A maximum-entropy model to predict 3D structural ensembles of chromatins from pairwise distances: Applications to Interphase Chromosomes and Structural Variants

Guang Shi[1, 2, *] and D. Thirumalai[3, †]

[1] *Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, USA*
[2] *Current Address: Department of Materials Science, University of Illinois, Urbana, Illinois 61801, USA*
[3] *Department of Chemistry and Department of Physics, University of Texas at Austin, Austin, Texas 78712, USA*

The principles that govern the organization of genomes, which are needed for a deeper understanding of how chromosomes are packaged and function in eukaryotic cells, could be deciphered if the three dimensional (3D) structures are known. Recently, single-cell imaging experiments have determined the 3D coordinates of a number of loci in a chromosome. Here, we introduce a computational method (Distance Matrix to Ensemble of Structures, DIMES), based on the maximum entropy principle, with experimental pair-wise distances between loci as constraints, to generate a unique ensemble of 3D chromatin structures. Using the ensemble of structures, we quantitatively account for the distribution of pair-wise distances, three-body co-localization and higher-order interactions. We demonstrate that the DIMES method can be applied to both small length-scale and chromosome-scale imaging data to quantify the extent of heterogeneity and fluctuations in the shapes on various length scales. We develop a perturbation method that is used in conjunction with DIMES to predict the changes in 3D structures from structural variations. Our method also reveals quantitative differences between the 3D structures inferred from Hi-C and the ones measured in imaging experiments. Finally, the physical interpretation of the parameters extracted from DIMES provides insights into the origin of phase separation between euchromatin and heterochromatin domains.

INTRODUCTION

In little over a decade, a variety of experimental techniques, combined with computational tools and physical modeling, have greatly contributed to our understanding of chromatin organization in a numerous cell types and species [1–9]. These studies have have paved the way towards a deeper understanding of the relation between genome structure and the gene expression [10–14]. The commonly used experimental techniques could be broadly classified as sequence-based or microscopy-based. The former include the Chromosome Conformation Capture (3C) [15] and its variants, including Hi-C [16] and Micro-C [17], which in concert with high-throughput sequencing provide population averaged data for the contact matrix or contact maps (CMs) [18, 19]. The elements of the CM are the average probabilities that two loci separated by a given genome length (s, a linear measure) are in spatial proximity. In order to reveal the cell-to-cell chromatin variations in chromatin conformations, single cell Hi-C (scHi-C) or similar techniques have also been developed [20–24]. These studies, along with methods that utilize a combination of Hi-C and imaging techniques [25], reveal the statistical and heterogeneous nature of chromatin organization. In addition, methods like SPRITE [26] and genome architecture mapping (GAM) [27, 28], which avoid the ligation step in Hi-C [9], have also revealed the organization of chromosomes, including features that are missed in conventional Hi-C methods, such as the higher-order contacts that go beyond pairwise interactions. How to utilize the data contained in the CM to *directly* reconstruct an ensemble of the three-dimensional structures of genome is a difficult inverse problem. Data-driven approaches [29–38] have been advanced to solve the complicated Hi-C to 3D structure problem (see the summary in Ref. [39] for additional related studies and Ref. [40] for a comprehensive overview of the existing methods).

The imaging-based technique is the most direct route for determining the 3D chromosome structures [41–43]. In combination with the fluorescence *in situ* hybridization (FISH) technique [44], imaging experiments have enabled direct imaging of the position of the genomic loci at the single-cell level. The FISH experiments have revealed global genome organization principles, such as the chromosome territories (CT) [44]. Recently developed multi-scale multiplexed DNA FISH-based imaging methods [25, 45–52] have further advanced the field, which have resulted in measurements of the spatial positions of many loci for a large number of cells, thus providing not only glimpses of the structures over a large length scale but also a quantitative assessment of the fluctuations in the cell-to-cell conformations. For instance, the imaging method was used to obtain the locations of ~ 65 loci in a 2 mega base pair (Mbp) region for chromosome 21 (Chr21) [47] for a large number of cells. More recently, the method was further improved to image over ~ 900 targeted loci spread uniformly across the entire chromosome scale (≈ 242 Mbp for Chr2) and over 1,000 genomic loci across all 23 chromosomes [51]. Compartment fea-

* guang.shi.gs@gmail.com
† dave.thirumalai@gmail.com

tures, long-range interactions between loci, and the distribution of the radius of gyration could be directly visualized or calculated using the coordinates of the imaged loci, thus providing direct quantitative information on the nuances of genome organization. Although the resolution in the imaging technique will doubtless increase in the future, currently Hi-C based methods provide higher resolution at the CM level but not at the 3D structural level. Thus, by combining the experimental data from a variety of sources and computational methods unexpected insights about chromosome organization could be gleaned.

Here, we introduce a new method, DIMES (from DIstance Matrix to Ensemble of Structures), that utilizes the mean distance matrix (DM) between loci as input to generate an ensemble of structures using the maximum entropy principle. The data in the two studies [47, 51] are used to quantitatively validate the DIMES method. In order to demonstrate the predictive power of the method, we used DIMES to determine the changes in the organization (expressed as CM) from structural variants (inversion) on Chr1 from the mouse cell line and the effect of single loci deletion on Chr2 from IMR90 cell line. Our approach, when applied to Hi-C (using the HIPPS method [53]) and imaging data, reveals important differences between the two methods in the finer details of the structural ensemble of Chr21.

METHOD

We developed the DIMES method, which utilizes the imaging data, to generate an ensemble of 3D chromosome conformations. The input for our theory is the pair-wise distances between the genomic loci (Fig.1). We seek to find a joint distribution of positions of loci, $P(\{\boldsymbol{x}_i\})$, which is consistent with the squared mean pair-wise distance $\langle ||\boldsymbol{x}_i - \boldsymbol{x}_j||^2 \rangle = \langle r^2_{ij,\mathrm{exp}} \rangle$, where $\langle r^2_{ij,\mathrm{exp}} \rangle$ is the experimentally measured average squared distance between two loci i and loci j. One could also use the average distance instead of the average squared distance as constraints. However, constraining average squared distances is computationally more efficicient because the resulted maximum entropy distribution is a multivariate Gaussian distribution which allows fast sampling.

In general, there are many, possibly infinite, number of such $P(\{\boldsymbol{x}_i\})$ which satisfy the constraints. Using the maximum entropy principle [54, 55], we can find a unique distribution $P^{\mathrm{MaxEnt}}(\{\boldsymbol{x}_i\})$ whose differential entropy is maximal among all possible distributions. We should point out that the maximum entropy principle has been previously used in the context of genome organization [56–58], principally to learn the values of the unknown parameters in a chosen energy function deemed to be appropriate for describing chromosome organization. Recently, Messelink et al showed that the maximum entropy distribution with the constraints of contact frequency can be mapped to a confined lattice polymer model [59]. The Lagrange multipliers that enforce the constraints are interpreted as the contact energies in the Hamiltonian of the polymer with position of each monomer occupying the lattice sites. Here, we use the pair-wise distances as constraints and derive the corresponding maximum entropy distribution, and from which the 3D structures may be readily obtained.

The maximum entropy distribution $P^{\mathrm{MaxEnt}}(\{\boldsymbol{x}_i\})$ is given by,

$$P^{\mathrm{MaxEnt}}(\{\boldsymbol{x}_i\}) = \frac{1}{Z}\exp\Big(-\sum_{i<j} k_{ij}||\boldsymbol{x}_i - \boldsymbol{x}_j||^2\Big), \quad (1)$$

where Z is the normalization factor, and k_{ij}'s are the Lagrange multipliers that are determined so that the average value $\langle ||\boldsymbol{x}_i - \boldsymbol{x}_j||^2 \rangle = \langle r^2_{ij,\mathrm{exp}} \rangle$. It can be proven that for any *valid* $r^2_{ij,\mathrm{exp}}$, there exists a unique set of k_{ij} (Supplementary Note 2). The values of k_{ij} can be determined using an iterative scaling algorithm [60] or a gradient descent algorithm (Supplementary Note 3). For later reference, we define the matrix with elements k_{ij}'s as $\boldsymbol{K}$. Note that Eq.1 has the same form as the Boltzmann distribution of the generalized Rouse model (GRM) with $k_{ij} \geq 0$, which has been applied to reconstruct chromosome structure by fitting to the Hi-C contact map [61, 62]. However, it is important to point out that Eq.1 is derived under the maximum entropy principle, which does not assume thermal equilibrium condition of the system.

The three steps in the DIMES to generate an ensemble of chromosome structures are: First, we compute the target mean squared spatial distance matrix from experimental measurements of the coordinates of genomic loci. Second, using an iterative scaling or gradient descent algorithm, we obtain the values of k_{ij}'s to match the experimental measured $\langle r^2_{ij,\mathrm{exp}} \rangle$. Third, using the values of k_{ij}, the coordinates of the 3D chromosome structures can be sampled from $P^{\mathrm{MaxEnt}}(\{\boldsymbol{x}_i\})$ - a multivariate normal distribution. The details of the procedures are described in the Supplementary Note 3.

RESULTS

Validating DIMES: In order to demonstrate the effectiveness of DIMES, we first used the experimental data [47] in which the authors reported, using highly multiplexed super-resolution imaging approach, coordinates of about 65 individual loci for the 2Mbp segment for Chromosome 21 (Chr21) from four different cell lines (IMR90, K562, HCT116, and A549). Using the calculated mean spatial distance matrices from the reported coordinates as targets, we determined k_{ij} (Eq. 1), which allowed us to generate an ensemble of structures for this 2Mbp segment. For all cell lines, the mean spatial distance matrices computed from the reconstructed ensemble of structures almost perfectly match the target distance maps (Fig.2(a-b) and Supplementary Fig. 9). In addition, we also applied the DIMES to the chromosome-scale data

(see a later section), and achieved the same level of accuracy.

We then perform cross-validation of the DIMES method. This is done as follows: a randomly chosen fraction of pairwise distances is from the distance map is deemed to be missing data. The new distance map containing the missing data is then used as input for DIMES. It is important to note that the missing data is not used to update k_{ij}. Finally, the predicted distances for the missing data are compared with the values obtained from the full distance map. Fig. 2(c-f) compare the input distance maps with missing data and the full predicted distance maps. Remarkably, DIMES quantitatively predicts pairwise distances for the missing data even if only 10% of the distance map is used. Together, these results demonstrate that the model is effective in producing the 3D structures that are consistent with the experimental input and is robust with respect to missing data.

Distribution of pair-wise distances

Next, we tested if DIMES could recover the properties of the genome organization that are not encoded in the mean spatial distances. We focused on reproducing the distributions of pairwise distances, which can be calculated because pairwise distance data are available for a large number of individual cells. It is worth emphasizing that the input in the DIMES method is the mean spatial distance, which does not contain any information about the distributions.

