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 genome architecture is important for understanding expression regulation. Additionally, Genome-wide association studies (GWASs) have uncovered thousands of loci related to complex traits. However, the functional roles of many located in non-coding regions like eQTLs are unclear. Current joint analyses of Hi-C and eQTLs data are limited by a lack of advanced computational tools, limiting what can behand from these data.

Result: In this work, we develop a computational method that can 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 ensemble (10^5) of high-resolution single-cell 3D chromatin conformations that accurately replicate Hi-C measurements and reveal complex many-body interactions. Our results show the presence of many-body

interactions in our single-cell chromatin conformation, offering specifics into how the three-dimensional structures of chromatin form the physical foundation for gene regulation, determine the impact of genetic variants, specifically eQTLs, on the expression level of their associated eGenes, as measured by the normalized effect size (NES). Furthermore, our method can quantify chromatin heterogeneity and examines the spatial associations of eQTLs and eGenes, revealing their regulatory impacts on gene expression. Together, Hi-C data and eQTLs provide unique opportunities for deciphering how chromatin 3D structures provide the physical basis for gene regulation, expression control, and 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. pairs of eQTLs and their associated genes, eGene-eQTLs) interactions play important roles in

genome organization and function (Hnisz et al. 2017), but they cannot be directly identified from 2D Hi-C heatmaps. Furthermore, while heatmaps of population Hi-C exhibit highly detailed patterns, the heterogeneities of the chromatin 3D structures in the underlying cell population are difficult to assess. While there are several approaches for modeling 3D chromatin from experimental Hi-C data, including consensus optimization [ShRec3D (Lesne et al. 2014), ChromSDE (Zhang et al. 2013) and Chrom3D (Paulsen et al. 2017)], ensemble optimization [(Baù and Marti-Renom 2012), and PGS (Tjong et al. 2016, Hua et al. 2018)] ,and block copolymer models (Di Pierro et al. 2016, Zhang and Wolynes 2017, Conte et al. 2020, Contessoto, Cheng and Onuchic 2022) (reviewed in (Liang and Perez-Rathke 2021)), adequately sampling diverse chromatin conformations remains challenging (Liang and Perez-Rathke 2021). As chromatin conformations in a cell population may fall into a set of distinct structural clusters (Gürsoy et al. 2017, Sun et al. 2021), where each cluster harbors conformations of similar topology, some clusters may have favorable 3D spatial arrangements of promoter, enhancer, gene, and other elements to facilitate gene expression, while others may not. However, such chromatin heterogeneity and analysis of cell subpopulations with relevant chromatin structural clusters have not been quantified in the eQTLs enriched loci.

In this study, we develop a new computational method for simultaneous analysis of both Hi-C and eQTLs data, overcoming difficulties listed above. Specifically, we construct a pipeline for identifying non-random Hi-C interactions around eQTLs/eGenes that are beyond polymer collision. We further generate large ensembles (5×10^4) of 3D single-cell chromatin conformations models for the selected locus. We then compute subpopulations of single-cell chromatin configurations at the locus and quantified the overall chromatin heterogeneity. Our

tools allow interrogations of the 3D chromatin structures of loci where eQTLs and/or eGenes reside to gain an understanding of how they are spatially associated, and how such spacial relationship may affect regulation of gene expression. Our tools also identify the participating genes, promoters, and other elements in the spatial context of the eQTLs. Furthermore, with 3D ensemble models of their spatial arrangement constructed, our results uncovered higher-order many-body interacting units important for gene regulation of an eQTLs-rich locus across lymphoblastoid cells (GM12878), human primary mammary epithelial cells (HMEC), and lung fibroblast cells (IMR-90). Our results demonstrate the benefits of integrating Hi-C and eQTL data through deep sampling of large ensemble models of single cell chromatin conformations, allowing us to explore critical issues related to the differences in genome 3D structure across tissues and the relationship between genome 3D structure and function.

2 MATERIALS AND METHODS

2.1 Dataset

In this study, we used bulk Hi-C data of lymphoblastoid cells (GM12878, 4DNES3JX38V5), human primary mammary epithelial cells (HMEC, 4DNESIE5R9HS) and lung fibroblast cells (IMR-90, 4DNES1ZEJNRU) from the 4D Nucleome Data Portal (the 4D Nucleome Network et al. 2017, Reiff et al. 2022). We also obtained eQTLs data and bulk tissue expression data for cells of EBV-transformed lymphocytes, breast mammary tissue, and lung tissues from the GTEx Portal (Data Source: GTEx Analysis Release V8) (Lonsdale et al. 2013). All data are based on the reference genome GRCh38.

