Structural Basis of Differential Gene Expression at eQTLs Loci from High-Resolution Ensemble Models of 3D Single-Cell Chromatin Conformations

Lin Du[1], Hammad Farooq[1], Pourya Delafrouz[1], and Jie Liang[1]

[1] Center for Bioinformatics and Quantitative Biology, and Richard and Loan Hill Department of Biomedical Engineering, University of Illinois at Chicago, IL

ABSTRACT

Motivation: Techniques such as high-throughput chromosome conformation capture (Hi-C) have provided a wealth of information on the organization of the nucleus and the genome important for understanding gene expression regulation. Additionally, Genome-Wide Association Studies (GWASs) have uncovered thousands of loci related to complex traits. Expression quantitative trait loci (eQTL) studies have further linked the genetic variants to alteration in expression levels of associated target genes across individuals. However, the functional roles of many eQTLs located in non-coding regions are unclear. Current joint analyses of Hi-C and eQTLs data lack advanced computational tools, limiting what can be learned from these data.

Result: In this work, we developed a computational method for simultaneous analysis of Hi-C and eQTL data. Our method can identify a small set of non-random interactions from all Hi-C interactions. Using these non-random interactions, we reconstruct large ensemble ($\times 10^5$) of high-resolution single-cell 3D chromatin conformations with thorough sampling, which accurately replicate Hi-C measurements. Our results revealed the presence of many-body interactions in

chromatin conformation at single-cell level in eQTL locus, offering detailed view into how three-dimensional structures of chromatin form the physical foundation for gene regulation, including how genetic variants of eQTLs affect the expression level of their associated eGenes. Furthermore, our method can deconvolve chromatin heterogeneity and investigate the spatial associations of eQTLs and eGenes at subpopulation level to reveal their regulatory impacts on gene expression. Together, ensemble modeling of thoroughly sampled single cell chromatin conformations from Hi-C, along with eQTL data, helps to decipher how chromatin 3D structures provide the physical basis for gene regulation, expression control, and aid in understanding of the overall structure-function relationships of genome organization.

Availability and implementation: It is available at https://github.com/uic-liang-lab/3DChromFolding-eQTL-Loci

1. INTRODUCTION

The spatial organization of genome plays important roles in gene expression and DNA replication (Fudenberg et al. 2011, Sexton and Cavalli 2015, Pope et al. 2014). Studies based on chromosome conformation capture techniques (Baù and Marti-Renom 2012, Simonis et al. 2006, Dostie et al. 2006) such as Hi-C (Lieberman-Aiden et al. 2009) contained in the 4D Nucleome Project (the 4D Nucleome Network et al. 2017, Reiff et al. 2022), have provided a wealth of information on the 3D organization of genomes of many tissues and cell types. At the same time, a large number of genetic variants statistically associated with gene expression across individuals

have been identified for 49 different tissues and are now available from the Genotype-Tissue Expression (GTEx) Consortium (Lonsdale et al. 2013, The GTEx Consortium et al. 2015, Aguet et al. 2017).

Joint analysis of eQTLs data and chromatin interactions has already revealed important insights (Aguet et al. 2017, Duggal, Wang and Kingsford 2014, Lu et al. 2020, Schmiedel et al. 2018, Yu, Hu and Li 2019). eQTLs are found to be spatially close to their target genes (Duggal, Wang and Kingsford 2014), and eQTLs enriched in cis regulatory elements tend to be in close spatial proximity with their target gene promoters (Aguet et al. 2017). A recent study showed that genes whose expression are significantly associated with eQTLs are positively correlated with chromatin contact frequencies. Further, eQTLs and their target genes are found to be more likely to co-localize within the same Topologically Associating Domains (TADs) (Yu, Hu and Li 2019).

However, current joint analyses that integrate Hi-C and eQTLs data have limitations. While there is abundant of Hi-C data, 3D spatial configurations of chromatin loci and interactions in individual cells cannot be directly inferred due to the intrinsic 2D and population-averaged nature of Hi-C frequency heatmaps, Hi-C data do not provide direct knowledge on how 3D structures of the chromatin determine gene expression in individual cells. Furthermore, it is also unclear that among the numerous chromatin interactions identified by Hi-C studies, which ones reflect functional associations, and which ones are due to by-stander effect, resulting from random collision owing to volume confinement and other factors (Belmont 2014, Gürsoy et al. 2014). In addition, many-body interactions in condensates involving more than pairwise (e.g.

70 pairs of eQTLs and their associated genes, eGene-eQTLs) interactions play important roles in
71 genome organization and function (Hnisz et al. 2017), but they cannot be directly identified from
72 2D Hi-C heatmaps. Furthermore, while heatmaps of population Hi-C exhibit highly detailed
73 patterns, the heterogeneities of the chromatin 3D structures in the underlying cell population are
74 difficult to assess. While there are several approaches for modeling 3D chromatin from
75 experimental Hi-C data, including consensus optimization [ShRec3D (Lesne et al. 2014),
76 ChromSDE (Zhang et al. 2013) and Chrom3D (Paulsen et al. 2017)], ensemble optimization
77 [(Baù and Marti-Renom 2012), and PGS (Tjong et al. 2016, Hua et al. 2018)] ,and block
78 copolymer models (Di Pierro et al. 2016, Zhang and Wolynes 2017, Conte et al. 2020,
79 Contessoto, Cheng and Onuchic 2022) (reviewed in (Liang and Perez-Rathke 2021)), adequately
80 sampling diverse chromatin conformations remains challenging (Liang and Perez-Rathke 2021).
81 As chromatin conformations in a cell population may fall into a set of distinct structural clusters
82 (Gürsoy et al. 2017, Sun et al. 2021), where each cluster harbors conformations of similar
83 topology, some clusters may have favorable 3D spatial arrangements of promoter, enhancer,
84 gene, and other elements to facilitate gene expression, while others may not. However, such
85 chromatin heterogeneity and analysis of cell subpopulations with relevant chromatin structural
86 clusters have not been quantified in the eQTLs enriched loci.
87
88 In this study, we develop a new computational method for simultaneous analysis of both Hi-C
89 and eQTLs data, overcoming difficulties listed above. Specifically, we construct a pipeline for
90 identifying non-random Hi-C interactions around eQTLs/eGenes that are beyond polymer
91 collision. We further generate large ensembles (5×10^4) of 3D single-cell chromatin
92 conformations models for the selected locus. We then compute subpopulations of single-cell

93 chromatin configurations at the locus and quantified the overall chromatin heterogeneity. Our
94 tools allow interrogations of the 3D chromatin structures of loci where eQTLs and/or eGenes
95 reside to gain an understanding of how they are spatially associated, and how such spacial
96 relationship may affect regulation of gene expression. Our tools also identify the participating
97 genes, promoters, and other elements in the spatial context of the eQTLs. Furthermore, with 3D
98 ensemble models of their spatial arrangement constructed, our results uncovered higher-order
99 many-body interacting units important for gene regulation of an eQTLs-rich locus across
100 lymphoblastoid cells (GM12878), human primary mammary epithelial cells (HMEC), and lung
101 fibroblast cells (IMR-90). Our results demonstrate the benefits of integrating Hi-C and eQTL
102 data through deep sampling of large ensemble models of single cell chromatin conformations,
103 allowing us to explore critical issues related to the differences in genome 3D structure across
104 tissues and the relationship between genome 3D structure and function.
105
106 **2 MATERIALS AND METHODS**
107
108 **2.1 Dataset**
109
110 In this study, we used bulk Hi-C data of lymphoblastoid cells (GM12878, 4DNES3JX38V5),
111 human primary mammary epithelial cells (HMEC, 4DNESIE5R9HS) and lung fibroblast cells
112 (IMR-90, 4DNES1ZEJNRU) from the 4D Nucleome Data Portal (the 4D Nucleome Network et
113 al. 2017, Reiff et al. 2022). We also obtained eQTLs data and bulk tissue expression data for
114 cells of EBV-transformed lymphocytes, breast mammary tissue, and lung tissues from the GTEx