To quantitatively measure the degree of agreement between the measured and calculated distance distributions using DIMES, we compute the Jensen-Shannon divergence (JSD), defined as, $(D(p||m) + D(q||m))/2$ where p and q are two probability vectors and $D(p||m)$ is the Kullback-Leibler divergence. The JSD value is bounded between 0 and 1. A zero value means that the two distributions are identical. For each loci pair (i, j), we calculated the JSD, thus generating the JSD matrix. Fig.3a shows the JSD matrix, and Fig.3b displays the histogram of the all the calculated JSDs. The average value of JSDs is merely 0.02, which shows that the overall agreement between the calculated and measured distributions of distances is excellent.

Upon closer inspection of Fig.3b, we find that the values of JSDs are not randomly distributed. We choose two pairs with relatively large and small JSD values. Comparison between the experimental and calculated $P(r_{ij})$ for one pair with JSD =0.08 (Fig.3c) and with =0.009 (Fig.3d) shows that the dispersions are substantial. The pair with JSD =0.08 (Fig.3c) samples distances that far exceed the mean value, which implies that there are substantial cell-to-cell variations in the organization of chromosomes [47, 51]. The percentage of the subpopulations, associated with different distances, can in principle be inferred by deconvolution of the full distance distribution [63].

Co-localization of three loci and biological significance

We next asked if the method accounts for higher-order structures, such as three way contacts, discovered in GAM [27, 28] and SPRITE experiments [26] and imaging experiments [47] , and predicted by the theory [64, 65]. First, we computed the probability of co-localization of loci triplets, $\pi_{ijk}(a) = \Pr(r_{ij} < a, r_{ik} < a, r_{jk} < a)$ where a is the distance threshold for contact formation ($r_{ij} < a$ implies a contact). To make quantitative comparison with experiments, we also computed $\pi^{\rm exp}_{ijk}(a)$ using the experimental data. We then calculated the Pearson correlation coefficient, ρ, between $\pi^{\rm exp}_{ijk}(a)$ and $\pi^{\rm sim}_{ijk}(a)$. Fig.4a shows that the degree of agreement between experiment and theory is best when a is in the range of 200 nm to 400 nm. We chose $a = 300$nm at which ρ is a maximum. The scatter plot of $\pi^{\rm DIMES}_{ijk}$ versus $\pi^{\rm exp}_{ijk}$ (Fig.4(a)) is in excellent agreement with $\rho(\pi^{DIMES}_{ijk}, \pi^{exp}_{ijk}) \approx 0.99$.

The good agreement is also reflected in the Fig.4(b), which compares the heat maps $\pi^{\rm exp}_{ijk}$ for $i = 11$ (lower triangle) and the heat map for $\pi^{\rm DIMES}_{ijk}$ with $i = 11$ (upper triangle). Due to the polymeric nature of the chromatin, the absolute value of π_{ijk} may not be instructive because π_{ijk} is usually highest when the three loci i, j, k are close along the sequence. To normalize the the genomic distance dependence, and capture the *significance* of the co-localization of triplets, we calculated the Z-score for π_{ijk} defined as $Z_{ijk} = (\pi_{ijk} - \mu(\pi_{ijk}))/\sigma(\pi_{ijk})$ where $\mu(\pi_{ijk}) = \sum_{m,n,q} \delta(|j-i||k-j| - |m-n||n-q|)\pi_{mnq} / \sum_{m,n,q} \delta(|j-i||k-j| - |m-n||n-q|)$, and $\sigma(\pi_{ijk})$ is the corresponding standard deviation. Positive Z-score implies that the corresponding triplet has a greater probability for co-localization with respect to the *expected* value. As an example, comparison of the Z-scores for $Z^{\rm exp}_{ijk}$ and $Z^{\rm DIMES}_{ijk}$ in Fig.4(c) for $i = 11$ shows excellent agreement. The scatter plot in Fig.4(d) of $Z^{\rm DIMES}_{ijk}$ versus $Z^{\rm exp}_{ijk}$ shows that the triplet of loci with index $(1, 11, 31)$ has highest value of Z_{ijk} both in the experimental data and the predictions based on DIMES. Three randomly selected individual conformations with $(1, 11, 31)$, co-localized within distance threshold $a = 300$ nm (Fig.4(e)) adopt diverse structures, attesting to the heterogeneity of chromosome organization [25].

To demonstrate the biological significance of the triplet with Z_{ijk} values, we overlay ten sets of three-way contacts with ten largest Z_{ijk} values on the mean spatial distance map (Fig.4(f), with orange circles representing the triplet loci). Interestingly, the three-way contacts are localized on the boundaries of the Topologically Associating Domains (TADs) [66, 67], which are enriched with CTCF motifs [66, 68]. Comparing the location of the triplets and the CTCF peak track (plotted using Chip-seq data [69]) we find that the spatial localization of the

triplets are highly correlated with the CTCF peaks. This implies that the CTCF/cohesin complex have tendency to co-localize in the form of triplets or possibly higher-order multiplets, which is consistent with the recent experiment studies demonstrating the presence of foci and clusters of CTCF and cohesin in cells [70, 71].

Higher-order structures

Because the DIMES method is quantitatively accurate, we could probe higher-order chromatin structures. To this end, we considered three aspects of chromatin organization, which can be calculated directly from the coordinates of the Chr21 loci.

TAD-like patterns in single cell: Imaging experiments [47] show that even in a single cell, domain-like or TAD-like structures may be discerned, even in inactive X chromosomes whose ensemble Hi-C map appears to be featureless [72]. These signatures are manifested as TAD-like patterns in the pairwise distance matrix (see the left panels for two cells in Figure 4B in [47]). Similar patterns are observed in the individual conformation generated by DIMES as well (Supplementary Fig.1). We then compute the boundary probabilities [47] of individual genomic loci using experimentally measured structures and the structures generated by DIMES. Fig. 5a shows that DIMES captures the profile of boundary probabilities nearly quantitatively. Interestingly, such TAD-like structures are present even in ideal homopolymer structures (Supplementary Fig.3). We surmise that the intrinsic features of fluctuating polymer conformations contribute to such TAD-like structures. These structures are dynamic in nature because they are a consequence of fluctuations. The specific interactions which distinguish chromosomes from an ideal homopolymer counterpart lead to the statistically preferred distributions of these TAD-like structures rather than being random.

Size and shapes: Using the ensemble of Chr21 structures, we wondered if the radius of gyration (R_g) and the shape of the genome organization could be accurately calculated. We determined the R_g distribution, $P(R_g)$, and the shape parameter κ^2 [73, 74], for the Chr21 in the 28 Mbp - 30 Mbp region. The results (Fig.5b, and Supplementary Fig. 10) show that the model achieves excellent agreement with experiment, both for $P(R_g)$ and $P(\kappa^2)$.

Given the excellent agreement, it is natural to ask whether the DIMES method can capture the size and shape on finer scales. To shed light on this issue, we calculated the distribution of R_g and κ^2 ($P(R_g; i, j)$ and $P(\kappa^2; i, j)$) for every sub-segment (i loci to j loci) over the 2 Mbp region. In order to assess the accuracy of the predictions, we calculated the JSD between $P^{\mathrm{DIMES}}(R_g; i, j)$ and $P^{\mathrm{exp}}(R_g; i, j)$, and between $P^{\mathrm{exp}}(\kappa^2; i, j)$ and $P^{\mathrm{DIMES}}(\kappa^2; i, j)$. On the finer scale, there are deviations between calculations based on DIMES and experiment. The JSD heat map (Fig.5c) for all pairs i and j shows that the deviation is not uniform throughout the 2 Mbp region. We picked two regions which show good agreement (red segment) and poor agreement (blue segment) in Fig.5c. Direct comparison of $P(R_g)$ and $P(\kappa^2)$ for these two segments is presented in Fig.5d. For the blue segment, both the predicted $P(R_g)$ and $P(\kappa^2)$ have less dispersion than the experimental data. This suggests that heterogeneity observed in experiments is even greater than predicted by the DIMES method. For the red segment, the predicted and experimentally measured are in good agreement. Visual inspection of Fig.5c suggests that the discrepancy is localized mostly in the TAD regions, which might be due to the dynamic nature of these sub-structures.

Ensemble of structures partition into clusters: We then compared the overall distributions of the ensemble of calculated structures with experiment data. To do this, we first performed t-SNE to project the coordinates of each conformation onto a two-dimensional manifold using the distance metric, D_{mn},

$$D_{mn} = \sqrt{\frac{1}{N^2} \sum_{i,j} \left(r_{ij}^{(m)} - r_{ij}^{(n)} \right)^2} \tag{2}$$

where $r_{ij}^{(m)}$ and $r_{ij}^{(n)}$ are the Euclidean distances between the ith and j^{th} loci in conformations m and n, respectively. Based on the density of the t-SNE projections (shown as contour lines in Fig. 6), it is easy to identify two peaks, implying that the space of structures partition into two major clusters. The points are then clustered into two major clusters (orange and blue) using Agglomerative Clustering with the Ward linkage [75]. The percentage of the cluster #1 (orange) in the experiment is 36% , which is in excellent agreement with the value (34%) predicted by the DIMES method. We display a representative structure with the lowest average distance to all the other members in the same cluster for the two clusters in Fig. 6. Based on our analyses of the experimental results (conformations on the left in Fig.6), the representative structure belonging to cluster #1 (in blue) is more compact compared to the one in cluster #2 (in orange). The same trend is found in the DIMES predictions (conformations on the right in Fig.6).

We also computed the mean distance maps from all the conformations from each cluster (shown on the left and right side in Fig.6). The distance maps show that the structures in cluster #2 adopt a dumbbell shape whereas those belonging to cluster #1 exhibit no such characteristic. The quantitative agreement between the distance maps from the experiment and the model is excellent.

Structures of the 242Mbp-long Chr2

We extend the DIMES method to chromosome-scale imaging data in order to compare with the recent super-resolution imaging experiments, which reported coordinates of 935 loci genomic segments with each locus being

50-kbp long spanning the entire 242-Mbp Chr2 of Human IMR90 cell [51]. Note that there are spaces between the loci that are not imaged. We computed the average squared distance matrix from the measured coordinates and then used DIMES to generate an ensemble of structures. Fig.7a compares the experimental and calculated mean distance matrices using the DIMES method (Fig.7a). As before, the agreement is excellent (see also Fig.2). A few randomly chosen conformations from the ensemble are shown in the Supplementary Fig.2a, demonstrating that there are large variations among the structures. Experimental single-cell distance maps (Supplementary Fig. 8a) show that similar variations are also observed *in vivo*. DIMES reproduces the distributions of pairwise distances at large length scales (Supplementary Fig. 8b).