115

116 **2.2 Overview of the pipeline**

117

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

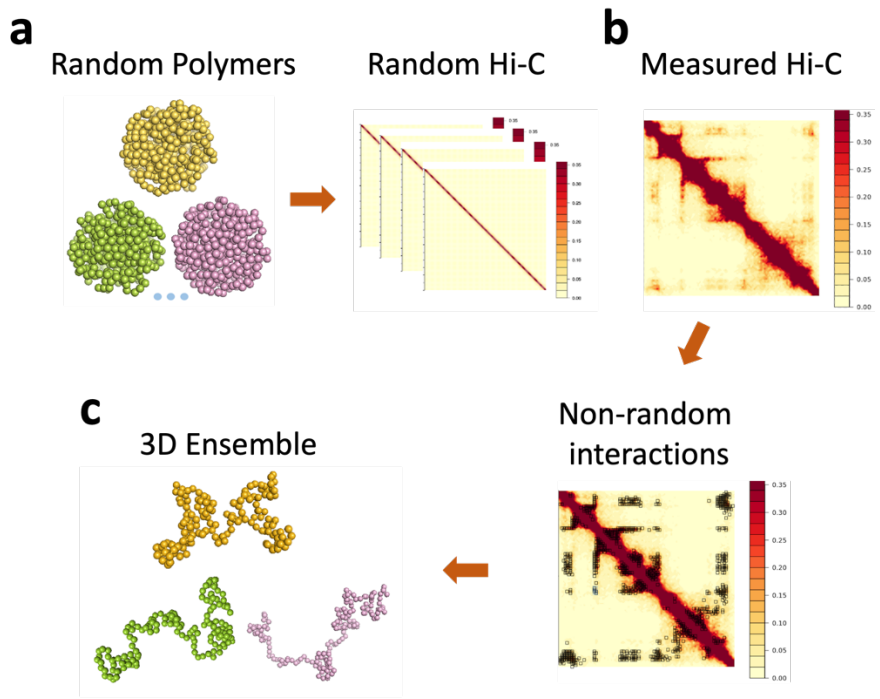


Figure. 1 The pipeline of our computational methods. **a.** A large ensemble of random polymer chains is generated as our background null model. **b.** The statistical significance of each Hi-C interaction is computed by bootstrapping the polymer ensemble for identifying non-random interactions from 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 et al. 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

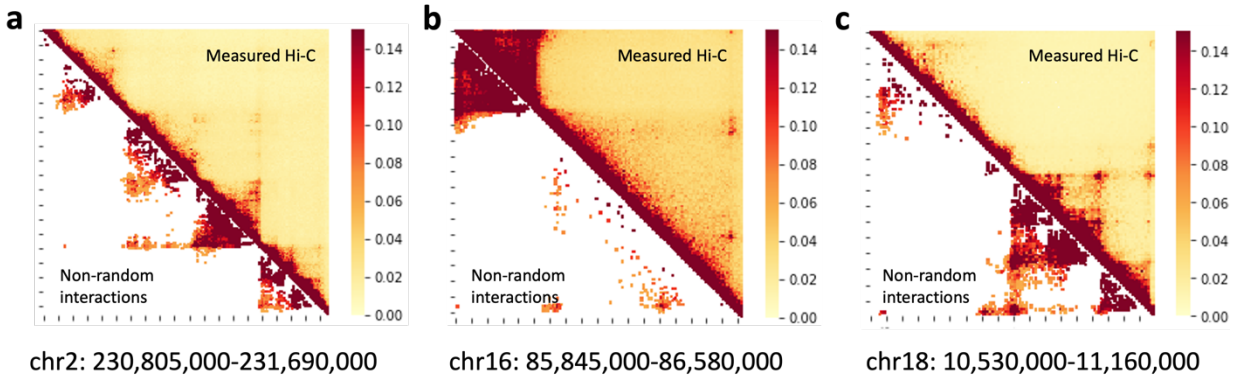
3.1 A small fraction of Hi-C contacts are non-random interactions.

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 containing eQTLs variants associated with ≥ 1 gene for at least one of these tissues. These loci are chr2: 230,805,000-231,690,000 (Locus I), chr16: 85,845,000-86,580,000 (Locus II) and chr18:10,530,000-11,160,000 (Locus III).

Our results show that just a fraction of Hi-C interactions can be considered as non-random. Specifically, 9.8%, 8.9%, and 13.1% of the total contacts for GM12878 cells are identified as

197 non-random interactions for these three loci (Figure 2). 10.1%, 12.9%, and 14.9% of the total
198 contacts for IMR90 cells are identified as non-random interactions and 13.4%, 22.4%, and 31.1%
199 of the total contacts for HMEC cells are identified as non-random interactions (see
200 Supplementary Figure 1). They all meet the specified criteria that adjust *p*-value is less than 0.05.
201 The high proportion of non-random interactions in HMEC cells is due to the low read count of
202 the HMEC Hi-C data. Although they conditionate of only a small fraction of Hi-C interactions,
203 the key structural patterns in the heatmap are well-presented by the non-random interactions in
204 all three loci (Figure 2).
205



206

207 **Figure. 2** Small fractions of Hi-C contacts are non-random interactions in GM12878 cells. The
208 non-random interactions of loci a. chr2: 230,805,000-231,690,000, b. chr16: 85,845,000-
209 86,580,000 and c. chr18: 10,530,000-11,160,000. The percentage of non-random interaction is
210 9.8%, 8.9% and 13.1%, respectively.
211

212 **3.2 Non-random interactions can reconstruct ensemble of single-cell chromatin**
213 **conformation.**
214

To evaluate the importance of these non-random interactions we uncovered, we use these non-random interactions as physical constraints to generate 3D chromatin conformations. We used previously developed efficient chromatin folding algorithms (Sun et al. 2021).

