. <https://doi.org/10.1209/0295-5075/23/5/012>
33. Grosberg AY, Khokhlov AR. *Statistical Physics of Macromolecules.* New York: AIP Press; 1994.
34. Rosa A, Everaers R. Structure and Dynamics of Interphase Chromosomes. *PLoS Comput Biol.* 2008; 4:e1000153. <https://doi.org/10.1371/journal.pcbi.1000153> PMID: 18725929
35. Tjong H, Gong K, Chen L, Alber F. Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Res.* 2012; 22(7):1295–1305. <https://doi.org/10.1101/gr.129437.111> PMID: 22619363
36. Wong H, Marie-Nelly H, Herbert S, Carrivain P, Blanc H, Koszul R, et al. A Predictive Computational Model of the Dynamic 3D Interphase Yeast Nucleus. *Curr Biol.* 2012; 22(20):1881–1890. <https://doi.org/10.1016/j.cub.2012.07.069> PMID: 22940469
37. Gürsoy G, Xu Y, Kenter AL, Liang J. Spatial confinement is a major determinant of the folding landscape of human chromosomes. *Nucleic Acids Res.* 2014; 42:8223–8230. <https://doi.org/10.1093/nar/gku462> PMID: 24990374
38. Kang H, Yoon YG, Thirumalai D, Hyeon C. Confinement-induced glassy dynamics in a model for chromosome organization. *Phys Rev Lett.* 2015; 115:198102. <https://doi.org/10.1103/PhysRevLett.115.198102> PMID: 26588418
39. Barbieri M, Chotalia M, Fraser J, Lavitas LM, Dostie J, Pombo A, et al. Complexity of chromatin folding is captured by the strings and binders switch model. *Proc Natl Acad Sci U S A.* 2012; 109(40):16173–16178. <https://doi.org/10.1073/pnas.1204799109> PMID: 22988072
40. Jost D, Carrivain P, Cavalli G, Vaillant C. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res.* 2014; 42(15):9553–9561. <https://doi.org/10.1093/nar/gku698> PMID: 25092923
41. Brackley CA, Brown JM, Waithe D, Babbs C, Davies J, Hughes JR, et al. Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models. *Genome Biol.* 2016; 17(1):59. <https://doi.org/10.1186/s13059-016-0909-0> PMID: 27036497
42. Brackley CA, Johnson J, Kelly S, Cook PR, Marenduzzo D. Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains. *Nucleic Acids Res.* 2016; 44(8):3503–3512. <https://doi.org/10.1093/nar/gkw135> PMID: 27060145
43. Di Pierro M, Zhang B, Aiden EL, Wolynes PG, Onuchic JN. Transferable model for chromosome architecture. *Proc Natl Acad Sci U S A.* 2016; 113(43):12168–12173. <https://doi.org/10.1073/pnas.1613607113> PMID: 27688758

44. Gursoy G, Xu Y, Liang J. Spatial organization of the budding yeast genome in the cell nucleus and identification of specific chromatin interactions from multi-chromosome constrained chromatin model. *PLoS Comput Biol.* 2017; 13(7):e1005658. <https://doi.org/10.1371/journal.pcbi.1005658> PMID: 28704374
45. Shi G, Liu L, Hyeon C, Thirumalai D. Interphase Human Chromosome Exhibits Out of Equilibrium Glassy Dynamics. *Nat Commun.* 2018; 9:3161. <https://doi.org/10.1038/s41467-018-05606-6> PMID: 30089831
46. Bianco S, Lupiáñez DG, Chiariello AM, Annunziatella C, Kraft K, Schöpflin R, et al. Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat Genetics.* 2018; 50:662–667. <https://doi.org/10.1038/s41588-018-0098-8> PMID: 29662163
47. Liu L, Shi G, Thirumalai D, Hyeon C. Chain organization of human interphase chromosome determines the spatiotemporal dynamics of chromatin loci. *PLoS Comp Biol.* 2018; 14(12):e1006617. <https://doi.org/10.1371/journal.pcbi.1006617> PMID: 30507936