In addition to recovering the experimental data, our approach produces genomic distance (s) dependent effective interaction strengths between the loci, which gives insights into the organization of Chr2 on genomic length scale. Because the parameters in the DIMES are k_{ij}'s, which could be interpreted as effective "interaction" strengths between the loci, we asked if the $\boldsymbol{K}$ matrix encode for the A/B compartments (the prominent checkerboard pattern in Hi-C experiments indicating phase separation between euchromatin and heterochromatin) observed in the distance matrix. (Although k_{ij}s may not represent the actual strength associated with interactions between i and j, we use this terminology for purposes of discussion.) Note that k_{ij} can be either negative or positive, with negative (positive) value representing the effective repulsion (attraction). The lower triangle in Fig.7b shows the matrix $\boldsymbol{K}$. We then computed the correlation matrix $\boldsymbol{\rho}$ from k_{ij} (see the Supplementary Note 4 for details), which is shown in the upper triangle in Fig.7b. The corresponding principal component dimension 1 (PC1) of ρ is shown in the top panel in Fig.7b. The negative (positive) PC1 corresponds to A (B) compartments. The results in Fig. 7b show that the A/B compartments can be inferred directly from $\boldsymbol{K}$, indicating that the parameters in DIMES correctly capture the underlying characteristics of the genome organization on all genomic length scales.

Given that the A/B type of each genomic loci are unambiguously identified, we then computed the histogram of k_{ij} and genomic-distance normalized $\langle k_{ij}(s)\rangle = (1/(N-s))\sum_{i<j}^{N}\delta(s-(j-i))k_{ij}$ for A-A, B-B, and A-B interactions. The results show that the mean A-B interactions are repulsive (negative $\langle k_{ij}(s)\rangle$) whereas A-A and B-B interactions are attractive (positive) (Fig.7c). This finding explains the compartment features observed in the Hi-C data and the distance maps. Furthermore, we find that, on an average, A-A interactions are more attractive than B-B interactions (Fig.7c), which seems counter intuitive because of the general lore that heterochromatin (formed by B locus) appears to be denser than euchromatin (composed of locus A) in microscopy experiments [41]. On the other hand, the same analysis on Chromosome 21 shows opposite results, with the B-B interactions being stronger on an average than A-A interactions (Supplementary Fig. 4b). The genomic-distance normalized $\langle k(s)\rangle$ also shows that over the range of $2-10$ Mbp, $\langle k(s)\rangle$ for B-B is consistently stronger than that for A-A (Supplementary Fig.4c,d). The results for Chr2 and Chr21 show that the comparison between the strength of A-A and B-B interactions are possibly chromosome-dependent. As we noted in a previous study [8], what is important is that the Flory $\chi = (\epsilon_{AA} + \epsilon_{BB} - 2\epsilon_{AB})/2$ is positive to ensure compartment formation. Here, ϵ_{AA}, ϵ_{BB}, and ϵ_{AB}, are the interaction energy scales involving A and B.

The genomic-distance normalized $\langle k_{ij}(s)\rangle$ for Chr2 shows that all the interaction pairs have the highest value at $s = 1$ – a manifestation of the polymeric nature of chromatin fiber. Beyond $s = 1$, all pairs have negative $k(s)$, indicating repulsive interaction on small length scale. At $s \approx 5$Mbp, $k(s)$ for all pairs develop positive peaks. At length scale $s > 10$Mbp, the B-B interactions decay as s increases whereas s-dependent A-A interactions fluctuate around a positive value. In addition, the histograms of k_{ij} (Fig.7c) suggest that the average differences among A-A, B-B, and A-B interactions are small ($\langle k_{AA}\rangle = 0.0014$,$\langle k_{BB}\rangle = 0.0005$, $\langle k_{AB}\rangle = -0.0009$), which is consistent with the recent liquid Hi-C experiment [76]. In summary, the results in Fig.7 demonstrate that the application of DIMES to large-length-scale imaging data explains the origin of compartments on large length scales. More importantly, the calculated values of k_{ij} provide insights into interactions between the genomic loci on all length scales, which cannot be inferred solely from experiments. Surprisingly, but in accord with experiments [76], the differences in the strengths of interaction between the distinct loci are relatively small.

Besides demonstrating the efficacy of DIMES in determining the ensemble of structures accurately, the calculated k_{ij}s explain micro phase segregation of active (A) and inactive (B) loci. That phase separation between A and B emerges from the calculated k_{ij}s, without a polymer model with an assumed energy function, is a surprise. Because k_{ij}s can be calculated using either the HIPPS [53] or the DIMES method, the differences in compartment formation (segregation between A and B loci) in various chromosomes can be quantitatively inferred.

Applications

Impact of genomic rearrangement on 3D organization

It is known that the 3D organization of chromosomes can change substantially upon genomic rearrangements, such as duplication (increases the length of the genome), deletion (decreases the genome length) or inversion (shuffling of genome sequence while preserving the length). Both deletion and duplication are drastic genomic changes that require recomputing the $\boldsymbol{K}$ using

either contact [53] or distance maps. In contrast, the inversion is a gentler perturbation, and hence the changes in chromosome folding compared to the wild type (WT) can be calculated by treating it as perturbations applied to the $\boldsymbol{K}$ for the WT.

Once the WT $\boldsymbol{K}$ is calculated, a perturbation method could be used to determine the variations in the ensemble of genome structures. In particular, we asked whether the 3D structural changes upon rearrangement in the genomic sequence could be predicted by accounting for the corresponding changes in k_{ij}. For instance, inversion (Fig. 8a) would correspond to an inversion on the $\boldsymbol{K}$ (see Supplementary Note 5). For illustration purposes, we applied our method to the experimental Hi-C maps for the WT and a variant with an inversion [77]. To apply DIMES to the two constructs, we first converted the contact probability to the mean spatial distance using the scaling relation $\langle r_{ij}\rangle = \Lambda p_{ij}^{-1/\alpha}$ where we take $\alpha = 4$ [53]. Fig. 8a shows excellent agreement between the experimentally measured Hi-C contact map with inversion and the one predicted by the combined HIPPS-DIMES approach. The more drastic structural variants (deletion and insertion) requires applying the HIPPS or the DIMES method directly to the mutated CMs or DMs.

Structural integrity is determined by loci at the CTCF anchors

and A/B boundary Does deletion of every single locus has the same effect on the 3D structures? To answer this question, we first investigated the effect of deletion of CTCF/cohesin anchors by applying DIMES to Chr21 28 Mbp-30 Mbp region [47]. The method for deletion of single locus is described in Supplementary Note 5. The 3D structural changes are quantified using the Perturbation Index (PI),

$$\mathrm{PI} = \sqrt{\left(\sum_{i<j}^{N}(\langle\tilde{r}_{ij}\rangle - \langle r_{ij}\rangle)^2\right) / \sum_{i<j}^{N}\langle\tilde{r}_{ij}\rangle^2} \qquad (3)$$

where $\langle\tilde{r}_{ij}\rangle$s is the WT average distance between loci i and j and $\langle r_{ij}\rangle$s are changed values after locus deletion. PI profile along the 2 Mbp region and the Chip-seq data for CTCF are shown in Fig. 8b top and middle panels. More importantly, on an average the PI profile exhibits higher values closer to the CTCF Chip-seq peaks (Fig. 8b bottom panel), demonstrating that the genomic loci associated with CTCF anchors have a more significant effect on the 3D structures upon their deletion.

We then probe the prediction of locus deletion for Chr2 imaging data [51] (Fig. 8c). The PI profile along Chr2 (Fig.8c top panel) shows that there are large variations among the individual loci, suggesting that deletion of some genomic locus have a larger impact on the 3D structures than others. To ascertain whether the variations in the PI are associated with the known chromosomal structural features such as A/B compartments, we compared the values of PI with the principal component dimension 1 (PC1) which are computed from correlation matrix $\boldsymbol{\rho}$ (Fig.8d). The loci with PC1 close to zero are interpreted as the A/B boundaries. We find that statistically the boundary elements between A/B compartments have higher PI values, indicated by the basin near $\ln(\mathrm{PI}) = -1$, whereas elements inside the A/B compartments have lower PIs $\ln(\mathrm{PI}) \approx -2$). From this finding, we propose that boundary loci are most important in maintaining chromosome structural integrity.

Comparing Hi-C and imaging experiment

Although there are several scHi-C experiments, the majority of the studies report Hi-C data as ensemble averaged contact maps. On the other hand, super-resolution imaging experiments directly measure the coordinates of loci for each cell. It is unclear how the chromosome structures inferred from Hi-C differs from the ones directly measured in imaging experiments. We compare the Hi-C data [1] and the imaging data [51] for Chr2 from the IMR90 cell line. Both he HIPPS and the DIMES methods first convert the contact probability p_{ij} to mean spatial distance $\langle r_{ij}\rangle$ using $\langle r_{ij}\rangle = \Lambda p_{ij}^{-1/\alpha}$. We determined the value of $\Lambda \approx 0.36\mu m$ and $\alpha \approx 5.26$ by minimizing the error between the distances inferred from Hi-C and the experimental measurements from imaging experiments, $\chi = (2/N(N-1))\sum_{i<j}^{N}(\langle r_{ij}^{\mathrm{Hi-C}}\rangle - \langle r_{ij}^{\mathrm{Imaging}}\rangle)^2$. Fig.9a compares the mean distance matrix inferred from Hi-C using the HIPPS method and the one computed using the coordinates directly measured in the imaging experiment. Visual inspection suggests that the distance matrix inferred from Hi-C shows stronger compartmental patterns compared to the imaging result even though on an average the mean pairwise distances r_{ij}'s obtained from the two methods are in good agreement with each other (Fig. 9b). We also find that the locations of A/B compartments obtained from both the experimental methods are in excellent agreement (Fig. 9c).

Although the positions of the A/B compartments obtained from both Hi-C and imaging data agree with each other, it is unclear whether these two methods could be used to generate the 3D structures that are consistent with each other. To ascertain the 3D structures inferred from Hi-C data and the imaging experiments are consistent with each other, we calculated Q_k and F_k [53]; Q_k measures the degree of spatial mixing between A and B compartments.

$$Q_k = \frac{1}{N}\sum_{i} |n_A(i;k)/\tilde{n}_A - n_B(i;k)/\tilde{n}_B|, \qquad (4)$$

where k is the number of the nearest neighbors of loci i. In Eq.4, $n_A(i;k)$ and $n_B(i;k)$ are the number of neighboring loci belonging to A compartment and B compart-

ment for loci i out of k nearest neighbors, respectively $[n_A(i;k) + n_B(i;k) = k]$. With $N = N_A + N_B$, the expected number of k neighboring loci in the A compartment with random mixing is $\tilde{n}_A = kN_A/N$ and $\tilde{n}_B = kN_B/N$ where N_A and N_B are the total number of A and B loci, respectively. With $k \ll N$, perfect mixing would result in $Q_k = 0$, and $Q_k \neq 0$ indicates demixing between the A and B compartments.