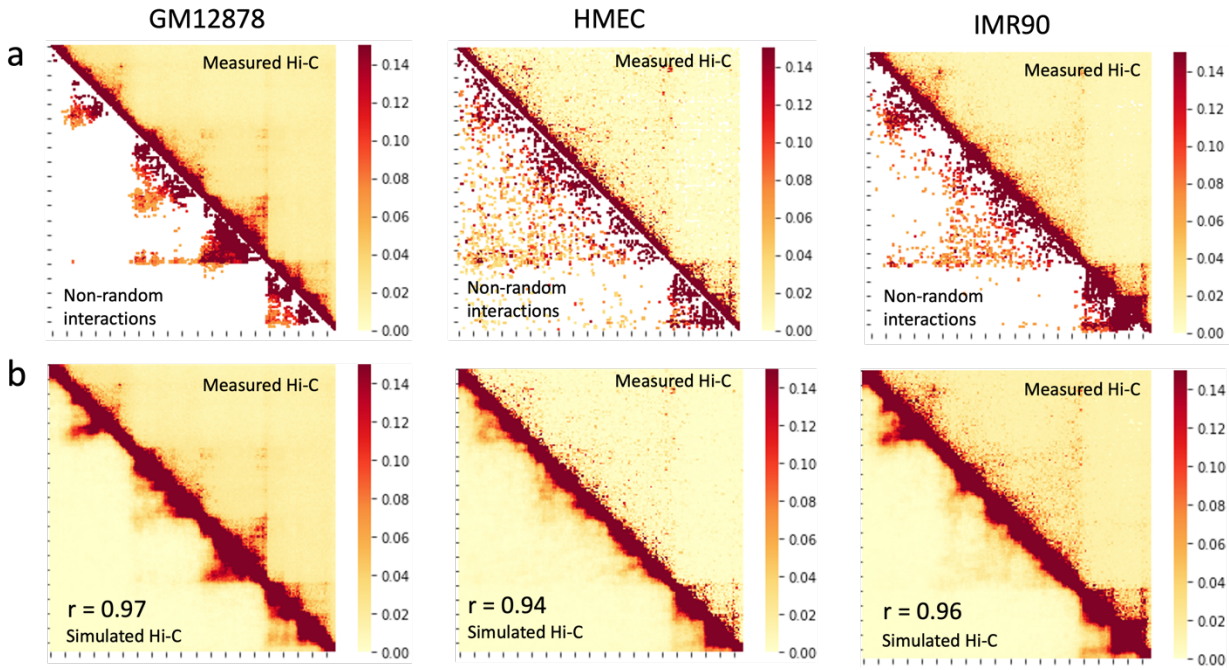
For Locus I, there are 10%-13% of the total contact pairs in GM12878, IMR90, and HMEC (Figure 3a). Using these non-random interactions, we reconstruct large ensembles of 5×10^5 single-cell chromatin conformations for each cell line at 5kb resolution. To compare our predicted ensemble of single-cell chromatin conformation with the experiment-measured Hi-C data, we calculate the interaction probabilities of simulated chromatin conformations. Taking the basic premise that DNA segments near each other are potential candidates for Hi-C ligations, we determined the interaction probabilities based on the frequency of 3D structures where the 3D distances between certain pairs of loci are less than 80 nanometers (Giorgetti et al. 2014, Gürsoy et al. 2017). As shown in Figure 3b, the simulated Hi-C map, obtained by aggregating 10,000 single-cell conformations, exhibits strong similarities with the measured Hi-C maps for GM12878, IMR90, and HMEC for Locus I with Pearson correlation coefficients range from 0.94 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

237 non-random interactions we uncovered are sufficient to guide chromatin folding of these loci in

238 all three cell types.

239



240

241 **Figure 3.** Precise reconstruction of 3D chromatin conformations with high resolution **a**. Non-

242 random interactions identified from our model for GM12878, HMEC, and IMR90 cell lines in

243 Locus I . **b**. Simulated Hi-C map compared to the measured Hi-C map for GM12878, HMEC,

244 and IMR90 cell lines within Locus I. The corresponding Pearson correlation coefficients (R) are

245 0.97, 0.96, and 0.94, respectively.