115 Portal (Data Source: GTEx Analysis Release V8) (Lonsdale et al. 2013). All data are based on
116 the reference genome GRCh38.
117
118 **2.2 Overview of the pipeline**
119
120 The overall computational pipeline of our method is illustrated in Figure 1. Our model builds on
121 the recent advancements in methods of deep sampling to generate random polymer 3D chromatin
122 conformations (Gürsoy et al. 2014, Gürsoy et al. 2017, Perez-Rathke et al. 2020, Sun et al. 2021,
123 Perez-Rathke et al. 2019). We first use a sequential Monte Carlo approach (Gürsoy et al. 2014,
124 Gürsoy et al. 2017, Perez-Rathke et al. 2020) to generate large ensembles of chromatin fibers
125 within the confines of the cell nucleus without any Hi-C information. These chromatin fibers are
126 modeled as self-avoiding polymer chains consisting of beads, each representing a 5kb genomic
127 region, at the same resolution of Hi-C measurements (Fig1a). These ensembles serve as our
128 background null model. By analyzing measured population Hi-C data and the random null
129 model, we identify a set of statistically significant non-random chromatin interactions that are
130 unlikely to be due to random collisions (Fig1b). Subsequently, we examined these non-random
131 interactions containing genetic variants of eQTLs and extracted relevant information from the
132 reference genome. Further, we use these non-random interactions as input constraints to
133 reconstruct ensembles of 10^5 single-cell chromatin configurations for selected genomic regions
134 (Fig1c). We then carry out structural analysis on the ensembles of modeled single-cell chromatin
135 configurations to characterize 3D structural properties of the eQTL loci, including conformation
136 compactness, the structure of spatial clusters, and their subpopulation distributions. This provides

us with a quantitative assessment of the 3D structural characteristics of the genome and can help us to understand the mechanisms regulating gene expression.

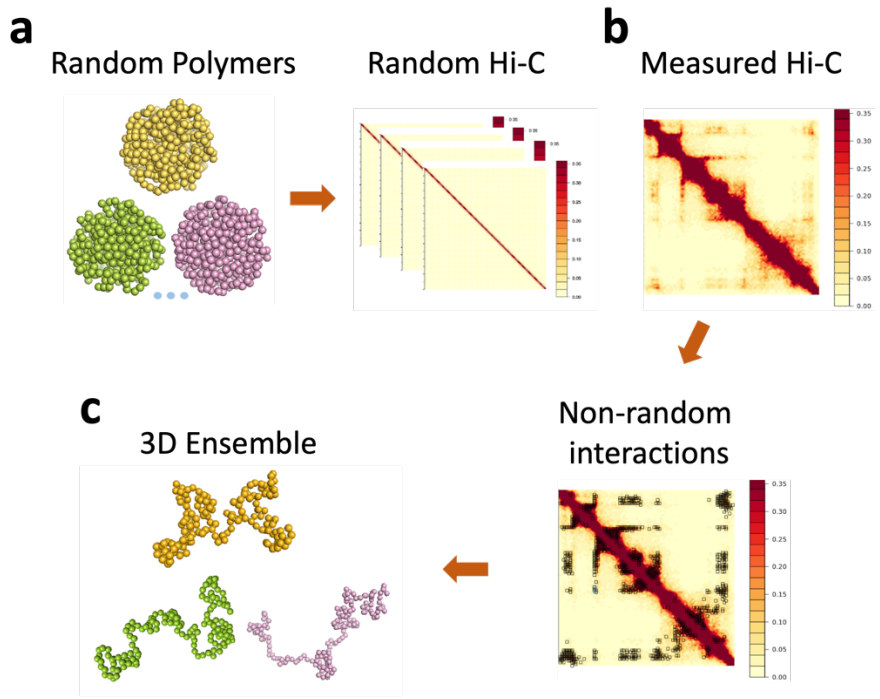


Figure. 1 The pipeline of our computational methods. **a.** A large ensemble of random polymer chains is generated as our background null model. **b.** The statistical significance of each Hi-C interaction is computed by bootstrapping the polymer ensemble for identifying non-random interactions from measured Hi-C data. **c.** A large ensemble of single-cell 3D chromatin conformation is reconstructed using the small set of non-random interactions as constrain.

2.3 Null model

Our null model was built without any Hi-C data. It comprised 10^6 random, self-avoiding 3D polymer chains of chromatin fiber within the confines of the cell nucleus. We utilized the fractal Monte Carlo method (Perez-Rathke et al. 2020), which extends the C-SAC technique (Gürsoy et al. 2014, Gürsoy et al. 2017), to construct this model. To address the challenges of sampling in compact space, this process is refined through a recursive resampling algorithm at set checkpoint lengths (Perez-Rathke et al. 2020). The goal is to sample all feasible self-avoiding chromatin chains equally within a defined volume confinement, which is strictly maintained through careful weighting during chain growth (Rosenbluth and Rosenbluth 1955, Grassberger 1997, Liang, Zhang and Chen 2002, Zhang et al. 2003, Liu 2004). The null model consists of 5×10^6 chromatin chains, each 4 Mb in length and is made up of 800 beads at 5 kb resolution. The volume confinement is proportional to the nuclear volume of cell types.

2.4 Non-random interactions

By comparing the Hi-C measurements to our null model, we were able to distinguish between the non-random interactions that may contribute to the 3D chromatin structures and the by-stander interactions that most likely are due to random polymer collision. To estimate the random contact probabilities of pairs of loci, we determine the frequency of 3D configurations where the spatial distance between the corresponding bead pair is less than 80 nm (Giorgetti et al. 2014, Gürsoy et al. 2017). The statistical significance of each Hi-C interaction is determined through bootstrapping the random ensemble (Kleiner et al. 2014). Hi-C interactions with a BH-FDR adjusted p-value below 0.05 are considered as non-random interactions (Benjamini and Hochberg 1995).

2.5 3D single-cell chromatin conformations

By taking these non-random interactions as physical constraints, we generate a large ensemble (5×10^4) of folded 3D chromatin conformations. All conformations are constructed by using our previously described chromatin folding algorithms, which are based on a novel approach of sequential Bayesian inference (Perez-Rathke et al. 2020, Sun et al. 2021).