The function, F_k, not unrelated to contact order, quantifies the multi-body long-range interactions of the chromosome structure. We define F_k as,

$$F_k = \frac{1}{kNF_{0,k}} \sum_i \sum_{j \in m_i(k)} |j - i| \tag{5}$$

where k again is the number of nearest neighbors and $m_i(k)$ is the set of loci that are k nearest neighbors of loci i; $F_{0,k}) = (1/2)(1 + k/2)$ is the value of F_k for a straight chain. Eq.5 implies that the presence of long-range interaction increases the value of F_k. In both Q_k and F_k, k is the number of nearest neighbors for a given locus.

Fig. 9d shows that, compared to the results obtained through imaging, the Hi-C method overestimates the extent of long-range interactions, and underestimates the spatial mixing between A and B compartments, which is reflected in the shift of the distribution of $P(Q_k)$ and $P(F_k)$ (we chose $k = 8$ without loss of generality). We also computed the interactions profiles of A-A, B-B, and A-B in the same fashion as shown in Fig.7c,d from the Hi-C data. The Hi-C data suggest that the B-B interactions are more attractive than A-A, which is opposite of the results obtained from the imaging data. Furthermore, the extracted $\langle k(s) \rangle$ for s between 0 and 10 Mbp differs in the two methods (Fig.9f and Fig.7d). The Hi-C data suggest that the interactions within the range of s between 0 and 10 Mbp are attractive (Fig.9f) whereas the imaging data suggest that interactions for $s \lesssim 5$ Mbp are repulsive, and become attractive for larger s values (Fig.7d).

Next, we compared the Hi-C and imaging techniques at a smaller scale. To this end, we applied HIPPS/DIMES to Hi-C data for Chr21 28-30 Mbp from the IMR90 cell line [1]. Supplementary Fig. 6a shows the comparison between the mean distance matrix inferred from Hi-C using HIPPS/DIMES method and that calculated from imaging data. The scatter plot between r_{ij}^{Imaging} and $r_{ij}^{\text{Hi-C}}$ (Supplementary Fig. 6b) shows higher degree of agreement compared to Chr2 (Fig. 9b). We then compute F_k and its distribution $P(F_k)$, which shows good agreement between Hi-C and imaging data (Supplementary Fig. 6c). These results show that structures inferred from Hi-C and imaging have higher degree of agreement on the length scale of Mbp compared to the scale of whole chromosome.

Discussion and Conclusion

We have developed a computational method (DIMES) that solves the following inverse problem: how to generate the three-dimensional conformations from the experimentally measured average distance matrix? First, applications to genome data on a length scale of a few TADs and on the scale of whole chromosome, show that DIMES correctly reproduces the pairwise distances. Second, we demonstrate that DIMES accurately accounts for the higher-order structures beyond pairwise contacts, such as the three-body interactions, radius of gyration, shapes, and the clustering of structures. These results for Chr21 on 2Mbp and the entire 242Mbp Chr2 agree quantitatively with multiplexed super-resolution data, thus setting the stage for a wide range of applications. Third, we also demonstrate that the DIMES accurately predicts the changes in the structures due to structural variants. We believe that this is a key prediction because the results for the wild-type suffices to predict 3D structures, thus eliminating the need to do additional experiments.

Implications of the DIMES method: Our method is based on the maximum entropy principle, which is used to find the optimal distribution over the coordinates of chromatin loci that are consistent with experimental data. With the choice of the average squared pairwise distances as constraints, the maximum entropy distribution has a special mathematical structure. The distribution in Eq. 1 shows that (1) it is a multivariate normal distribution whose properties are analytically known. Thus, finding the values of parameters k_{ij} does not require simulations but only an optimization procedure. (2) It has the same mathematical structure as GRM [78, 79] if one sets $1/k_BT = 1$. Hence, all the properties of the GRM [63] also hold for Eq.1. This analogy provides a physical interpretation of k_{ij}, allowing us to explain phase separation between A and B compartments without appealing to polymer models.

Interpretation of k_{ij}'s: We wondered whether k_{ij}'s could be decomposed into two additive terms representing the polymer component and epigenetic component, $k_{ij \in \alpha\beta} = k_0(i,j) + k_e(\alpha,\beta)$; $k_0(i,j)$ is the polymer contribution to k_{ij} which only depends on i and j. $k_e(\alpha,\beta)$ is the epigenetic contribution to k_{ij} that only depends on epigenetic types, e.g. $k_e(A,A)$, $k_e(B,B)$ and $k_e(A,B)$. Assuming that $k_{ij \in \alpha\beta} = k_0(i,j) + k_e(\alpha,\beta)$ holds, by averaging over $s = |i-j|$, it follows that $\langle k_{AA}(s) \rangle - \langle k_{AB}(s) \rangle$, $\langle k_{AA}(s) \rangle - \langle k_{BB}(s) \rangle$, and $\langle k_{BB}(s) \rangle - \langle k_{AB}(s) \rangle$ are independent of s. Supplementary Fig. 7 shows that these quantities fluctuate approximately around constant values for 60 Mbp $< s <$ 200 Mbp. This result suggests that, to a first approximation, the parameters k_{ij}'s may be decoupled into contributions from polymer connectivity and that arising from epigenetic states in an additive manner.

Predictive power of DIMES: One could legitimately wonder about the utility of DIMES, especially if future imaging techniques generate the coordinates of in-

dividual loci at high resolution. Of course, if this were to occur then it would make all computational approaches as well as Hi-C experiments for genome organization irrelevant. However, what is worth noting is that using the DIMES method one can also predict the 3D structures of structural variants accurately. These applications show that the DIMES method accurately accounts for the data generated by high resolution imaging experiments for the WT. With the WT k_{ij}'s at hand, certain mutational effects could be predicted without having to repeat the imaging experiments, which may not become routine for the foreseeable future. Such high throughput calculations, which can be performed using DIMES, would be particularly useful when analyzing cancer data from different tissues. We believe this is the major advantage of our computational approach.

Data Availability: All relevant data supporting the findings of this study are available within the article and its Supplementary Information. The Multiplexed FISH imaging data used in this study are publicly available from the GitHub repository at `https://github.com/BogdanBintu/ChromatinImaging` and Zenodo repository at `https://zenodo.org/record/3928890#.Yizd1xDMKFF`. The Hi-C data used in this study is publicly available from GEO database under accession number GSE92294 and GSE63525.

Code Availability: The code for the DIMES method presented in this work and its detailed user instruction can be accessed at the Github repository `https://github.com/anyuzx/HIPPS-DIMES`

Acknowledgments: We thank Davin Jeong and Sucheol Shin for several pertinent comments on the work. This work was supported by a grant from the National Science Foundation (CHE 19-00033) and the Welch Foundation theough the Collie-Welch Chair (F-0019).

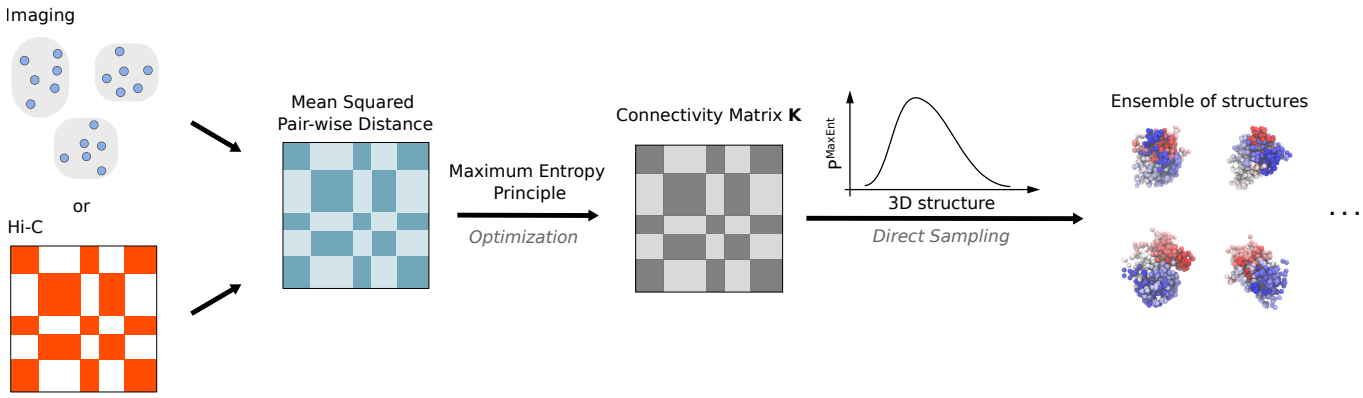

FIG. 1. Schematic flowchart for DIMES. Either imaging (measurements of chromatin loci coordinates) or Hi-C data (contact map) may be used to compute or infer the mean pairwise distance matrix, which is used as constraints to determine the maximum entropy distribution P^{MaxEnt}. The parameters, which we refer to as the connectivity matrix $\boldsymbol{K}$, in the P^{MaxEnt}, are obtained through an optimization procedure using either iterative scaling or gradient descent algorithm, as explained in the Supplementary Note 3. The ensemble of structures (coordinates of chromatin loci) can be randomly sampled from the distribution P^{MaxEnt}.