246

247 **3.3 Uncovering physical contacts between eGene and eQTLs and their strong influence on**

248 **gene expression.**

249

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

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

252 2014). Additionally, eQTLs enriched with cis-regulatory elements are commonly situated in
253 close spatial proximity to the promoters of the genes they target (Aguet et al. 2017, Duggal,
254 Wang and Kingsford 2014). With the 3D single-cell chromatin conformations now at hand, we
255 examined the 3D structures of our selected eQTL containing loci.
256
257 We first analyzed modeled single-cell conformations to determine the occurrence where two 5kb
258 genomic regions represented by polymer beads, one containing eGene and another containing
259 eQTLs are in physical interaction using the criteria that the median distance between them is less
260 than 80 nm (Giorgetti et al. 2014, Gürsoy et al. 2014). Next, we examine all eGene-eQTLs pairs
261 derived from the GTEx database to determine whether they are in physical contacts among all
262 three loci for EBV-transformed lymphocytes (GM12878), breast–mammary tissue (HMEC), and
263 lung cells (IMR90). eGene-eQTLs pairs are then categorized as either in physical contact or are
264 free of such contacts.
265
266 Among the three loci, we found 153 physical eGene-eQTLs contacts from 798 eGene-eQTLs
267 pairs in EBV-transformed lymphocytes, 365 physical eGene-eQTLs contacts from 2,024 eGene-
268 eQTLs pairs in Brest-mammary tissue, and 535 physical eGene-eQTLs contacts from 2,678
269 eGene-eQTLs pairs in lung tissue cells. Our method can reveal pairs of eGenes and eQTLs in
270 physical contact. (For more details on these physically contacted eGene-eQTL pairs, see
271 Supplementary Data.)
272
273 Intriguingly, our results show that physically contacted eQTLs exhibit a stronger influence on
274 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.

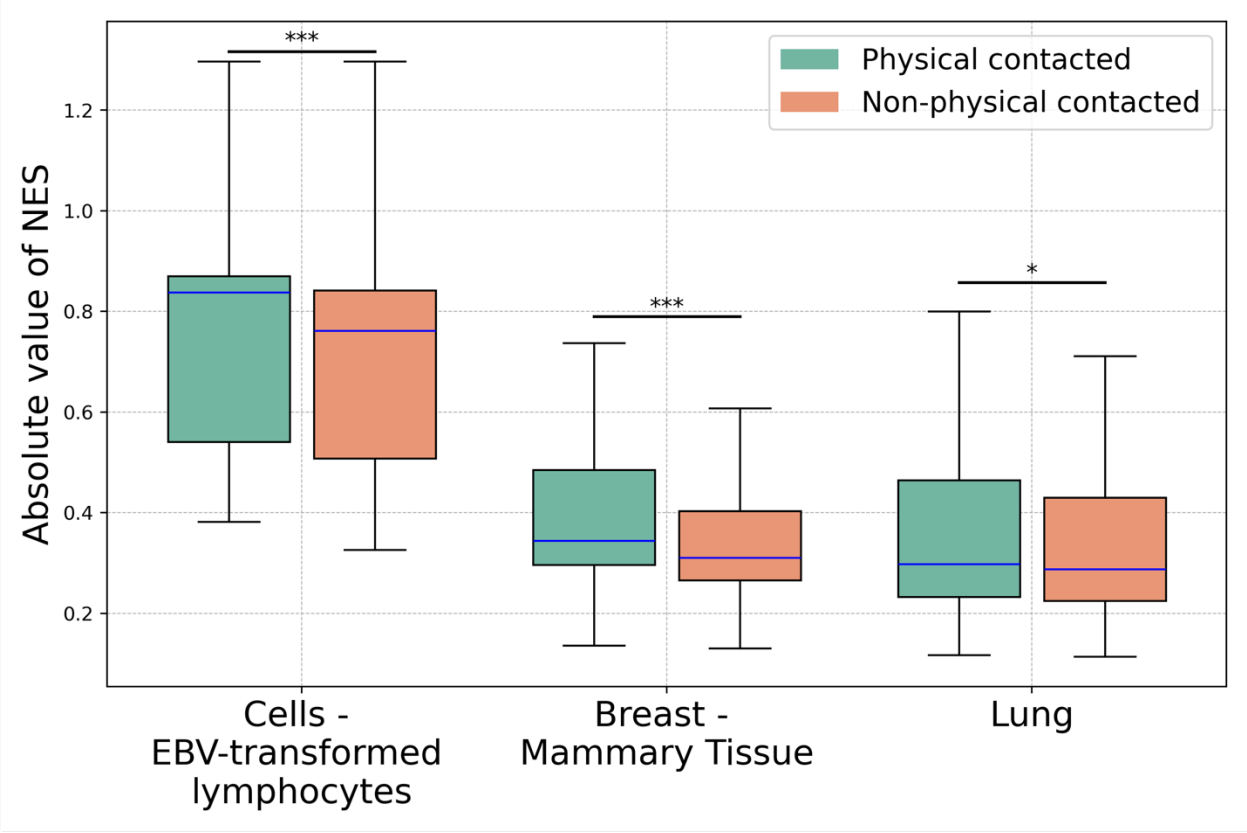


Figure 4. Display of the absolute values of NES for eGene-eQTL pairings across all three examined loci for EBV-transformed lymphocytes (GM12878), breast - mammary tissue (HMEC)