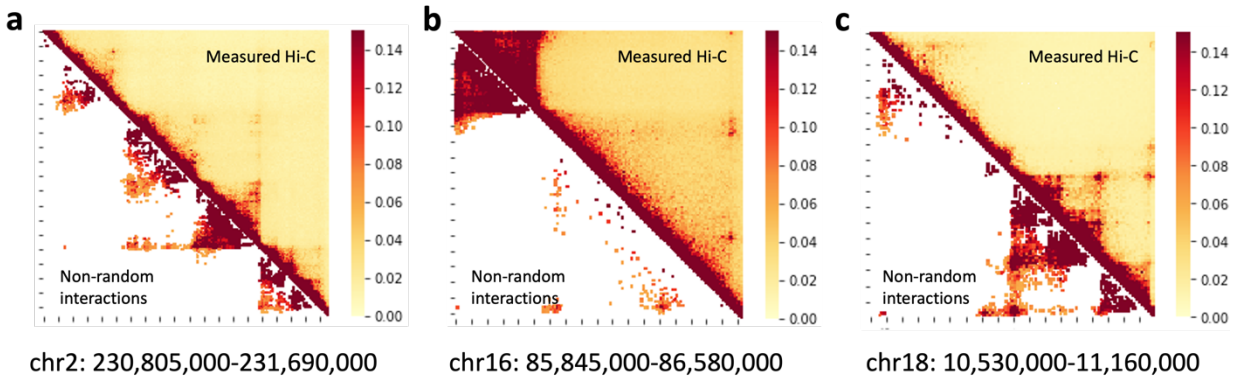
3. RESULTS

3.1 A small fraction of Hi-C contacts are non-random interactions.

It is well known that many of the Hi-C measurements may be by-stander contacts resulting from random collisions of chromatin fibers within the limited space of the nucleus(Belmont 2014, Gürsoy et al. 2017). Our null model of random polymer ensumbles can help to eliminate those spurious by-stander contacts and identify important Hi-C contacts that contribute to the formation of 3D chromatin structure.

We take Hi-C data of three loci from lymphoblastoid cells (GM12878), human primary mammary epithelial cells (HMEC) and lung fibroblast cells (IMR-90) from 4DN. Each loci

196 containing eQTLs variants associated with ≥ 1 gene for at least one of these tissues. These loci
197 are chr2: 230,805,000-231,690,000 (Locus I), chr16: 85,845,000-86,580,000 (Locus II) and
198 chr18:10,530,000-11,160,000 (Locus III).
199
200 Our results show that just a fraction of Hi-C interactions can be considered as non-random.
201 Specifically, 9.8%, 8.9%, and 13.1% of the total contacts for GM12878 cells are identified as
202 non-random interactions for these three loci (Figure 2). 10.1%, 12.9%, and 14.9% of the total
203 contacts for IMR90 cells are identified as non-random interactions and 13.4%, 22.4%, and 31.1%
204 of the total contacts for HMEC cells are identified as non-random interactions (see
205 Supplementary Figure 1). They all meet the specified criteria that adjust *p*-value is less than 0.05.
206 The high proportion of non-random interactions in HMEC cells is due to the low read count of
207 the HMEC Hi-C data. Although they constitute of only a small fraction of Hi-C interactions, the
208 key structural patterns in the heatmap are well-presented by the non-random interactions in all
209 three loci (Figure 2).
210



211

212 **Figure. 2** Small fractions of Hi-C contacts are non-random interactions in GM12878 cells. The
213 non-random interactions of loci a. chr2: 230,805,000-231,690,000, b. chr16: 85,845,000-

214 86,580,000 and c. chr18: 10,530,000-11,160,000. The percentage of non-random interaction is
215 9.8%, 8.9% and 13.1%, respectively.
216
217 **3.2 Non-random interactions can reconstruct ensembles of single-cell chromatin**
218 **conformation.**
219
220 To evaluate the importance of these non-random interactions we uncovered, we use these non-
221 random interactions as physical constraints to generate 3D chromatin conformations. We used
222 previously developed efficient chromatin folding algorithms (Sun et al. 2021).
223
224 For Locus I, there are 10%-13% of the total contact pairs in GM12878, IMR90, and HMEC
225 (Figure 3a). Using these non-random interactions, we reconstruct large ensembles of 5×10^5
226 single-cell chromatin conformations for each cell line at 5kb resolution. To compare our
227 predicted ensemble of single-cell chromatin conformation with the experiment-measured Hi-C
228 data, we calculate the interaction probabilities of simulated chromatin conformations. Taking the
229 basic premise that DNA segments near each other are potential candidates for Hi-C ligations, we
230 determined the interaction probabilities based on the frequency of 3D structures where the 3D
231 distances between certain pairs of loci are less than 80 nanometers (Giorgetti et al. 2014, Gürsoy
232 et al. 2017). As shown in Figure 3b, the simulated Hi-C map, obtained by aggregating 10,000
233 single-cell conformations, exhibits strong similarities with the measured Hi-C maps for
234 GM12878, IMR90, and HMEC for Locus I with Pearson correlation coefficients range from 0.94
235 to 0.97. Similar results are observed for the other two loci, where Pearson correlation coefficients

are 0.95, 0.9, and 0.96 for Locus II , and 0.96, 0.87, and 0.95 for Locus III. (see Supplementary Figure 2-3)

These findings demonstrate that by utilizing just a fraction of the Hi-C interactions identified as non-random interactions, we can accurately reconstruct experimental Hi-C contact heatmaps for different regions of chromosomes at high resolution with accuracy. Our results show that the non-random interactions we uncovered are sufficient to guide chromatin folding of these loci in all three cell types.

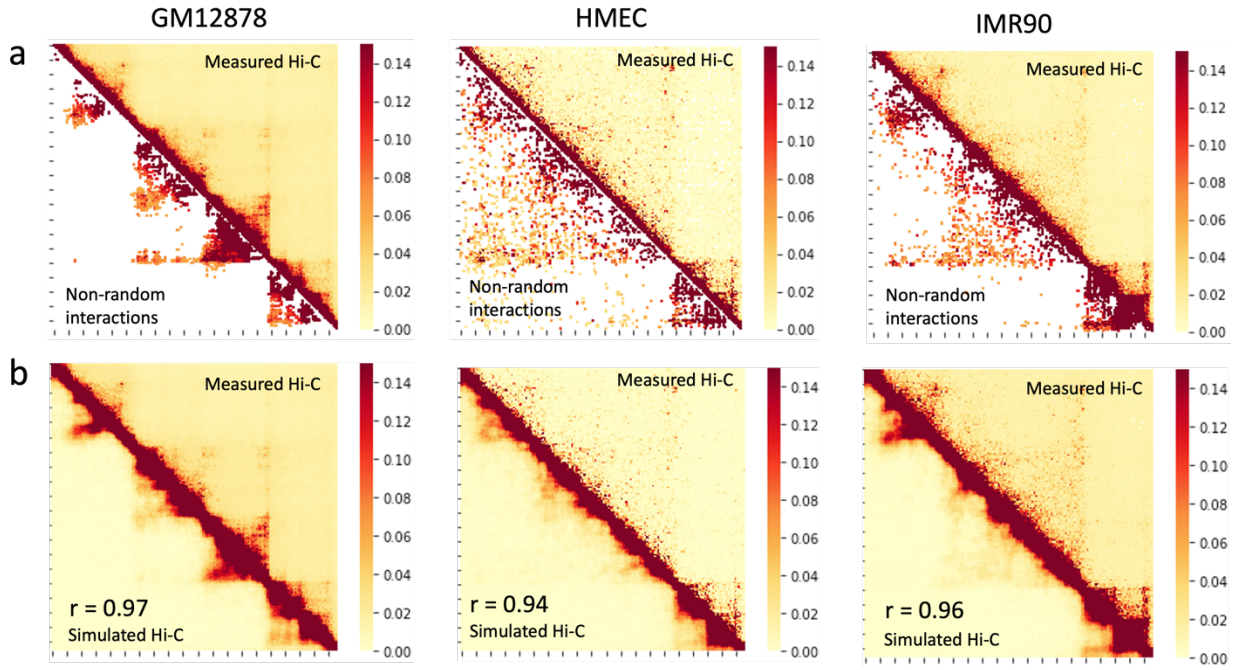


Figure 3. Precise reconstruction of 3D chromatin conformations with high resolution **a**. Non-random interactions identified from our model for GM12878, HMEC, and IMR90 cell lines in Locus I . **b**. Simulated Hi-C map compared to the measured Hi-C map for GM12878, HMEC, and IMR90 cell lines within Locus I. The corresponding Pearson correlation coefficients (R) are 0.97, 0.96, and 0.94, respectively.