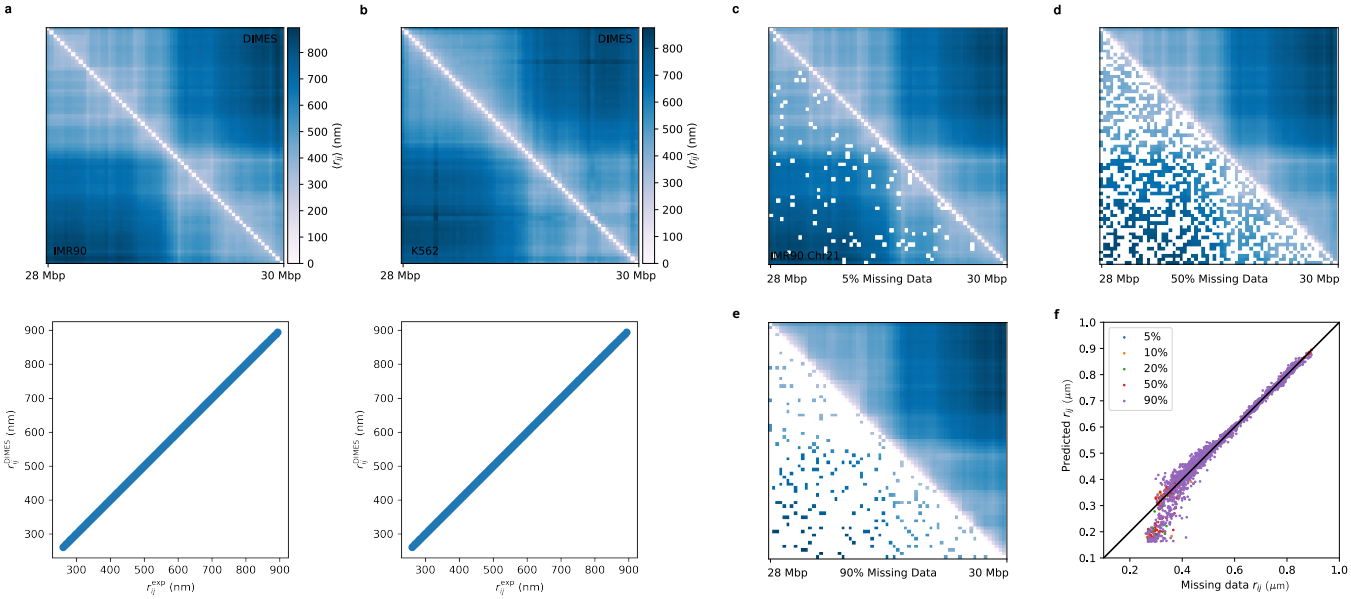

FIG. 2. Effectiveness of DIMES in matching the targets generated using experimental imaging data. Comparison between the mean spatial distances computed from the reconstructed structures and the experimental data in cell lines: IMR90 (**a**), K562 (**b**), A549 (Supplementary Fig. 10a), and HCT116 Supplementary Fig. 10b). The upper panel shows side-by-side comparisons of the distance matrices, and the lower panel displays the scatter plot between individual pairwise distances r_{ij}'s. The Pearson correlation coefficient is near unity (> 0.99) for all of the cell types shows the accuracy of the DIMES method. **(c-e)** Comparison between input distance map with missing data (lower triangle) and the full predicted distance map (upper triangle). A percentage of pairwise distances in the distance map are randomly chosen and set to be missing data (displayed in white). **(f)** r_{ij} for missing data versus the predicted values. Black line has a slope of unity.

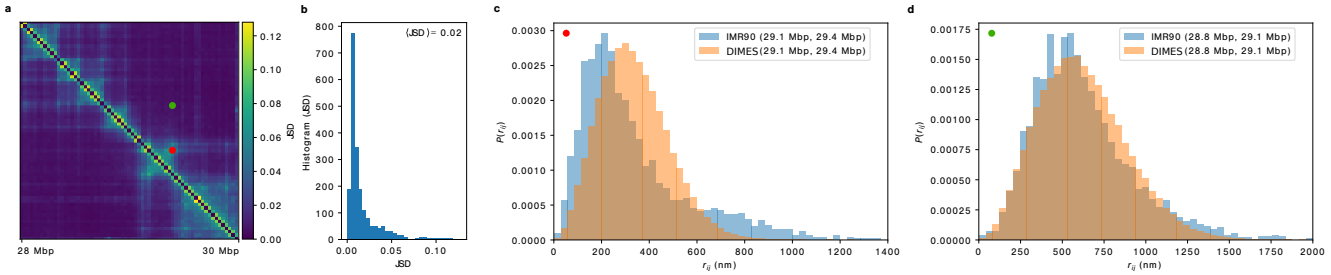

FIG. 3. Comparison between the calculated and measured $P(r_{ij})$. **(a)** The Jensen-Shannon divergence matrix computed from experiments (top half) and using DIMES (bottom half). The scale on the right shows that the maximum JSD value is $\approx$ only 0.12, thus establishing the effectiveness of DIMES in the calculating $P(r_{ij})$. **(b)** Histogram of the JSD. **(c)** Comparison between $P(r_{ij})$ for the pair (29.1 Mbp, 29.4 Mbp) for experiment and the model (corresponding to the red dot in **(a)**). **(d)** Same plot as in **(c)** but for the pair (28.8 Mbp, 29.1 Mbp), which corresponds to the green dot in **(a)**.

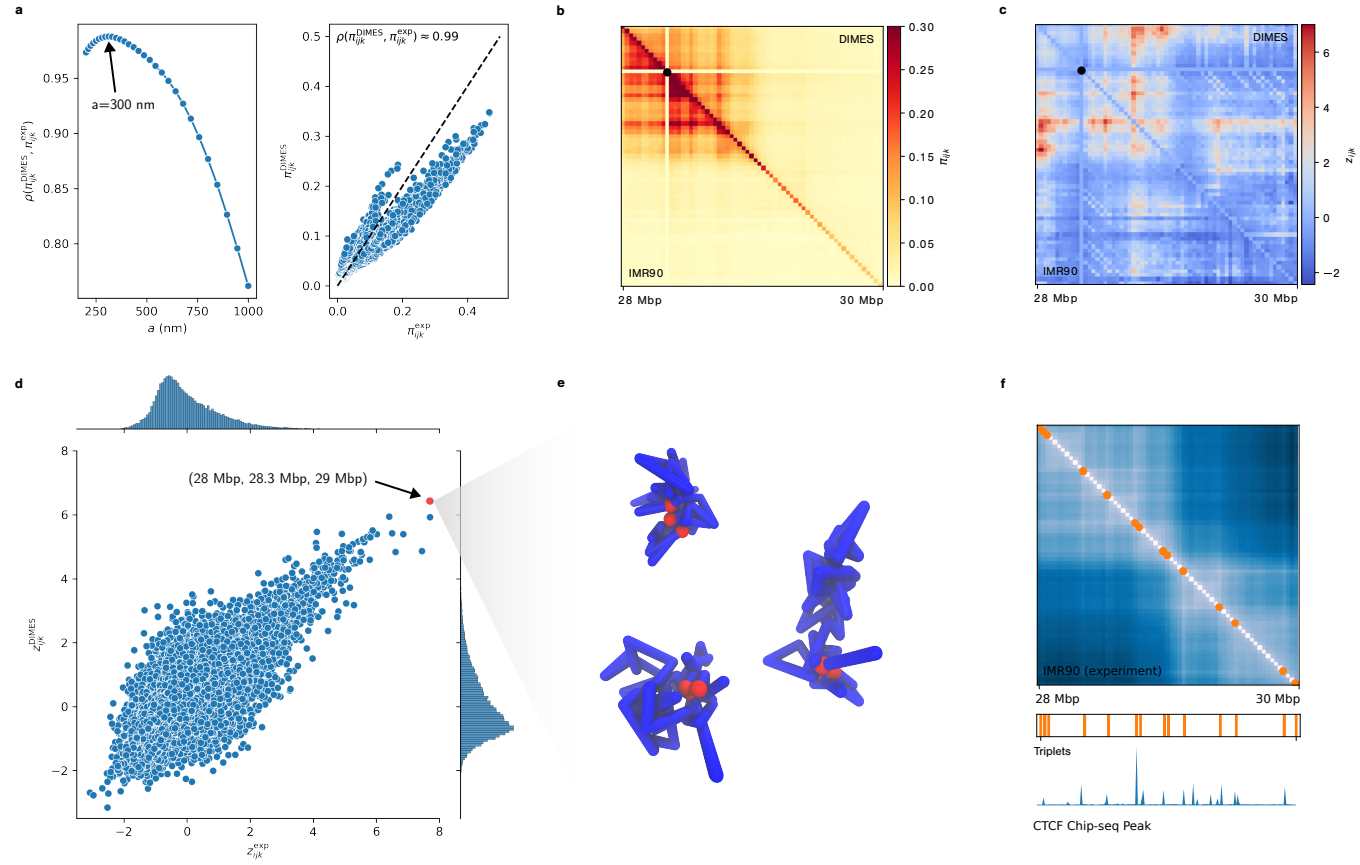

FIG. 4. Comparison of three-way contacts predicted by DIMES with super-resolution experimental data. **(a)** (Left) Pearson correlation coefficient, $\rho(\pi_{ijk}^{\mathrm{exp}}, \pi_{ijk}^{\mathrm{model}})$ versus a, where a is the threshold distance for contact formation. π_{ijk}^{exp} ($\pi_{ijk}^{\mathrm{DIMES}}$) are the probabilities of co-localization of three loci (i, j, k), defined as $\pi_{ijk} = \Pr(r_{ij} < a, r_{jk} < a, r_{ik} < a)$. (Right) Plot of $\pi_{ijk}^{\mathrm{DIMES}}$ versus π_{ijk}^{exp} for $a = 300$nm at which the $\rho(\pi_{ijk}^{\mathrm{exp}}, \pi_{ijk}^{\mathrm{DIMES}})$ is a maximum. **(b)** Comparison between the heatmaps for π_{ijk}^{exp} (lower triangle) and $\pi_{ijk}^{\mathrm{DIMES}}$ (upper triangle) for $i = 11$. The scale on the right gives $\pi_{11,j,k}$. **(c)** Same as **(b)** except it compares Z_{ijk}^{exp} (lower triangle) Z_{ijk}^{DIMES}. $Z_{ijk} = (\pi_{ijk} - \mu(\pi_{ijk}))/\sigma(\pi_{ijk})$ where $\mu(\pi_{ijk}) = \sum_{m,n,q} \delta(|j-i||k-j| - |m-n||n-q|)\pi_{mnq} / \sum_{m,n,q} \delta(|j-i||k-j| - |m-n||n-q|)$, and $\sigma(\pi_{ijk})$ is the standard deviation. The $Z_{11,j,k}$ scale is on the right. **(d)** Scatter plot of Z_{ijk}^{DIMES} versus Z_{ijk}^{exp} for $i = 11$. **(e)** Three individual chromosome conformations with the constraint that loci $(1, 11, 31)$ be colocalized ($a = 300$nm). **(f)** (Top) Mean distance map for cell-line IMR90 with 10 triplet sets (orange circles) with ten largest Z_{ijk} values. (Middle) Same as the orange circles in the Top panel, but plotted horizontally for easier comparison with the bottom panel. (Bottom) The CTCF Peak track plotted using Chip-seq data [69].