48. Nir G, Farabella I, Estrada CP, Ebeling CG, Beliveau BJ, Sasaki HM, et al. Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS genetics.* 2018; 14(12):e1007872. <https://doi.org/10.1371/journal.pgen.1007872> PMID: 30586358
49. Zhang S, Chen F, Bahar I. Differences in the intrinsic spatial dynamics of the chromatin contribute to cell differentiation. *Nucl Acids Res.* 2020; 48:1131–1145. <https://doi.org/10.1093/nar/gkz1102> PMID: 31828312
50. Liu L, Kim MH, Hyeon C. Heterogeneous loop model to infer 3D chromosome structures from Hi-C. *Bio-phys J.* 2019; 117(3):613–625. <https://doi.org/10.1016/j.bpj.2019.06.032> PMID: 31337548
51. Liu L, Hyeon C. Revisiting the organization of Polycomb-repressed domains: 3D chromatin models from Hi-C compared with super-resolution imaging. *Nucleic Acids Res.* 2020; 48(20):11486–11494. <https://doi.org/10.1093/nar/gkaa932> PMID: 33095877
52. Chu X, Wang J. Microscopic Chromosomal Structural and Dynamical Origin of Cell Differentiation and Reprogramming. *Adv Sci.* 2020; 7(20):2001572. <https://doi.org/10.1002/adv.202001572> PMID: 33101859
53. Di Stefano M, Nützmänn HW, Marti-Renom MA, Jost D. Polymer modelling unveils the roles of heterochromatin and nucleolar organizing regions in shaping 3D genome organization in *Arabidopsis thaliana*. *Nucleic Acids Res.* 2021; 49(4):1840–1858. <https://doi.org/10.1093/nar/gkaa1275> PMID: 33444439
54. Shi G, Thirumalai D. From Hi-C Contact Map to Three-dimensional Organization of Interphase Human Chromosomes. *Phys Rev X.* 2021; 11(1):011051.
55. Bianco S, Annunziatella C, Andrey G, Chiariello AM, Esposito A, Fiorillo L, et al. Modeling Single-Molecule Conformations of the HoxD Region in Mouse Embryonic Stem and Cortical Neuronal Cells. *Cell Rep.* 2019; 28(6):1574–1583. <https://doi.org/10.1016/j.celrep.2019.07.013> PMID: 31390570
56. Chiariello AM, Bianco S, Oudelaar AM, Esposito A, Annunziatella C, Fiorillo L, et al. A Dynamic Folded Hairpin Conformation Is Associated with alpha-Globin Activation in Erythroid Cells. *Cell Rep.* 2020; 30(7):2125–2135. <https://doi.org/10.1016/j.celrep.2020.01.044> PMID: 32075757
57. Bak JH, Kim MH, Liu L, Hyeon C. A unified framework for inferring the multi-scale organization of chromatin domains from Hi-C. *PLoS Comput Biol.* 2021; 17(3):e1008834. <https://doi.org/10.1371/journal.pcbi.1008834> PMID: 33724986
58. Bohn M, Heermann DW, van Driel R. Random loop model for long polymers. *Phys Rev E.* 2007; 76(5):051805. <https://doi.org/10.1103/PhysRevE.76.051805> PMID: 18233679
59. Nitzsche A, Paszkowski-Rogacz M, Matarese F, Janssen-Megens EM, Hubner NC, Schulz H, et al. RAD21 Cooperates with Pluripotency Transcription Factors in the Maintenance of Embryonic Stem Cell Identity. *PLoS One.* 2011; 6(5):e19470. <https://doi.org/10.1371/journal.pone.0019470> PMID: 21589869
60. The ENCODE Project Consortium. An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature.* 2012; 489(7414):57–74. <https://doi.org/10.1038/nature11247>
61. Gürsoy G, Xu Y, Kenter AL, Liang J. Computational construction of 3D chromatin ensembles and prediction of functional interactions of alpha-globin locus from 5C data. *Nucleic Acids Res.* 2017; 45(20):11547–11558. <https://doi.org/10.1093/nar/gkx784> PMID: 28981716
62. Yang T, Zhang F, Yardimci GG, Song F, Hardison RC, Noble WS, et al. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* 2017; 27:1939–1949. <https://doi.org/10.1101/gr.220640.117> PMID: 28855260