251

252 **3.3 Uncovering physical contacts between eGene and eQTLs and their strong influence on**

253 **gene expression.**

254

255 Previous studies showed that eQTLs tend to be located near their target genes in space and loci

256 containing eQTLs show a higher frequency of contact interactions (Duggal, Wang and Kingsford

257 2014). Additionally, eQTLs enriched with cis-regulatory elements are commonly situated in

258 close spatial proximity to the promoters of the genes they target (Aguet et al. 2017, Duggal,

259 Wang and Kingsford 2014). With the 3D single-cell chromatin conformations now at hand, we

260 examined the 3D structures of our selected eQTL containing loci.

261

262 We first analyzed modeled single-cell conformations to determine the occurrence where two 5kb

263 genomic regions represented by polymer beads, one containing eGene and another containing

264 eQTLs, are in physical interaction using the criteria that the median distance between them is

265 less than 80 nm (Giorgetti et al. 2014, Gürsoy et al. 2014). Next, we examine all eGene-eQTLs

266 pairs derived from the GTEx database to determine whether they are in physical contacts among

267 all three loci for EBV-transformed lymphocytes (GM12878), breast–mammary tissue (HMEC),

268 and lung cells (IMR90). eGene-eQTLs pairs are then categorized as either in physical contact or

269 are free of such contacts.

270

271 Among the three loci, we found 153 physical eGene-eQTLs contacts from 798 eGene-eQTLs

272 pairs in EBV-transformed lymphocytes, 365 physical eGene-eQTLs contacts from 2,024 eGene-

273 eQTLs pairs in Brest-mammary tissue, and 535 physical eGene-eQTLs contacts from 2,678

eGene-eQTLs pairs in lung tissue cells. Our method can reveal pairs of eGenes and eQTLs in physical contact. (For more details on these physically contacted eGene-eQTL pairs, see Supplementary Data.)

Intriguingly, our results show that physically contacting eQTLs exhibit a stronger influence on eGene than those without physical contact, as they have larger normalized effect size (NES) values (Lonsdale et al. 2013). Among the three tissues, physically contacting eGene-eQTLs pairs have larger absolute values of NES than non-physically contacted pairs (Figure 4). Here, the NES value of an eQTLs is defined as the slope of the linear regression and is taken as a measure of the effect of the alternative allele (ALT) relative to the reference allele (REF) in the human genome reference following (Lonsdale et al. 2013) . A higher normalized effect size suggests that the genetic variant has a stronger influence on gene expression (Lonsdale et al. 2013, Zeng et al. 2022). Among all three tissues, the NES absolute values difference between physical contact pairs and non-physical contact pairs is significant, which the *p*-value is 1.0×10^{-3}, 3.05×10^{-11} and 2.2×10^{-2}, respectively.