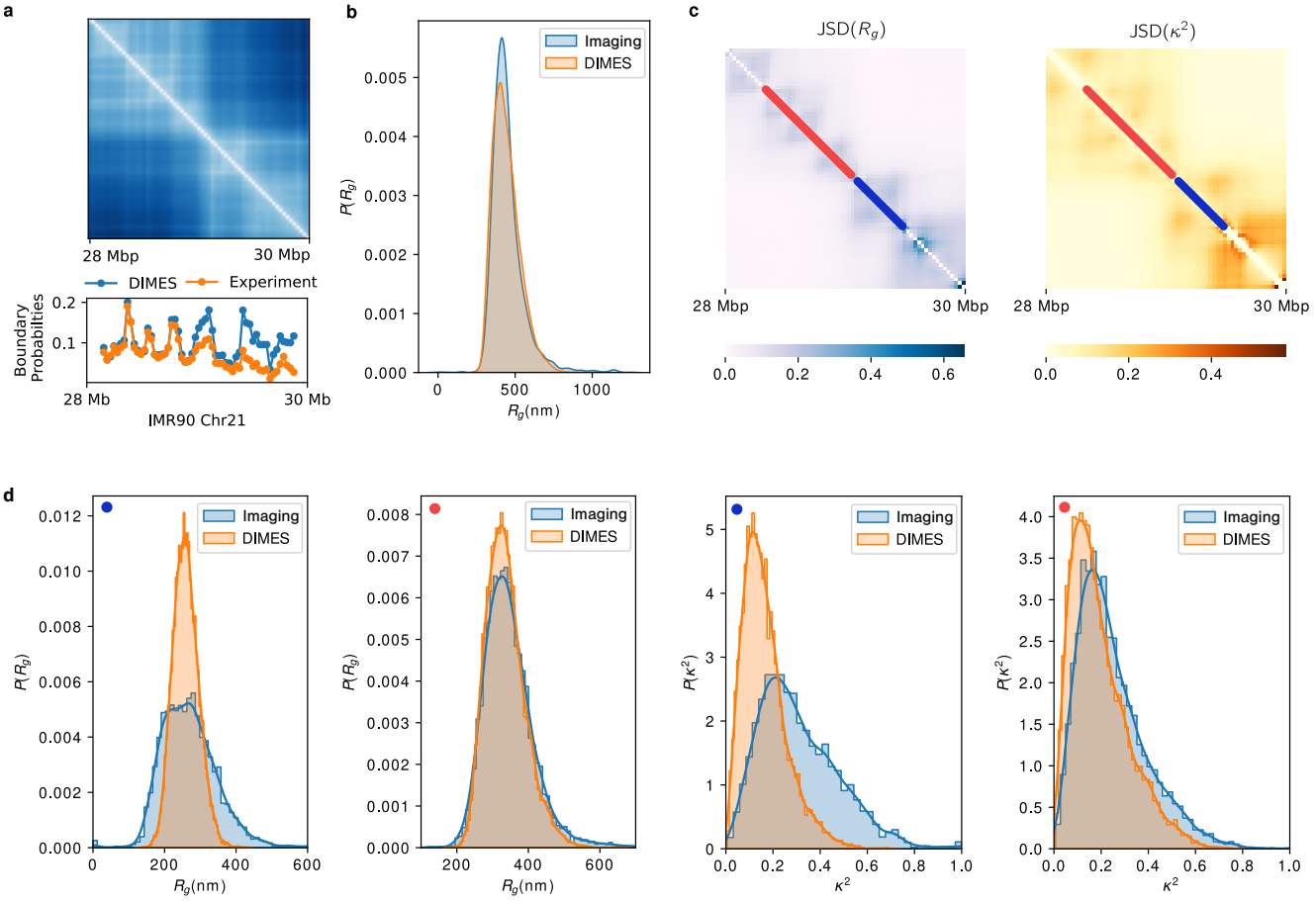

FIG. 5. TAD-like structures and shape characteristics of Chromosome 21. **(a)** Comparison of boundary probabilities between experimental imaging data and prediction from DIMES. Boundary probability measures the probability that a genomic loci acts as a single-cell domain boundary. **(b)** Comparison of $P(R_g)$ between experiment and the DIMES predictions. $P(R_g)$ is the probability density distribution of the radius of gyration R_g for the 28 Mbp - 30 Mbp region of Chr21. Comparison of $P(\kappa^2)$ where κ^2 is the shape parameter is shown in Supplementary Fig. 10. **(c)** The heatmaps of the JSD of the distribution of R_g and κ^2 between the experiment and the calculations. Each element (i, j) in the heatmap is the value of JSD for the segment that starts from i^{th} loci and ends at j^{th} loci). Red and blue lines represent two such segments. **(d)** Comparison of $P(R_g)$ and $P(\kappa^2)$ between the predictions using DIMES and those calculated using experiments for the segments marked in **(c)**. The blue (red) dot on the left corner in each sub-figure indicates the locations of the segments.

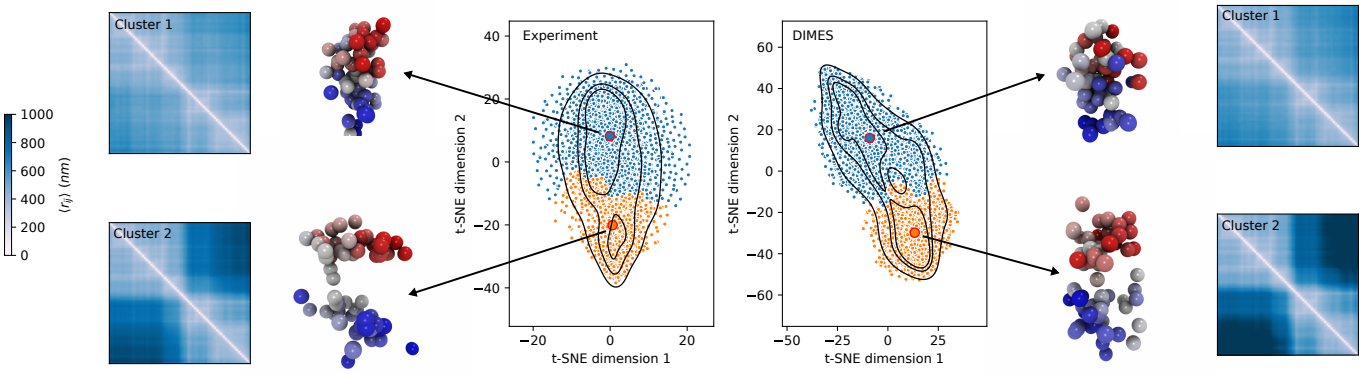

FIG. 6. Phase space structure of the 2Mbp Chr21 organization. t-SNE projections calculated from the conformations and the agglomerative clustering results. Individual conformation is projected onto a two-dimensional manifold using t-SNE. Contour lines of the density of t-SNE projections are shown to reflect the underlying clusters of the conformations. The conformations naturally partition into two into two clusters (cluster #1 and cluster # 2 marked by blue and orange, respectively). A representative conformation from each cluster and the mean distance map computed from the conformations belonging to each cluster are also displayed.

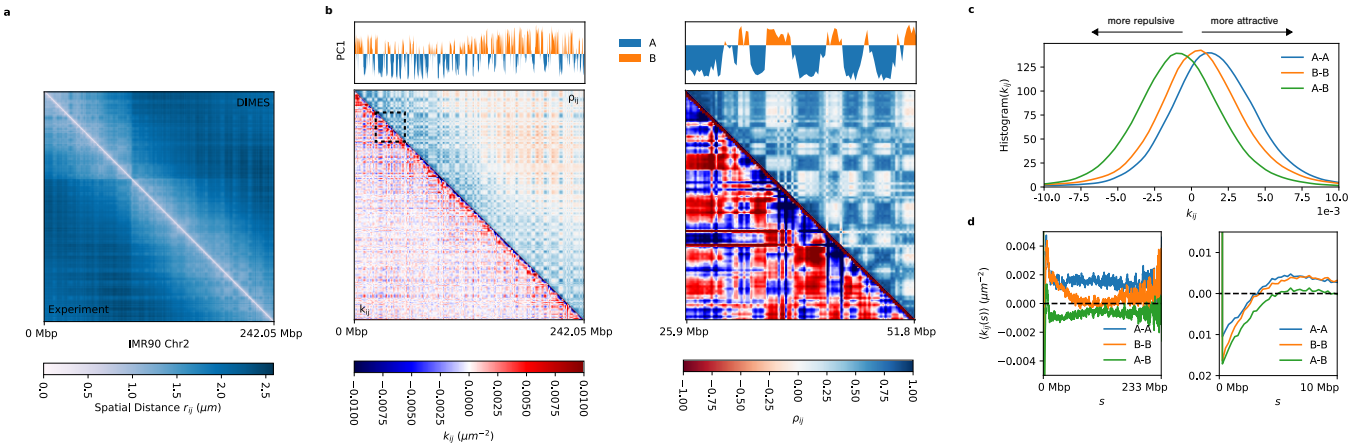

FIG. 7. Features of the Chr2 organization. **(a)** The average distance matrix between experiment (lower triangle) and the DIMES (upper triangle) show excellent agreement. **(b)** The connectivity matrix $\boldsymbol{K}$, whose elements are k_{ij} (lower triangle), and the correlation matrix $\boldsymbol{\rho}$ (upper triangle) computed from $\boldsymbol{K}$. The top track shows the principal component dimension 1 (PC1) computed using principal component analysis (PCA) from $\boldsymbol{\rho}$. A(B) compartments correspond to negative (positive) PC1. **(c)** Histogram of k_{ij} for A-A, B-B, and A-B. **(d)** Genomic-distance normalized $\langle k_{ij}(s) \rangle = (1/(N-s)) \sum_{i<j}^{N} \delta(s-(j-i)) k_{ij}$ for A-A, B-B, and A-B. $\langle k_{ij}(s) \rangle$ are shown for s between 0 and 233 Mbp (left) and for between 0 and 10 Mbp (right).