288 and lung (IMR90) cells, respectively. A larger absolute NES value indicates a more substantial
289 genotype impact on gene expression. Significance levels are indicated as follows: ***$p < 0.001$,
290 **$p < 0.01$, *$p < 0.05$.
291
292 **3.4 Discovery of many-body interactions at eGene and eQTLs loci**
293
294 Previous studies found that many-body chromatin interactions such as those in condensates
295 likely play important roles in gene regulations (Gürsoy et al. 2017, Sabari et al. 2018, Quinodoz
296 et al. 2018, Perez-Rathke et al. 2020). However, the presence of many-body interactions within
297 individual cells are masked in the bulk Hi-C measurements, as they capture average pairwise
298 genomic interactions across populations. In order to determine whether many-body interactions
299 occur in eGenes and eQTLs loci, we investigate Locus II (chr16: 85,845,000-86,580,000) across
300 GM12878, HMEC and IMR90 cells, which contains eGenes RP (RP11-463O9.9) and MT
301 (MTHFSD), both are associated with the eQTL of chr16_86531581_A_T_b38 (Figure 5a, b).
302 We examined the spatial structure of genetic interactions among these eGenes and eQTL at this
303 locus.
304
305 Analysis of the 3D chromatin structure shows that three-body interactions do exist at this locus.
306 Here, we define three-body interactions as a complex of 5kb chromatin regions, where the
307 median Euclidean distances between all pairs of regions in the complex of the ensemble of
308 chromatin conformations are shorter or approximately equal to a cross-linking threshold of 80 ±
309 5 nm (Giorgetti et al. 2014, Gürsoy et al. 2017) (Figure 5b). The interactomes among eGenes RP,
310 MT and the eQTL are three-body in both GM12878 and HMEC. However, the degree of

influence of the eQTL variant on gene expression is different among these three tissues. This is indicated by the different values of NES for the same eQTL. The absolute value of NES of this eQTL on eGene MT in lymphocytes (GM12878) is the largest (1.9). The NES values are only 0.64 for breast-mammary tissue (HMEC) and 0.53 for lung tissue (IMR90), respectively. Furthermore, the NES of the same eQTL with eGene RP is also the largest (1.6) in lymphocytes (GM12878). The NES value is 1.4 for breast-mammary tissue (HMEC) and 1.0 for Lung tissue (IMR90), respectively.

To understand the reason why the impact of the variant on gene expression as measured by NES value varies across these three tissues, we ask whether it can be explained by their spacial distance relationship. For this, we randomly selected 5,000 simulated single-cell conformations from the ensembles for GM12878, HMEC, and IMR90 cells and calculated the Euclidean distance between eQTL and the eGenes. From the distributions of the Euclidean distance between eGene and eQTLs, we found that in GM12878 the median spatial distance between the promoter of the eGene MT and the eQTL is 55.74 nm. The distance between the two promoters of the eGenes MT and RP is 82.42 nm. These are closer than the distances in HMEC (61.24 nm, 83.61 nm) and IMR90 (62.88 nm, 87.68 nm). For eGene RP, the spatial distance between its promoter and the eQTL is similar among all three cell types; this is consistent with their shared close genomic distances between the eQTL and eGene RP (<10kb). The shorter 3D physical distances in GM12878 (MT to the eQTL and MT to RP) among the eGenes MT, RP and the eQTL is closer than the corresponding 3D physical distance in HMEC and in IMR90, leading to stronger effects of eQTL on gene expression in GM12878. These results demonstrate that closer

333 spatial proximity between eQTL and the target gene promoters can result in stronger eQTL
334 effects on gene expression, as revealed by the larger absolute value of NES.
335
336 The measured expression level of eGene MT in these tissues provide additional direct support to
337 our findings (Figure 5c). The eQTL with closer spatial distance has stronger effects to the gene
338 expression. The median TPM of gene MT in lymphocytes is the largest at 8.55, followed by 7.27
339 in breast-mammary tissue and 5.97 in lung tissue. Overall, these results show that there exist
340 three-body interactions among eGenes RP, MT and eQTL at Locus II, and the spatial pattern of
341 interactions among the eQTLs and eGenes play important roles in gene regulation and genome
342 functions.