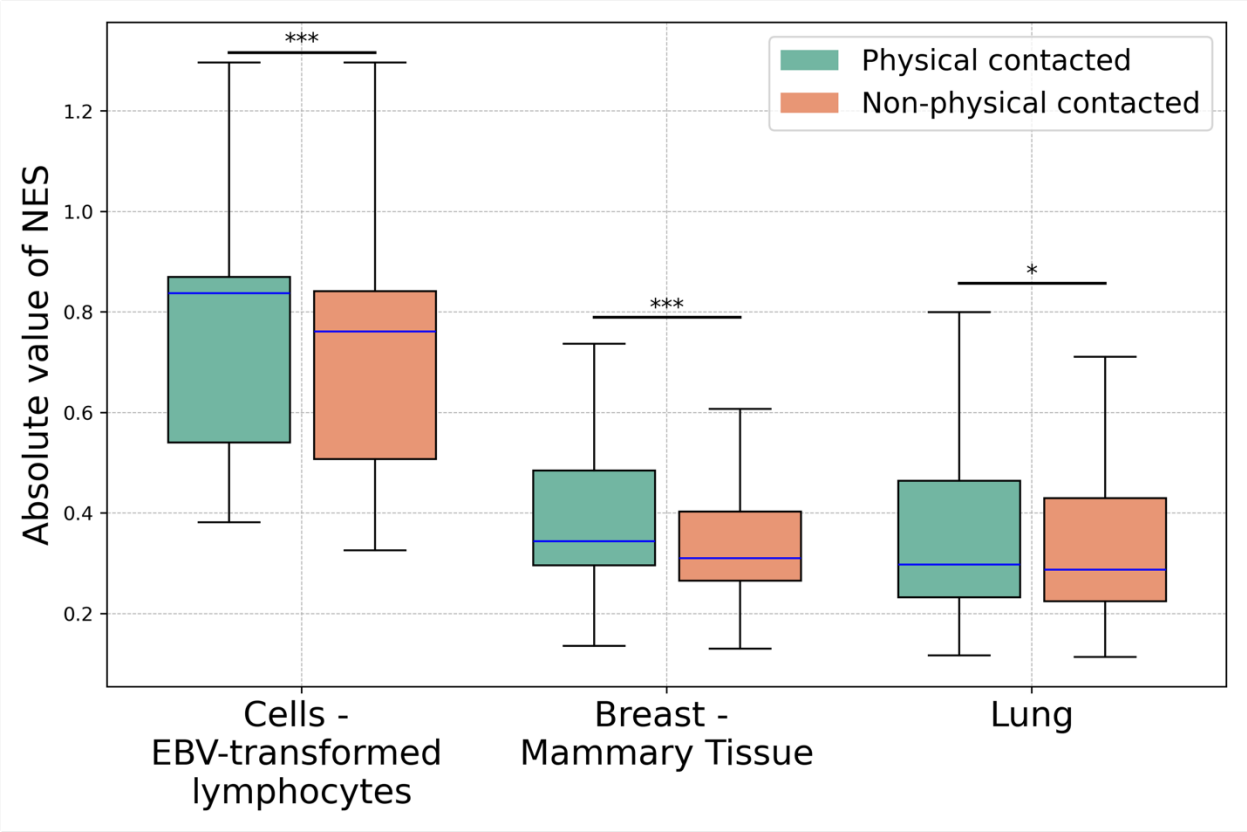


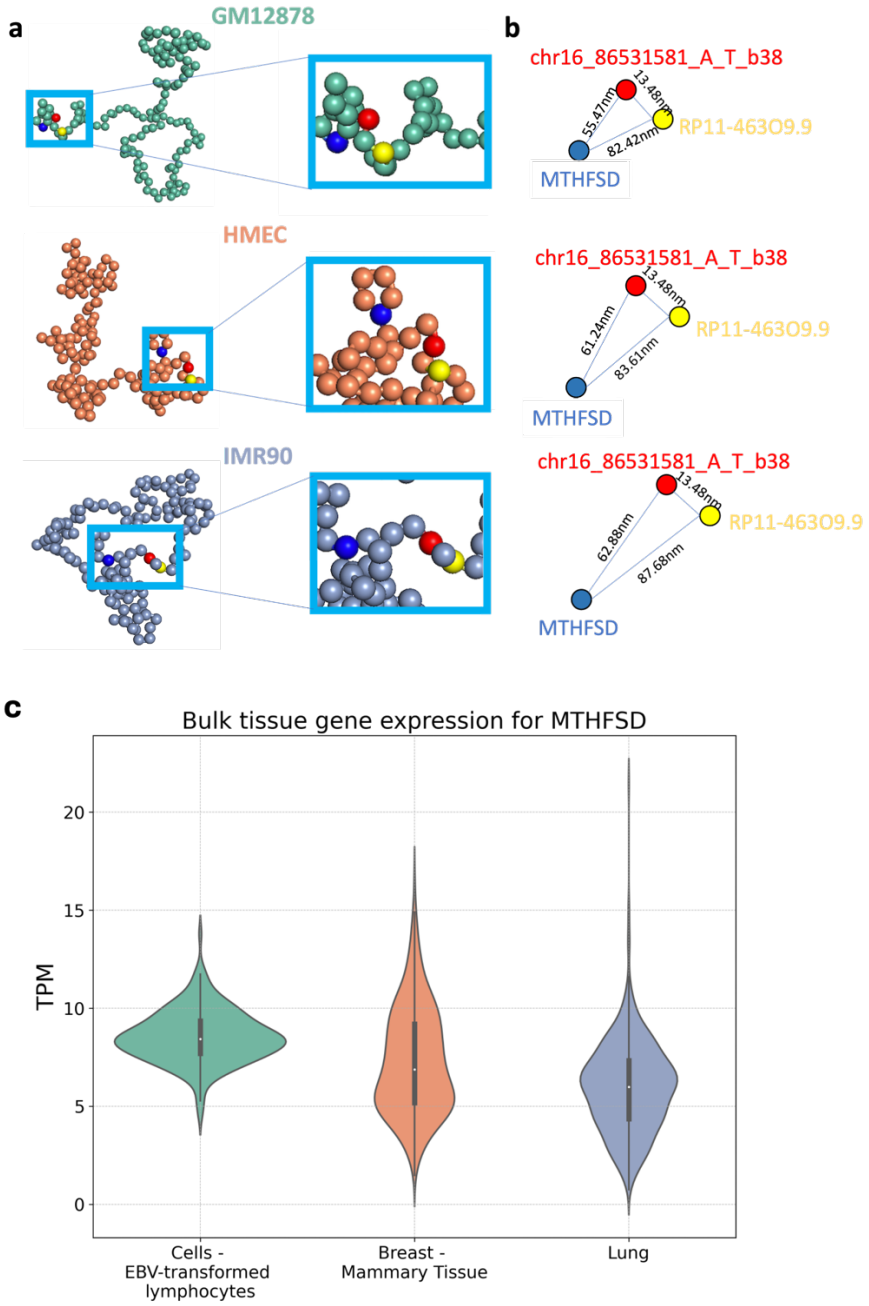
Figure 4. Display of the absolute values of NES for eGene-eQTL pairings across all three examined loci for EBV-transformed lymphocytes (GM12878), breast - mammary tissue (HMEC) and lung (IMR90) cells, respectively. A larger absolute NES value indicates a more substantial genotype impact on gene expression. Significance levels are indicated as follows: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

3.4 Discovery of many-body interactions at eGene and eQTLs loci

Previous studies found that many-body chromatin interactions such as those in condensates likely play important roles in gene regulations (Hnisz et al. 2013, Gürsoy et al. 2017, Sabari et al. 2018, Quinodoz et al. 2018, Perez-Rathke et al. 2020). However, the presence of many-body interactions within individual cells are masked in the bulk Hi-C measurements, as they capture

303 average pairwise genomic interactions across populations. In order to determine whether many-
304 body interactions occur in eGenes and eQTLs loci, we investigate Locus II (chr16: 85,845,000-
305 86,580,000) across GM12878, HMEC and IMR90 cells, which contains eGenes RP (RP11-
306 463O9.9) and MT (MTHFSD), both are associated with the eQTL of chr16_86531581_A_T_b38
307 (Figure 5a, b). We examined the spatial structure of genetic interactions among these eGenes and
308 eQTL at this locus.
309
310 Analysis of the 3D chromatin structure shows that three-body interactions do exist at this locus.
311 Here, we define three-body interactions as a complex of 5kb chromatin regions, where the
312 median Euclidean distances between all pairs of regions in the complex of the ensemble of
313 chromatin conformations are shorter or approximately equal to a cross-linking threshold of 80 ±
314 5 nm (Giorgetti et al. 2014, Gürsoy et al. 2017) (Figure 5b). The interactomes among eGenes RP,
315 MT and the eQTL are three-body in both GM12878 and HMEC. However, the degree of
316 influence of the eQTL variant on gene expression is different among these three tissues. This is
317 indicated by the different values of NES for the same eQTL. The absolute value of NES of this
318 eQTL on eGene MT in lymphocytes (GM12878) is the largest (1.9). The NES values are only
319 0.64 for breast-mammary tissue (HMEC) and 0.53 for lung tissue (IMR90), respectively.
320 Furthermore, the NES of the same eQTL with eGene RP is also the largest (1.6) in lymphocytes
321 (GM12878). The NES value is 1.4 for breast-mammary tissue (HMEC) and 1.0 for Lung tissue
322 (IMR90), respectively.
323
324 To understand the reason why the impact of the variant on gene expression as measured by NES
325 value varies across these three tissues, we ask whether it can be explained by their spacial

326 distance relationship. For this, we randomly selected 5,000 simulated single-cell conformations
327 from the ensembles for GM12878, HMEC, and IMR90 cells and calculated the Euclidean
328 distance between eQTL and the eGenes. From the distributions of the Euclidean distance
329 between eGene and eQTLs, we found that in GM12878 the median spatial distance between the
330 promoter of the eGene MT and the eQTL is 55.74 nm. The distance between the two promoters
331 of the eGenes MT and RP is 82.42 nm. These are closer than the distances in HMEC (61.24 nm,
332 83.61 nm) and IMR90 (62.88 nm, 87.68 nm). For eGene RP, the spatial distance between its
333 promoter and the eQTL is similar among all three cell types; this is consistent with their shared
334 close genomic distances between the eQTL and the eGene RP (<10kb). The shorter 3D physical
335 distances in GM12878 (MT to the eQTL and MT to RP) among the eGenes MT, RP and the
336 eQTL is closer than the corresponding 3D physical distance in HMEC and in IMR90, leading to
337 stronger effects of eQTL on gene expression in GM12878. These results demonstrate that closer
338 spatial proximity between eQTL and the target gene promoters can result in stronger eQTL
339 effects on gene expression, as revealed by the larger absolute value of NES.
340
341 The measured expression level of eGene MT in these tissues provide additional direct support to
342 our findings (Figure 5c). When eQTL has closer spatial distance to MT in GM12878 cells (55.73
343 nm vs 61.24nm in HMEC and 62.88 nm in IMR90), it has stronger effects to the gene
344 expression: The median TPM of gene MT in lymphocytes is the largest at 8.55, followed by 7.27
345 in breast-mammary tissue and 5.97 in lung tissue. Overall, these results show that there exist
346 three-body interactions among eGenes RP, MT and eQTL at Locus II, and the spatial pattern of
347 interactions among the eQTLs and eGenes play important roles in gene regulation and genome
348 functions.



349

350

351 **Figure 5.** The physical interactome of many-body genetic interactions between eGene and eQTL

352 at Locus II. **a**. Representative spatial structures of the locus for lymphocytes (GM12878), breast-

353 mammary tissue (HMEC) and lung fibroblasts (IMR90), with eGenes MTHFSD (blue), RP11-

354 463O9.9 (yellow) and eQTLs chr16_86531581_A_T_b38 (red) depicted. Note the spacial

distances among eGenes and eQTL are different. **b**. The spatial distances between the eGenes MTHFSD (blue), RP11-463O9.9 (yellow), and the eQTLs (red) of the many-body interaction unit in the representative structures shown in (**a**). The spacial distances among the many-bodies are the closest in GM12878, followed by HMEC and then IMR90 **c.** Bulk tissue gene expression level of eGene MTHFSD among the three tissues of lymphocytes (GM12878), breast-mammary tissue (HMEC), and lung (IMR90). The expression level of MTHFSD is the highest in the transformed lymphocytes cells, followed by breast-mammary cells and then lung cells.