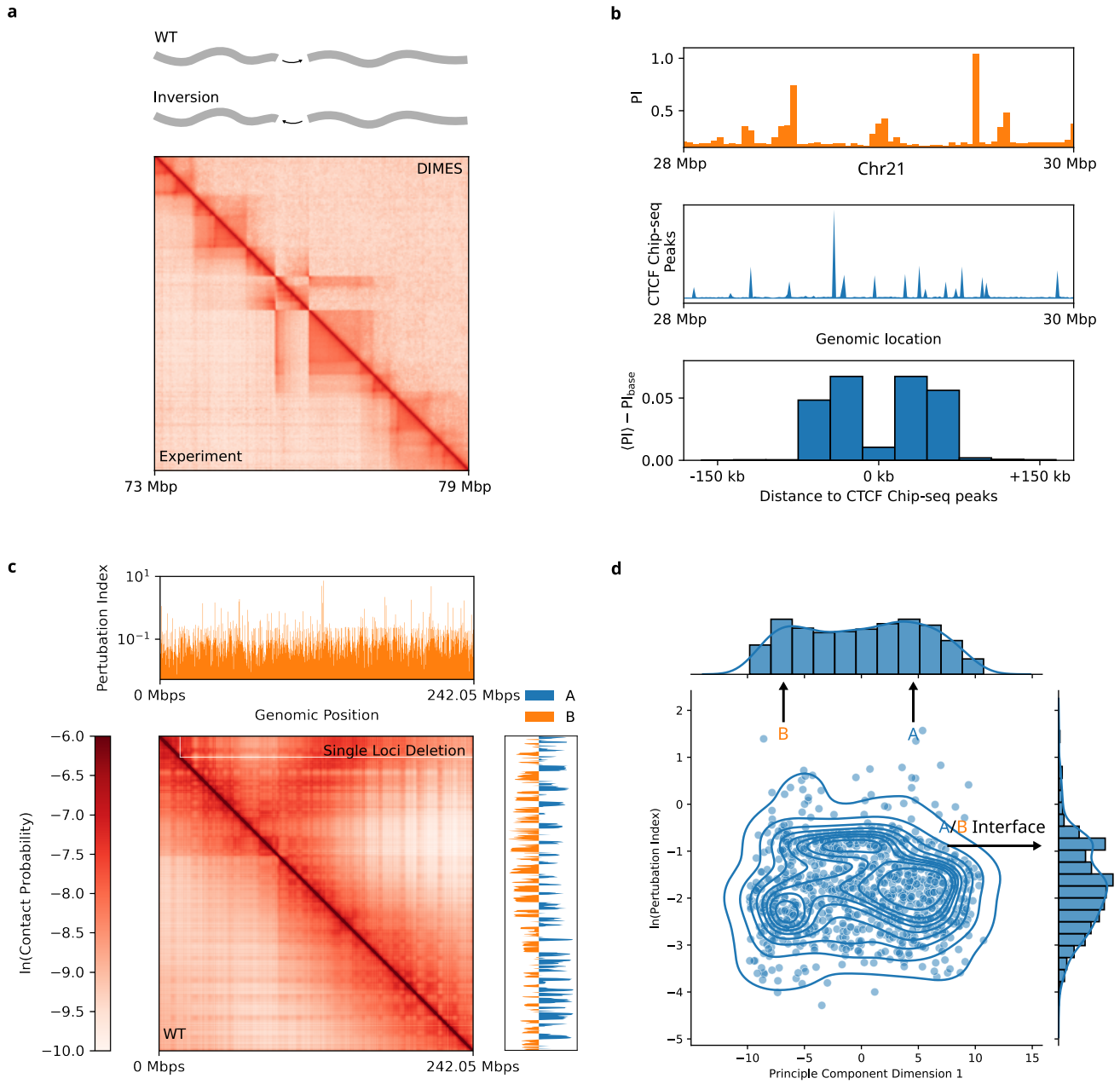

FIG. 8. Predictions for Structural Variants. **(a)** Experimentally measured Hi-C contact map for the 1.1-Mb homozygous inversion for Chr1 from the mouse cell line E11.5 (lower triangle). The position of the segment that is inverted is shown on the track in the top panel. The predicted contact map using DIMES with the perturbation on the WT Hi-C data (see main text and Supplementary Note 5) is shown for comparison (upper triangle). **(b)** Top: Perturbation Index (PI) for Chr21 28 Mbp - 30 Mbp. Middle: Chip-seq data for CTCF. Bottom: average value of PI as a function of genomic distance from CTCF Chip-seq peaks. Average PI values up to 150 kbp on either side of CTCF Chip-seq peaks are calculated at 30 kbp resolution. $\mathrm{PI}_{\mathrm{base}} = 0.2$. **(c)** An example of the contact map with a single locus deletion (upper triangle) for Chr2 from the IMR90 cell line. The top panel shows track plot of the perturbation index (PI) computed using Eq.3. **(d)** Plot of principal component dimension 1 (computed from $\boldsymbol{\rho}$) versus the logarithm of the perturbation index. A, B and A/B boundaries are marked. Histograms of PC1 and the logarithm of the perturbation index are shown on the top and side, respectively.

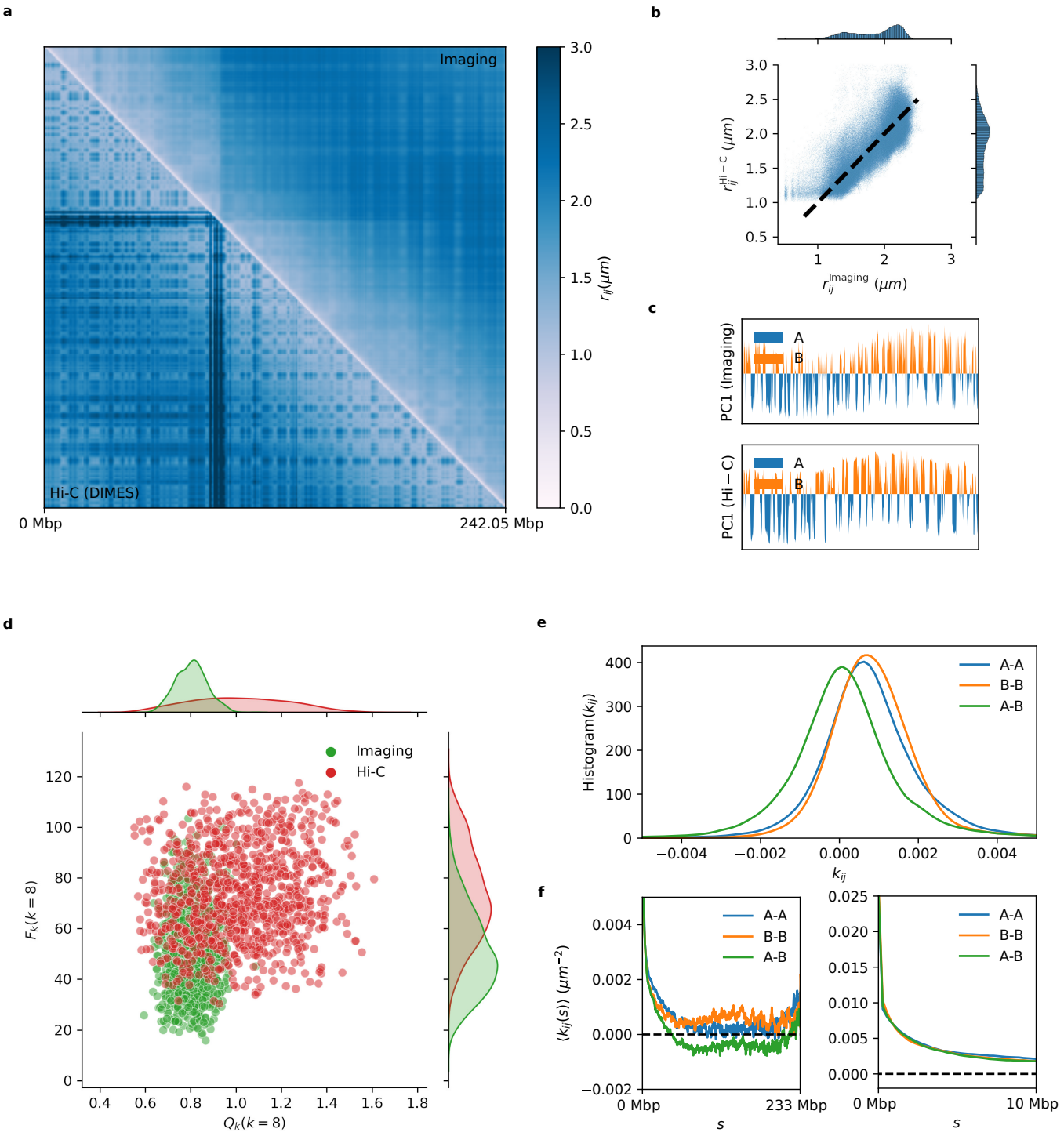

FIG. 9. Structural organization calculated from Hi-C and imaging data. **(a)** Comparison between the mean distance matrix inferred from Hi-C contact map (lower triangle) and the experimental measured average distance matrix (upper triangle) for Chr2 from the cell line IMR90. The distance scale is given on the right. **(b)** Direct comparison of pairwise distances, $\langle r_{ij}^{\text{Imaging}}\rangle$ versus $\langle r_{ij}^{\text{Hi-C}}\rangle$. Each dot represents a pair (i, j). Dashed line, with a slope of unity, is a guide to the eye. **(c)** Principal component dimension 1 (PC1) for imaging and Hi-C data. The correlation matrix is computed from the connectivity matrix $\boldsymbol{K}$, and then PCA is performed on the resulting correlation matrix. **(d)** Scatter plot of $Q_k(k = 8)$ and $F_k(k = 8)$ for 1,000 conformations. The conformations are randomly chosen from the total of $\sim$ 3,000 conformations measured in the imaging experiment (green). For Hi-C, 1,000 conformations are randomly generated using HIPPS/DIMES (red). **(e)** Histogram of k_{ij} for A-A, B-B, and A-B. k_{ij}s are obtained using Hi-C contact map. **(f)** Genomic-distance normalized $\langle k_{ij}(s)\rangle = (1/(N-s))\sum_{i<j}^{N}\delta(s-(j-i))k_{ij}$ for A-A, B-B, and A-B. $\langle k_{ij}(s)\rangle$ are shown for s between 0 and 233 Mbp (left) and for between 0 and 10 Mbp (right).

[1] Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell **159**, 1665–1680 (2014).

[2] Giorgetti, L. et al. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. Cell **157**, 950–963 (2014).

[3] Finn, E. H. & Misteli, T. Molecular basis and biological function of variability in spatial genome organization. Science **365** (2019).

[4] Nichols, M. H. & Corces, V. G. Principles of 3D compartmentalization of the human genome. Cell Reports **35**, 109330 (2021).

[5] Jerkovic´, I. & Cavalli, G. Understanding 3D genome organization by multidisciplinary methods. Nature Reviews Molecular Cell Biology **22**, 511–528 (2021).

[6] Barbieri, M. et al. Complexity of chromatin folding is captured by the strings and binders switch model. Proceedings of the National Academy of Sciences **109**, 16173–16178 (2012).

[7] Zhang, B. & Wolynes, P. G. Topology, structures, and energy landscapes of human chromosomes. Proceedings of the National Academy of Sciences **112**, 6062–6067 (2015).

[8] Shi, G., Liu, L., Hyeon, C. & Thirumalai, D. Interphase human chromosome exhibits out of equilibrium glassy dynamics. Nature Communications **9** (2018).