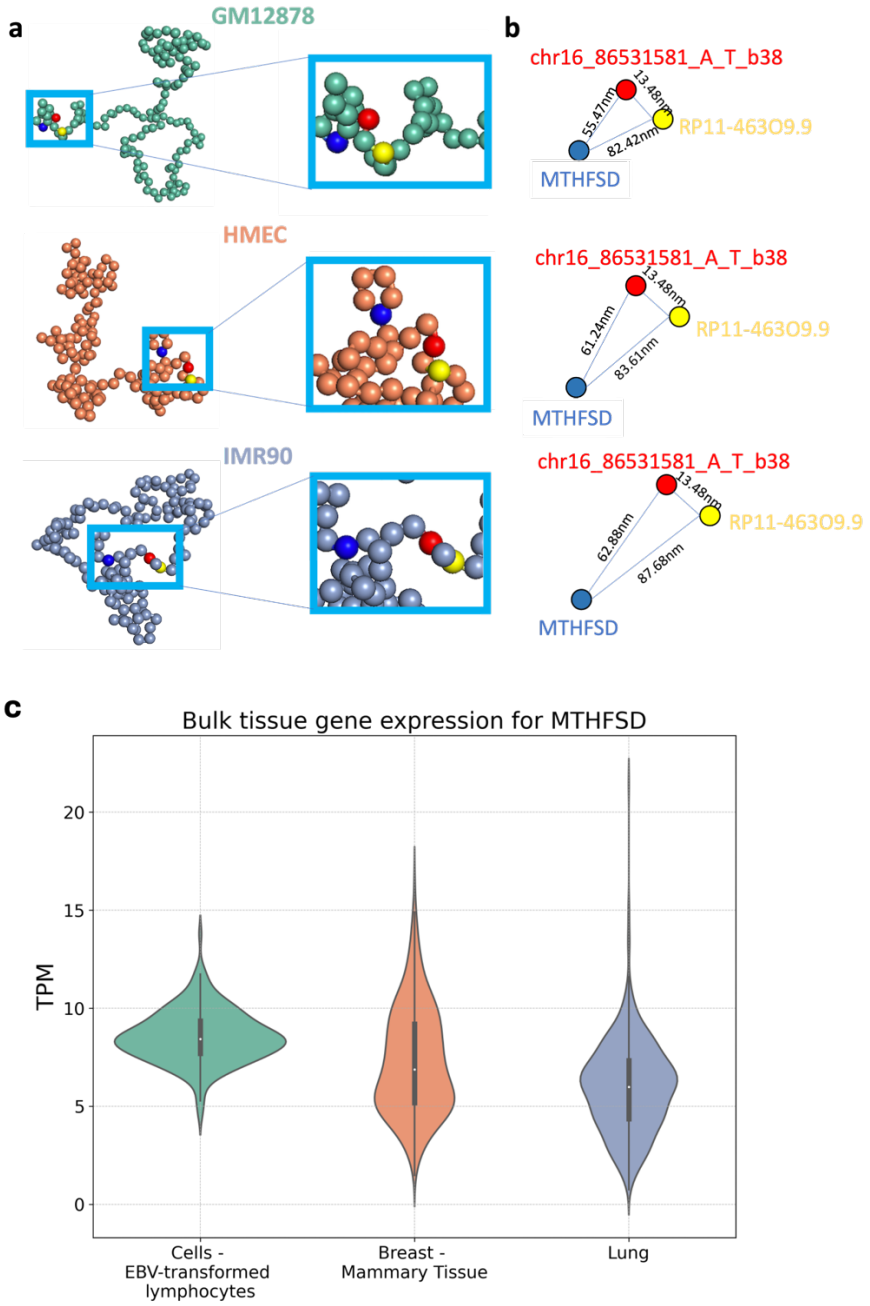


Figure 5. The physical interactome of many-body genetic interactions between eGene and eQTL at Locus II. **a**. Representative spatial structures of the locus for GM12878, HMEC and IMR90, with eGenes MTHFSD (blue), RP11-463O9.9 (yellow) and eQTLs chr16_86531581_A_T_b38 (red) depicted. Note the spacial distances among eGenes and eQTL are different. **b**. The spatial

349 distances between the eGenes MTHFSD (blue), RP11-463O9.9 (yellow), and the eQTLs (red) of
350 the many-body interaction unit in the representative structures shown in (**a**). The spacial
351 distances among the many-bodies are the closest in GM12878, followed by HMEC and then
352 IMR90 **c.** Bulk tissue gene expression level of eGene MTHFSD among the three tissues of
353 transformed lymphocytes (GM12878), breast-mammary tissue (HMEC), and lung (IMR90). The
354 expression level of MTHFSD is the highest in the transformed lymphocytes cells, followed by
355 breast-mammary cells and then lung cells.
356
357 **3.5 3D chromatin structure difference and heterogeneities at eGene and eQTLs loci among**
358 **tissues**
359
360 Quantitatively accessing the chromatin heterogeneities of cells within a tissue is challenging, as
361 chromatin in individual cells may take different configurations, and the overall ensemble may be
362 very heterogeneous. Because of this, independent initial condition through sampling is crucial.
363 With a large of diverse ensemble of single-cell conformations generated, we can group them into
364 clusters by their 3D topology to quantify their heterogeneity (Perez-Rathke et al. 2020, Sun et al.
365 2021). We use the K-means clustering method to cluster the 3D chromatin configurations into
366 major subpopulations of chromatins with similar topologies. For Locus I, we used 10,000
367 simulated single-cell chromatin conformations for each of GM12878, IMR90, and HMEC cells,
368 totaling 30,000 conformations, and were able to group them into four clusters (Figure 6). There
369 are notable distinctions in both the contacts heatmap and compactness among the four
370 subclusters. Importantly, the distribution of subpopulations varies significantly across different
371 cell lines, with HMEC and IMR90 predominantly featuring subclusters 1 and 3, respectively,

while GM12878 enriched with subclusters 2 and 4 cells. That shows GM12878 cells are more heterogeneous compared HMEC and IMR90 cells. Our simulated chromatin conformations Uncover 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).