3.5 3D chromatin structure difference and heterogeneities at eGene and eQTLs loci among tissues

Quantitatively assessing the chromatin heterogeneities of cells within a tissue is challenging, as chromatin in individual cells may take different configurations, and the overall ensemble may be very heterogeneous. Because of this, initial condition independent thorough sampling is crititcal. With a large and diverse ensemble of single-cell conformations generated, we can group them into clusters by their 3D topology to quantify their heterogeneity (Perez-Rathke et al. 2020, Sun et al. 2021). We use the K-means clustering method to cluster the 3D chromatin configurations into major subpopulations of chromatins with similar topologies. For Locus I, we used 10,000 simulated single-cell chromatin conformations for each of GM12878, IMR90, and HMEC cells, totaling 30,000 conformations, and were able to group them into four clusters (Figure 6). There are notable distinctions in both the contacts heatmap and compactness among the four clusters. Importantly, the distribution of subpopulations varies significantly across different cell lines,

with HMEC and IMR90 predominantly featuring subclusters 1 and 3, respectively, while GM12878 enriched with subclusters 2 and 4 cells. That is, GM12878 cells are more heterogeneous compared to HMEC and IMR90 cells. Our simulated chromatin conformations uncovered significant structural differences among different tissues. Similar results are also observed for Locus II and Locus III (see Supplementary Figure 4-5). These findings highlight substantial differences in the conformation of different cell lines in the same locus and reveal strong tissue-specific 3D chromatin structures, which are not evident in the population Hi-C heatmap directly (Figure 6).

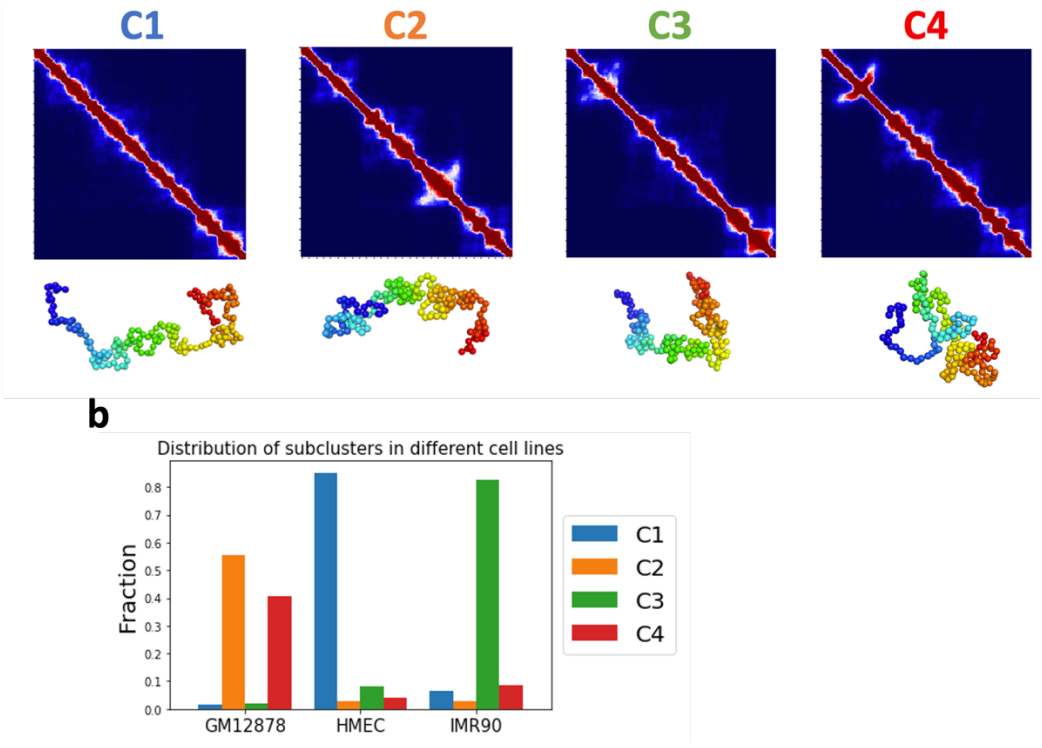


Figure. 6 Uncovering the tissue-specific subpopulations of 3D chromatin structures. **a** 3D chromatin conformation of GM12878, IMR90 and HMEC at Locus I (chr2: 230,805,000-231,690,000). The resulting aggregated contact heatmaps of the four clusters are arranged in

order of their compactness. Representative conformations of four clusters are shown below the heatmap. b Proportions of the four subpopulations in each cell types.

4 CONCLUSION AND DISCUSSION

In this study, we have developed a set of tools to study the relationship between eQTLs and 3D genome organization, through modeling of spatial and physical chromatin interactions at the single-cell level, enabling a general understanding of how eQTLs affect gene expression. Our method identified a small set of non-random interactions from the measured Hi-C data that effectively captured the key patterns from the population Hi-C heatmap. Using these non-random interactions, we reconstruct the 3D single-cell chromatin conformations with accuracy. These simulated chromatin conformations provide detailed structural insights that enable the examination of the 3D spatial pattern of chromatin at loci containing eQTLs and eGenes.

Furthermore, our work enables new discoveries as we can now directly investigate the 3D spatial relationship between eGenes and eQTLs across multiple tissues. Results obtained from the analysis of 3D single-cell chromatin conformations of loci across different cell types showed that the physically interacted eQTLs have stronger effects on gene expression. The specific spatial arrangements of eQTLs in cell subpopulations allow a quantitative approach to understanding the regulatory mechanisms controlling gene expression. In addition, our tools enable novel and compelling biological questions to be formulated, such as possible roles of specific promoters,

414 enhancers, super-enhancers, genes, and other elements relating to specific eQTLs in multi-way
415 interactions.
416
417 While we have shown a promising approach to investigate genome structure and function
418 relationship, there are limitation. Firstly, the availability of high-quality Hi-C data is limited.
419 With more high-quality data from the 4DN becoming available recently, this will become less of
420 an issue. Secondly, we analyzed only three loci among three cell lines. It is desirable to expand
421 the scale to analyze additional eQTLs and eGene-associated loci among many more tissues and
422 cell lines. Furthermore, the advancement of single-cell sequencing technologies, such as scRNA-
423 seq (Tang et al. 2009), scATAC-seq (Buenrostro et al. 2015), and scHi-C (Ramani et al. 2017)
424 provided additional data at single cell level. Integrating these data with 3D chromatin modeling
425 should lead to better understanding of genome structure-function relationship at the single-cell
426 level.
427
428 **Acknowledgment:**
429
430 This work is supported by NIH grants 1R03OD032628-01, 1R03OD036492-01 and
431 2R35GM127084-06. An award for computer time was provided by the U.S. Department of
432 Energy's (DOE) Innovative and Novel Computational Impact on Theory and Experiment
433 (INCITE) Program. This research used resources from the Argonne Leadership Computing
434 Facility, a U.S. DOE Office of Science user facility at Argonne National Laboratory, which is
435 supported by the Office of Science of the U.S. DOE under Contract No. DE-AC02-06CH11357.
436

Data availability:

Non-random interactions, single-cell conformations and physically contacted eGene-eQTLs pairs are available on UIC Indigo (https://indigo.uic.edu/projects/Constructing_High-Resolution_Ensemble_Models_of_3D_Single-Cell_Chromatin_Conformations_of_eQTL_Loci_from_Integrated_Analysis_of_4DN-GTEx_Data_towards_Structural_Basis_of_Differential_Gene_Expression/190164)

REFERENCES

Aguet, François, Brown, Andrew A., Castel, Stephane E., Davis, Joe R., He, Yuan, Jo, Brian, et al., 'Genetic Effects on Gene Expression across Human Tissues', *Nature*, 550/7675 (2017), 204–13

Baù, Davide, and Marti-Renom, Marc A., 'Genome Structure Determination via 3C-Based Data Integration by the Integrative Modeling Platform', *Methods*, 58/3 (2012), 300–306

Belmont, Andrew S, 'Large-Scale Chromatin Organization: The Good, the Surprising, and the Still Perplexing', *Current Opinion in Cell Biology*, 26 (2014), 69–78

Benjamini, Yoav, and Hochberg, Yosef, 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society: Series B (Methodological)*, 57/1 (1995), 289–300

Buenrostro, Jason D., Wu, Beijing, Litzenburger, Ulrike M., Ruff, Dave, Gonzales, Michael L., Snyder, Michael P., et al., 'Single-Cell Chromatin Accessibility Reveals Principles of Regulatory Variation', *Nature*, 523/7561 (2015), 486–90

Conte, Mattia, Fiorillo, Luca, Bianco, Simona, Chiariello, Andrea M., Esposito, Andrea, and Nicodemi, Mario, 'Polymer Physics Indicates Chromatin Folding Variability across Single-Cells Results from State Degeneracy in Phase Separation', *Nature Communications*, 11/1 (2020), 3289

Contessoto, Vinícius G., Cheng, Ryan R., and Onuchic, José N., 'Uncovering the Statistical Physics of 3D Chromosomal Organization Using Data-Driven Modeling', *Current Opinion in Structural Biology*, 75 (2022), 102418

Di Pierro, Michele, Zhang, Bin, Aiden, Erez Lieberman, Wolynes, Peter G., and Onuchic, José N., 'Transferable Model for Chromosome Architecture', *Proceedings of the National Academy of Sciences*, 113/43 (2016), 12168–73

Dostie, Josée, Richmond, Todd A., Arnaout, Ramy A., Selzer, Rebecca R., Lee, William L., Honan, Tracey A., et al., 'Chromosome Conformation Capture Carbon Copy (5C): A Massively Parallel Solution for Mapping Interactions between Genomic Elements', *Genome Research*, 16/10 (2006), 1299–1309

Duggal, Geet, Wang, Hao, and Kingsford, Carl, 'Higher-Order Chromatin Domains Link eQTLs with the Expression of Far-Away Genes', *Nucleic Acids Research*, 42/1 (2014), 87–96

Fudenberg, Geoff, Getz, Gad, Meyerson, Matthew, and Mirny, Leonid A, 'High Order Chromatin Architecture Shapes the Landscape of Chromosomal Alterations in Cancer', *Nature Biotechnology*, 29/12 (2011), 1109–13

Giorgetti, Luca, Galupa, Rafael, Nora, Elphège P., Piolot, Tristan, Lam, France, Dekker, Job, et al., 'Predictive Polymer Modeling Reveals Coupled Fluctuations in Chromosome Conformation and Transcription', *Cell*, 157/4 (2014), 950–63

Grassberger, Peter, 'Pruned-Enriched Rosenbluth Method: Simulations of θ Polymers of Chain Length up to 1 000 000', *Physical Review E*, 56/3 (1997), 3682–93

Gürsoy, Gamze, Xu, Yun, Kenter, Amy L., and Liang, Jie, 'Computational Construction of 3D Chromatin Ensembles and Prediction of Functional Interactions of Alpha-Globin Locus from 5C Data', *Nucleic Acids Research*, 45/20 (2017), 11547–58

———, 'Spatial Confinement Is a Major Determinant of the Folding Landscape of Human Chromosomes', *Nucleic Acids Research*, 42/13 (2014), 8223–30

Gürsoy, Gamze, Xu, Yun, and Liang, Jie, 'Spatial Organization of the Budding Yeast Genome in the Cell Nucleus and Identification of Specific Chromatin Interactions from Multi-Chromosome Constrained Chromatin Model', ed. by Robert L Jernigan, *PLOS Computational Biology*, 13/7 (2017), e1005658

Hnisz, Denes, Abraham, Brian J., Lee, Tong Ihn, Lau, Ashley, Saint-André, Violaine, Sigova, Alla A., et al., 'Super-Enhancers in the Control of Cell Identity and Disease', *Cell*, 155/4 (2013), 934–47

Hnisz, Denes, Shrinivas, Krishna, Young, Richard A., Chakraborty, Arup K., and Sharp, Phillip A., 'A Phase Separation Model for Transcriptional Control', *Cell*, 169/1 (2017), 13–23

499 Hua, Nan, Tjong, Harianto, Shin, Hanjun, Gong, Ke, Zhou, Xianghong Jasmine, and Alber,
500 Frank, 'Producing Genome Structure Populations with the Dynamic and Automated PGS
501 Software', *Nature Protocols*, 13/5 (2018), 915–26

502 Kinney, Nicholas Allen, Sharakhov, Igor V., and Onufriev, Alexey V., 'Chromosome–Nuclear
503 Envelope Attachments Affect Interphase Chromosome Territories and Entanglement',
504 *Epigenetics & Chromatin*, 11/1 (2018), 3

505 Kleiner, Ariel, Talwalkar, Ameet, Sarkar, Purnamrita, and Jordan, Michael I., 'A Scalable
506 Bootstrap for Massive Data', *Journal of the Royal Statistical Society Series B: Statistical*
507 *Methodology*, 76/4 (2014), 795–816

508 Lesne, Annick, Riposo, Julien, Roger, Paul, Cournac, Axel, and Mozziconacci, Julien, '3D
509 Genome Reconstruction from Chromosomal Contacts', *Nature Methods*, 11/11 (2014),
510 1141–43

511 Liang, Jie, and Perez-Rathke, Alan, 'Minimalistic 3D Chromatin Models: Sparse Interactions in
512 Single Cells Drive the Chromatin Fold and Form Many-Body Units', *Current Opinion in*
513 *Structural Biology*, 71 (2021), 200–214

514 Liang, Jie, Zhang, Jinfeng, and Chen, Rong, 'Statistical Geometry of Packing Defects of Lattice
515 Chain Polymer from Enumeration and Sequential Monte Carlo Method', *The Journal of*
516 *Chemical Physics*, 117/7 (2002), 3511–21

517 Lieberman-Aiden, Erez, Van Berkum, Nynke L., Williams, Louise, Imakaev, Maxim, Ragoczy,
518 Tobias, Telling, Agnes, et al., 'Comprehensive Mapping of Long-Range Interactions
519 Reveals Folding Principles of the Human Genome', *Science*, 326/5950 (2009), 289–93

520 Liu, Jun S., *Monte Carlo Strategies in Scientific Computing*, Springer Series in Statistics (New
521 York, NY, 2004) <http://link.springer.com/10.1007/978-0-387-76371-2> [accessed 11
522 March 2024]

523 Lonsdale, John, Thomas, Jeffrey, Salvatore, Mike, Phillips, Rebecca, Lo, Edmund, Shad, Saboor,
524 et al., 'The Genotype-Tissue Expression (GTEx) Project', *Nature Genetics*, 45/6 (2013),
525 580–85