[9] Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nature Reviews Genetics **14**, 390–403 (2013).

[10] Rao, S. S. et al. Cohesin loss eliminates all loop domains. Cell **171**, 305–320.e24 (2017).

[11] Chen, H. et al. Dynamic interplay between enhancer–promoter topology and gene activity. Nature Genetics **50**, 1296–1303 (2018).

[12] Ghavi-Helm, Y. et al. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. Nature Genetics **51**, 1272–1282 (2019).

[13] Delaneau, O. et al. Chromatin three-dimensional interactions mediate genetic effects on gene expression. Science **364** (2019).

[14] Zuin, J. et al. Nonlinear control of transcription through enhancer-promoter interactions. Biorxiv (2021).

[15] Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. Science **295**, 1306–1311 (2002).

[16] Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science **326**, 289–293 (2009).

[17] Hsieh, T.-H. S. et al. Mapping nucleosome resolution chromosome folding in yeast by Micro-C. Cell **162**, 108–119 (2015).

[18] Bonev, B. & Cavalli, G. Organization and function of the 3D genome. Nature Reviews Genetics **17**, 661–678 (2016).

[19] Yu, M. & Ren, B. The three-dimensional organization of mammalian genomes. Annual Review of Cell and Developmental Biology **33**, 265–289 (2017).

[20] Nagano, T. et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature **502**, 59–64 (2013).

[21] Flyamer, I. M. et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. Nature **544**, 110–114 (2017).

[22] Stevens, T. J. et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature **544**, 59–64 (2017).

[23] Ramani, V. et al. Massively multiplex single-cell Hi-C. Nature Methods **14**, 263–266 (2017).

[24] Tan, L., Xing, D., Chang, C.-H., Li, H. & Xie, X. S. Three-dimensional genome structures of single diploid human cells. Science **361**, 924–928 (2018).

[25] Finn, E. H. et al. Extensive heterogeneity and intrinsic variation in spatial genome organization. Cell **176**, 1502–1515.e10 (2019).

[26] Quinodoz, S. A. et al. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. Cell **174**, 744–757.e24 (2018).

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

[28] Kempfer, R. & Pombo, A. Methods for mapping 3D chromosome architecture. Nature Reviews Genetics **21**, 207–226 (2019).

[29] Duan, Z. et al. A three-dimensional model of the yeast genome. Nature **465**, 363–367 (2010). URL `https://doi.org/10.1038/nature08973`.

[30] Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F. & Chen, L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nature Biotechnology **30**, 90–98 (2011). URL `https://doi.org/10.1038/nbt.2057`.

[31] Rousseau, M., Fraser, J., Ferraiuolo, M. A., Dostie, J. & Blanchette, M. Three-dimensional modeling of chromatin structure from interaction frequency data using markov chain monte carlo sampling. BMC Bioinformatics **12** (2011). URL `https://doi.org/10.1186/1471-2105-12-414`.

[32] Zhang, Z., Li, G., Toh, K.-C. & Sung, W.-K. 3d chromosome modeling with semi-definite programming and hi-c data. Journal of Computational Biology **20**, 831–846 (2013). URL `https://doi.org/10.1089/cmb.2013.0076`.

[33] Hu, M. et al. Bayesian inference of spatial organizations of chromosomes. PLoS Computational Biology **9**, e1002893 (2013). URL `https://doi.org/10.1371/journal.pcbi.1002893`.

[34] Varoquaux, N., Ay, F., Noble, W. S. & Vert, J.-P. A statistical approach for inferring the 3d structure of the genome. Bioinformatics **30**, i26–i33 (2014). URL `https://doi.org/10.1093/bioinformatics/btu268`.

[35] Lesne, A., Riposo, J., Roger, P., Cournac, A. & Mozziconacci, J. 3D genome reconstruction from chromosomal contacts. Nature Methods **11**, 1141–1143 (2014).

[36] Tjong, H. et al. Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. Proceedings of the National Academy of Sciences **113**, E1663–E1672 (2016).

[37] Oluwadare, O., Zhang, Y. & Cheng, J. A maximum likelihood algorithm for reconstructing 3d structures of

human chromosomes from chromosomal contact data. BMC Genomics **19** (2018). URL https://doi.org/10.1186/s12864-018-4546-8.

[38] Wang, H., Yang, J., Zhang, Y., Qian, J. & Wang, J. Reconstruct high-resolution 3d genome structures for diverse cell-types using FLAMINGO. Nature Communications **13** (2022). URL https://doi.org/10.1038/s41467-022-30270-2.

[39] Hua, N. et al. Producing genome structure populations with the dynamic and automated PGS software. Nature Protocols **13**, 915–926 (2018). URL https://doi.org/10.1038/nprot.2018.008.

[40] Oluwadare, O., Highsmith, M. & Cheng, J. An overview of methods for reconstructing 3-d chromosome and genome structures from hi-c data. Biological Procedures Online **21** (2019). URL https://doi.org/10.1186/s12575-019-0094-0.

[41] Ou, H. D. et al. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science **357**, eaag0025 (2017).

[42] Boettiger, A. & Murphy, S. Advances in chromatin imaging at kilobase-scale resolution. Trends in Genetics **36**, 273–287 (2020). URL https://doi.org/10.1016/j.tig.2019.12.010.

[43] Li, Y. et al. Nanoscale chromatin imaging and analysis platform bridges 4d chromatin organization with molecular function. Science Advances **7** (2021). URL https://doi.org/10.1126/sciadv.abe4310.

[44] Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nature Reviews Genetics **2**, 292–301 (2001).

[45] Wang, S. et al. Spatial organization of chromatin domains and compartments in single chromosomes. Science **353**, 598–602 (2016).

[46] Cattoni, D. I. et al. Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. Nature Communications **8** (2017).

[47] Bintu, B. et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science **362**, eaau1783 (2018).

[48] Nir, G. et al. Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. PLOS Genetics **14**, e1007872 (2018).

[49] Szabo, Q. et al. TADs are 3D structural units of higher-order chromosome organization in drosophila. Science Advances **4** (2018).

[50] Mateo, L. J. et al. Visualizing DNA folding and RNA in embryos at single-cell resolution. Nature **568**, 49–54 (2019).

[51] Su, J.-H., Zheng, P., Kinrot, S. S., Bintu, B. & Zhuang, X. Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. Cell **182**, 1641–1659.e26 (2020).

[52] Liu, M. et al. Multiplexed imaging of nucleome architectures in single cells of mammalian tissue. Nature Communications **11** (2020).

[53] Shi, G. & Thirumalai, D. From Hi-C contact map to three-dimensional organization of interphase human chromosomes. Physical Review X **11** (2021).

[54] Jaynes, E. T. Information theory and statistical mechanics. Physical Review **106**, 620–630 (1957).

[55] Pressé, S., Ghosh, K., Lee, J. & Dill, K. A. Principles of maximum entropy and maximum caliber in statistical physics. Reviews of Modern Physics **85**, 1115–1141 (2013).

[56] Pierro, M. D., Zhang, B., Aiden, E. L., Wolynes, P. G. & Onuchic, J. N. Transferable model for chromosome architecture. Proceedings of the National Academy of Sciences **113**, 12168–12173 (2016).

[57] Farré, P. & Emberly, E. A maximum-entropy model for predicting chromatin contacts. PLOS Computational Biology **14**, e1005956 (2018).

[58] Lin, X., Qi, Y., Latham, A. P. & Zhang, B. Multiscale modeling of genome organization with maximum entropy optimization. The Journal of Chemical Physics **155**, 010901 (2021).

[59] Messelink, J. J. B., van Teeseling, M. C. F., Janssen, J., Thanbichler, M. & Broedersz, C. P. Learning the distribution of single-cell chromosome conformations in bacteria reveals emergent order across genomic scales. Nature Communications **12** (2021).

[60] Malouf, R. A comparison of algorithms for maximum entropy parameter estimation. In COLING-02: The 6th Conference on Natural Language Learning 200 (2002).

[61] Treut, G. L., Képès, F. & Orland, H. A polymer model for the quantitative reconstruction of chromosome architecture from HiC and GAM data. Biophysical Journal **115**, 2286–2294 (2018).

[62] Shinkai, S. et al. PHi-C: deciphering Hi-C data into polymer dynamics. NAR Genomics and Bioinformatics **2** (2020).

[63] Shi, G. & Thirumalai, D. Conformational heterogeneity in human interphase chromosome organization reconciles the FISH and Hi-C paradox. Nature Communications **10** (2019).

[64] Liu, L., Zhang, B. & Hyeon, C. Extracting multi-way chromatin contacts from Hi-C data. PLOS Computational Biology **17**, e1009669 (2021).

[65] Harju, J., Messelink, J. J. & Broedersz, C. P. Multi-contact statistics distinguish models of chromosome organization (2022). URL https://doi.org/10.1101/2022.05.17.492279.

[66] Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature **485**, 376–380 (2012).

[67] Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin domains: The unit of chromosome organization. Molecular Cell **62**, 668–680 (2016).

[68] Ong, C.-T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. Nature Reviews Genetics **15**, 234–246 (2014).

[69] Zhang, J. et al. An integrative ENCODE resource for cancer genomics. Nature Communications **11** (2020).

[70] Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. eLife **6** (2017).

[71] Zirkel, A. et al. HMGB2 loss upon senescence entry disrupts genomic organization and induces CTCF clustering across cell types. Molecular Cell **70**, 730–744.e6 (2018).

[72] Cheng, Y., Liu, M., Hu, M. & Wang, S. TAD-like single-cell domain structures exist on both active and inactive x chromosomes and persist under epigenetic perturbations. Genome Biology **22** (2021).

[73] Aronovitz, J. & Nelson, D. Universal features of polymer shapes. Journal de Physique **47**, 1445–1456 (1986).

[74] Dima, R. I. & Thirumalai, D. Asymmetry in the shapes of folded and denatured states of proteins. The Journal of Physical Chemistry B **108**, 6564–6570 (2004).

[75] Virtanen, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods **17**, 261–272 (2020).

[76] Belaghzal, H. et al. Liquid chromatin Hi-C characterizes compartment-dependent chromatin interaction dynamics. Nature Genetics **53**, 367–378 (2021).

[77] Bianco, S. et al. Polymer physics predicts the effects of structural variants on chromatin architecture. Nature Genetics **50**, 662–667 (2018).

[78] Bryngelson, J. D. & Thirumalai, D. Internal constraints induce localization in an isolated polymer molecule. Physical Review Letters **76**, 542–545 (1996).

[79] Doi, M., Edwards, S. F. & Edwards, S. F. The theory of polymer dynamics, vol. 73 (oxford university press, 1988).