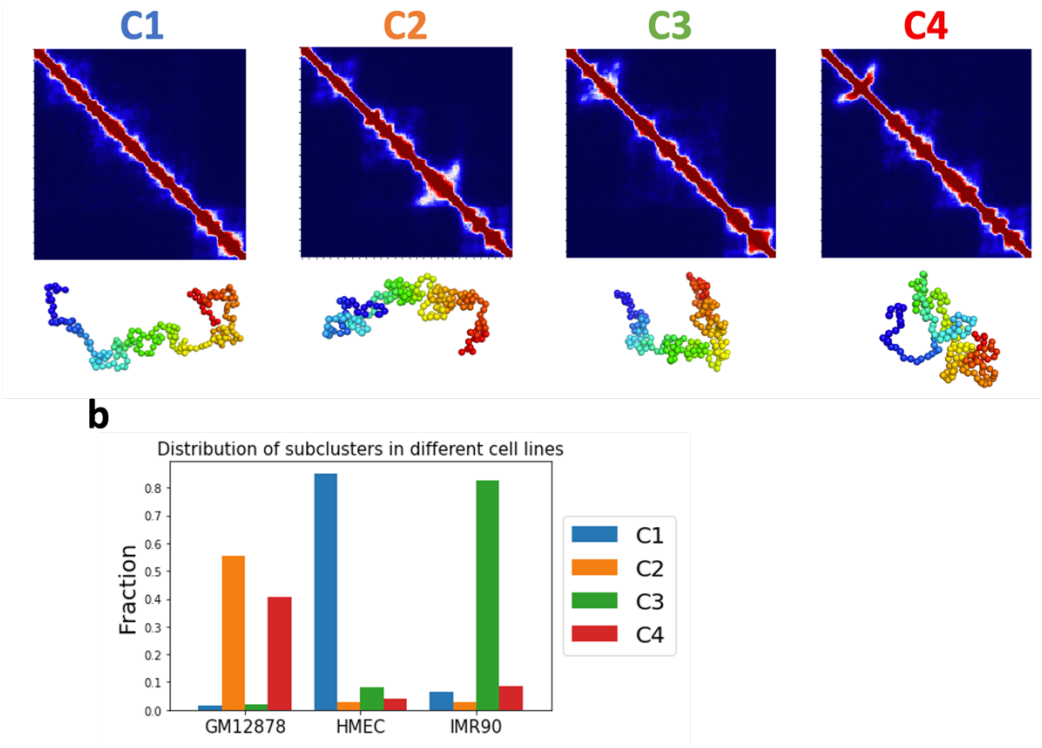


Figure. 6 Uncovering the tissue-specific subpopulations of 3D chromatin structures. **a** 3D chromatin conformation of GM12878, IMR90 and HMEC at Locus I (chr2: 230,805,000-231,690,000). The resulting aggregated contact heatmaps of the four clusters are arranged in order of their compactness. Representative conformations of four clusters are shown below the heatmap. b Proportions of the four subpopulations in each cell types.

386

387 **4 CONCLUSION AND DISCUSSION**

388

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

397

398 Furthermore, our work enables new discoveries as we can now directly investigate the 3D spatial
399 relationship between eGenes and eQTLs across multiple tissues. Results obtained from the
400 analysis of 3D single-cell chromatin conformations of loci across different cell types showed that
401 the physically interacted eQTLs have stronger effects on gene expression. The specific spatial
402 arrangements of eQTLs in cell subpopulations allow a quantity approach to understanding the
403 regulatory mechanisms controlling gene expression. In addition, our tools enable novel and
404 compelling biological questions to be formulated, such as possible roles of specific promoters,
405 enhancers, super-enhancers, genes, and other elements relating to specific eQTLs in multi-way
406 interactions.

407

While we have shown a promising approach to investigate genome structure and function relationship, there are limitation. Firstly, the availability of high-quality Hi-C data is limited. With more high-quality data from the 4DN becoming available recently, this will become less of an issue. Secondly, we analyzed only three loci among three cell lines. It is desirable to expand the scale to analyze additional eQTLs and eGene-associated loci among many more tissues and cell lines. Furthermore, the advancement of single-cell sequencing technologies, such as scRNA-seq (Tang et al. 2009), scATAC-seq (Buenrostro et al. 2015), and scHi-C (Ramani et al. 2017) provided additional data at single cell level. Integrating with these data with 3D chromatin modeling should lead to better understanding of genome structure-function relationship at the single-cell level.