526 Lu, Leina, Liu, Xiaoxiao, Huang, Wei-Kai, Giusti-Rodríguez, Paola, Cui, Jian, Zhang, Shanshan,
527 et al., 'Robust Hi-C Maps of Enhancer-Promoter Interactions Reveal the Function of
528 Non-Coding Genome in Neural Development and Diseases', *Molecular Cell*, 79/3
529 (2020), 521-534.e15

530 Paulsen, Jonas, Sekelja, Monika, Oldenburg, Anja R., Barateau, Alice, Briand, Nolwenn,
531 Delbarre, Erwan, et al., 'Chrom3D: Three-Dimensional Genome Modeling from Hi-C
532 and Nuclear Lamin-Genome Contacts', *Genome Biology*, 18/1 (2017), 21

533 Perez-Rathke, Alan, Mali, Samira, Du, Lin, and Liang, Jie, 'Alterations in Chromatin Folding
534 Patterns in Cancer Variant-Enriched Loci', in *2019 IEEE EMBS International*
535 *Conference on Biomedical & Health Informatics (BHI)* (presented at the 2019 IEEE

536 EMBS International Conference on Biomedical & Health Informatics (BHI), Chicago,
537 IL, USA, 2019), 1–4 <https://ieeexplore.ieee.org/document/8834565/> [accessed 11
538 March 2024]

539 Perez-Rathke, Alan, Sun, Qiu, Wang, Boshen, Boeva, Valentina, Shao, Zhifeng, and Liang, Jie,
540 'CHROMATIX: Computing the Functional Landscape of Many-Body Chromatin
541 Interactions in Transcriptionally Active Loci from Deconvolved Single Cells', *Genome*
542 *Biology*, 21/1 (2020), 13

543 Pope, Benjamin D., Ryba, Tyrone, Dileep, Vishnu, Yue, Feng, Wu, Weisheng, Denas, Olgert, et
544 al., 'Topologically Associating Domains Are Stable Units of Replication-Timing
545 Regulation', *Nature*, 515/7527 (2014), 402–5

546 Quinodoz, Sofia A., Ollikainen, Noah, Tabak, Barbara, Palla, Ali, Schmidt, Jan Marten, Detmar,
547 Elizabeth, et al., 'Higher-Order Inter-Chromosomal Hubs Shape 3D Genome
548 Organization in the Nucleus', *Cell*, 174/3 (2018), 744-757.e24

549 Ramani, Vijay, Deng, Xinxian, Qiu, Ruolan, Gunderson, Kevin L., Steemers, Frank J., Disteche,
550 Christine M., et al., 'Massively Multiplex Single-Cell Hi-C', *Nature Methods*, 14/3
551 (2017), 263–66

552 Reiff, Sarah B., Schroeder, Andrew J., Kırlı, Koray, Cosolo, Andrea, Bakker, Clara, Mercado,
553 Luisa, et al., 'The 4D Nucleome Data Portal as a Resource for Searching and Visualizing
554 Curated Nucleomics Data', *Nature Communications*, 13/1 (2022), 2365

555 Rosenbluth, Marshall N., and Rosenbluth, Arianna W., 'Monte Carlo Calculation of the Average
556 Extension of Molecular Chains', *The Journal of Chemical Physics*, 23/2 (1955), 356–59

557 Sabari, Benjamin R., Dall'Agnese, Alessandra, Boija, Ann, Klein, Isaac A., Coffey, Eliot L.,
558 Shrinivas, Krishna, et al., 'Coactivator Condensation at Super-Enhancers Links Phase
559 Separation and Gene Control', *Science*, 361/6400 (2018), eaar3958

560 Schmiedel, Benjamin J., Singh, Divya, Madrigal, Ariel, Valdovino-Gonzalez, Alan G., White,
561 Brandie M., Zapardiel-Gonzalo, Jose, et al., 'Impact of Genetic Polymorphisms on
562 Human Immune Cell Gene Expression', *Cell*, 175/6 (2018), 1701-1715.e16

563 Sexton, Tom, and Cavalli, Giacomo, 'The Role of Chromosome Domains in Shaping the
564 Functional Genome', *Cell*, 160/6 (2015), 1049–59

565 Simonis, Marieke, Klous, Petra, Splinter, Erik, Moshkin, Yuri, Willemsen, Rob, de Wit, Elzo, et
566 al., 'Nuclear Organization of Active and Inactive Chromatin Domains Uncovered by
567 Chromosome Conformation Capture-on-Chip (4C)', *Nature Genetics*, 38/11 (2006),
568 1348–54

569 Sun, Qiu, Perez-Rathke, Alan, Czajkowsky, Daniel M., Shao, Zhifeng, and Liang, Jie, 'High-
570 Resolution Single-Cell 3D-Models of Chromatin Ensembles during Drosophila
571 Embryogenesis', *Nature Communications*, 12/1 (2021), 205

572 Tang, Fuchou, Barbacioru, Catalin, Wang, Yangzhou, Nordman, Ellen, Lee, Clarence, Xu,
573 Nanlan, et al., 'mRNA-Seq Whole-Transcriptome Analysis of a Single Cell', *Nature*
574 *Methods*, 6/5 (2009), 377–82

575 the 4D Nucleome Network, Dekker, Job, Belmont, Andrew S., Guttman, Mitchell, Leshyk,
576 Victor O., Lis, John T., et al., 'The 4D Nucleome Project', *Nature*, 549/7671 (2017),
577 219–26

578 The GTEx Consortium, Ardlie, Kristin G., Deluca, David S., Segrè, Ayellet V., Sullivan,
579 Timothy J., Young, Taylor R., et al., 'The Genotype-Tissue Expression (GTEx) Pilot
580 Analysis: Multitissue Gene Regulation in Humans', *Science*, 348/6235 (2015), 648–60

581 Tjong, Harianto, Li, Wenyuan, Kalhor, Reza, Dai, Chao, Hao, Shengli, Gong, Ke, et al.,
582 'Population-Based 3D Genome Structure Analysis Reveals Driving Forces in Spatial
583 Genome Organization', *Proceedings of the National Academy of Sciences*, 113/12 (2016)
584 <https://pnas.org/doi/full/10.1073/pnas.1512577113> [accessed 11 March 2024]

585 Yu, Jingting, Hu, Ming, and Li, Chun, 'Joint Analyses of Multi-Tissue Hi-C and eQTL Data
586 Demonstrate Close Spatial Proximity between eQTLs and Their Target Genes', *BMC*
587 *Genetics*, 20/1 (2019), 43

588 Zeng, Ruijie, Wang, Jinghua, Jiang, Rui, Yang, Jie, Zheng, Chunwen, Wu, Huihuan, et al.,
589 'Investigating Causality and Shared Genetic Architecture between Neurodegenerative
590 Disorders and Inflammatory Bowel Disease', *Aging and Disease*, 2022, 0

591 Zhang, Bin, and Wolynes, Peter G., 'Genomic Energy Landscapes', *Biophysical Journal*, 112/3
592 (2017), 427–33

593 Zhang, Jinfeng, Chen, Rong, Tang, Chao, and Liang, Jie, 'Origin of Scaling Behavior of Protein
594 Packing Density: A Sequential Monte Carlo Study of Compact Long Chain Polymers',
595 *The Journal of Chemical Physics*, 118/13 (2003), 6102–9

596 Zhang, ZhiZhuo, Li, Guoliang, Toh, Kim-Chuan, and Sung, Wing-Kin, '3D Chromosome
597 Modeling with Semi-Definite Programming and Hi-C Data', *Journal of Computational*
598 *Biology*, 20/11 (2013), 831–46

599

600