63. Haarhuis JHI, van der Weide RH, Vincent Blomen A, Yáñez-Cuna JO, Amendola M, van Ruiten MS, et al. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell.* 2017; 169(4):693–707. <https://doi.org/10.1016/j.cell.2017.04.013> PMID: 28475897
64. Sanborn AL, Rao SSP, Huang SC, Durand NC, Huntley MH, Jewett AI, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl*

- Acad Sci USA. 2015; 112(47):E6456–E6465. <https://doi.org/10.1073/pnas.1518552112> PMID: [26499245](https://pubmed.ncbi.nlm.nih.gov/26499245/)
65. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* 2016; 15(9):2038–2049. <https://doi.org/10.1016/j.celrep.2016.04.085> PMID: [27210764](https://pubmed.ncbi.nlm.nih.gov/27210764/)
 66. Banigan EJ, Mirny LA. Loop extrusion: theory meets single-molecule experiments. *Curr Opin Cell Biol.* 2020; 64:124–138. <https://doi.org/10.1016/j.ceb.2020.04.011> PMID: [32534241](https://pubmed.ncbi.nlm.nih.gov/32534241/)
 67. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. *Nature.* 2012; 489(7414):101–108. <https://doi.org/10.1038/nature11233> PMID: [22955620](https://pubmed.ncbi.nlm.nih.gov/22955620/)
 68. Polovnikov KE, Nechaev S, Tamm MV. Many-body contacts in fractal polymer chains and fractional Brownian trajectories. *Phys Rev E.* 2019; 99(3):032501. <https://doi.org/10.1103/PhysRevE.99.032501> PMID: [30999417](https://pubmed.ncbi.nlm.nih.gov/30999417/)
 69. Hsu HP, Kremer K. Static and dynamic properties of large polymer melts in equilibrium. *J Chem Phys.* 2016; 144(15):154907. <https://doi.org/10.1063/1.4946033> PMID: [27389240](https://pubmed.ncbi.nlm.nih.gov/27389240/)
 70. Bonev B, Cohen NM, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, et al. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell.* 2017; 171(3):557–572.e24. <https://doi.org/10.1016/j.cell.2017.09.043> PMID: [29053968](https://pubmed.ncbi.nlm.nih.gov/29053968/)
 71. Le Treut G, Képès F, Orland H. A polymer model for the quantitative reconstruction of chromosome architecture from HiC and GAM data. *Biophys J.* 2018; 115(12):2286–2294. <https://doi.org/10.1016/j.bpj.2018.10.032> PMID: [30527448](https://pubmed.ncbi.nlm.nih.gov/30527448/)
 72. Shinkai S, Nakagawa M, Sugawara T, Togashi Y, Ochiai H, Nakato R, et al. PHI-C: deciphering Hi-C data into polymer dynamics. *NAR Genom Bioinf.* 2020; 2(2):lqaa020. <https://doi.org/10.1093/nargab/lqaa020> PMID: [33575580](https://pubmed.ncbi.nlm.nih.gov/33575580/)
 73. Shinkai S, Sugawara T, Miura H, Hiratani I, Onami S. Microrheology for Hi-C Data Reveals the Spectrum of the Dynamic 3D Genome Organization. *Biophys J.* 2020; 118(9):2220–2228. <https://doi.org/10.1016/j.bpj.2020.02.020> PMID: [32191860](https://pubmed.ncbi.nlm.nih.gov/32191860/)
 74. Tieleman T, Hinton G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA.* 2012; 4(2):26–31.
 75. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv.* 2014; p. 1412.6980.
 76. Mehta P, Bukov M, Wang CH, Day AGR, Richardson C, Fisher CK, et al. A high-bias, low-variance introduction to Machine Learning for physicists. *Phys Rep.* 2019; 810:1–124. <https://doi.org/10.1016/j.physrep.2019.03.001> PMID: [31404441](https://pubmed.ncbi.nlm.nih.gov/31404441/)
 77. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell.* 2014; 159(7):1665–1680. <https://doi.org/10.1016/j.cell.2014.11.021> PMID: [25497547](https://pubmed.ncbi.nlm.nih.gov/25497547/)