Acknowledgment:

This work is supported by NIH grants 1R03OD032628-01, 1R03OD036492-01 and 2R35GM127084-06. An award for computer time was provided by the U.S. Department of Energy's (DOE) Innovative and Novel Computational Impact on Theory and Experiment (INCITE) Program. This research used resources from the Argonne Leadership Computing Facility, a U.S. DOE Office of Science user facility at Argonne National Laboratory, which is supported by the Office of Science of the U.S. DOE under Contract No. DE-AC02-06CH11357.

Data availability:

430 Non-random interactions, single-cell conformations and physically contacted eGene-eQTLs

431 pairs are available on UIC Indigo (https://indigo.uic.edu/projects/Constructing_High-

432 Resolution_Ensemble_Models_of_3D_Single-

433 Cell_Chromatin_Conformations_of_eQTL_Loci_from_Integrated_Analysis_of_4DN-

434 GTEx_Data_towards_Structural_Basis_of_Differential_Gene_Expression/190164)

435

436 REFERENCES

437

438 Aguet, François, Brown, Andrew A., Castel, Stephane E., Davis, Joe R., He, Yuan, Jo, Brian, et al.,
439 'Genetic Effects on Gene Expression across Human Tissues', *Nature*, 550/7675 (2017),
440 204–13

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

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

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

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

451 Conte, Mattia, Fiorillo, Luca, Bianco, Simona, Chiariello, Andrea M., Esposito, Andrea, and
452 Nicodemi, Mario, 'Polymer Physics Indicates Chromatin Folding Variability across Single-
453 Cells Results from State Degeneracy in Phase Separation', *Nature Communications*, 11/1
454 (2020), 3289

455 Contessoto, Vinícius G., Cheng, Ryan R., and Onuchic, José N., 'Uncovering the Statistical Physics
456 of 3D Chromosomal Organization Using Data-Driven Modeling', *Current Opinion in*
457 *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

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, 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

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

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

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

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

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

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

501 Lonsdale, John, Thomas, Jeffrey, Salvatore, Mike, Phillips, Rebecca, Lo, Edmund, Shad, Saboor,
502 et al., 'The Genotype-Tissue Expression (GTEx) Project', *Nature Genetics*, 45/6 (2013),
503 580–85

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

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

511 Perez-Rathke, Alan, Mali, Samira, Du, Lin, and Liang, Jie, 'Alterations in Chromatin Folding
512 Patterns in Cancer Variant-Enriched Loci', in *2019 IEEE EMBS International Conference*
513 *on Biomedical & Health Informatics (BHI)* (presented at the 2019 IEEE EMBS
514 International Conference on Biomedical & Health Informatics (BHI), Chicago, IL, USA,
515 2019), 1–4 <https://ieeexplore.ieee.org/document/8834565/> [accessed 11 March
516 2024]

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

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

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

527 Ramani, Vijay, Deng, Xinxian, Qiu, Ruolan, Gunderson, Kevin L., Steemers, Frank J., Disteche,
528 Christine M., et al., 'Massively Multiplex Single-Cell Hi-C', *Nature Methods*, 14/3 (2017),
529 263–66

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

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

536 Sanborn, Adrian L., Rao, Suhas S. P., Huang, Su-Chen, Durand, Neva C., Huntley, Miriam H.,
537 Jewett, Andrew I., et al., 'Chromatin Extrusion Explains Key Features of Loop and
538 Domain Formation in Wild-Type and Engineered Genomes', *Proceedings of the National*
539 *Academy of Sciences of the United States of America*, 112/47 (2015), E6456-6465

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

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

545 Simonis, Marieke, Klous, Petra, Splinter, Erik, Moshkin, Yuri, Willemsen, Rob, de Wit, Elzo, et al.,
546 'Nuclear Organization of Active and Inactive Chromatin Domains Uncovered by
547 Chromosome Conformation Capture-on-Chip (4C)', *Nature Genetics*, 38/11 (2006),
548 1348–54

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

552 Tang, Fuchou, Barbacioru, Catalin, Wang, Yangzhou, Nordman, Ellen, Lee, Clarence, Xu, Nanlan,
553 et al., 'mRNA-Seq Whole-Transcriptome Analysis of a Single Cell', *Nature Methods*, 6/5
554 (2009), 377–82

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

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

560 Tjong, Harianto, Li, Wenyuan, Kalhor, Reza, Dai, Chao, Hao, Shengli, Gong, Ke, et al.,
561 'Population-Based 3D Genome Structure Analysis Reveals Driving Forces in Spatial

562 Genome Organization', *Proceedings of the National Academy of Sciences*, 113/12 (2016)
563 <https://pnas.org/doi/full/10.1073/pnas.1512577113> [accessed 11 March 2024]

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

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

570 Zhang, Bin, and Wolynes, Peter G., 'Genomic Energy Landscapes', *Biophysical Journal*, 112/3
571 (2017), 427–33

572 Zhang, ZhiZhuo, Li, Guoliang, Toh, Kim-Chuan, and Sung, Wing-Kin, '3D Chromosome Modeling
573 with Semi-Definite Programming and Hi-C Data', *Journal of Computational Biology*,
574 20/11 (2013), 831–46

575

576