Characterizing the variation in chromosome structure ensembles in the context of the nuclear microenvironment

Priyojit Das[1], Tongye Shen[1,2] and Rachel Patton McCord[1,2,*]

[1] UT-ORNL Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, Tennessee, 37996, USA

[2] Biochemistry & Cellular and Molecular Biology, University of Tennessee, Knoxville, Tennessee, 37996, USA

* To whom correspondence should be addressed. Tel: 865-974-3149; Email: rmccord@utk.edu

ABSTRACT

Inside the nucleus, chromosomes are subjected to direct physical interaction between different components, active forces, and thermal noise, leading to the formation of an ensemble of three-dimensional structures. However, it is still not well understood to what extent and how the structural ensemble varies from one chromosome region or cell-type to another. We designed a statistical analysis technique and applied it to single-cell chromosome imaging data to reveal the fluctuation of individual chromosome structures. By analyzing the resulting structural landscape, we find that the largest dynamic variation is the overall radius of gyration of the chromatin region, followed by domain reorganization within the region. By comparing different human cell-lines and experimental perturbations data using this statistical analysis technique and a network entropy approach, we identify both cell-type and condition-specific features of the structural landscapes. We identify a relationship between epigenetic state and the properties of chromosome structure fluctuation and validate this relationship through polymer simulations. Overall, our study suggests that the types of variation in a chromosome structure ensemble are cell-type as well as region-specific and can be attributed to constraints placed on the structure by factors such as variation in epigenetic state.

INTRODUCTION

In eukaryotic cells, long genomic DNA is packed within a relatively small nucleus, requiring chromosomes to fold into a multi-layered 3D spatial organization (1). At the kilobase length scale, loops form between different chromatin regions (2-4). Some loops are formed by the interaction of cohesin-mediated loop extrusion and boundary factors, a process which can create domains of enriched contacts called Topologically Associated Domains (TADs) at the scale of 100s of kilobases (5-7). Through processes such as phase separation and tethering to nuclear structures the chromatin is further organized at the megabase scale into separate euchromatin and heterochromatin spatial regions, known as A/B compartments (8-12). Finally at the individual chromosome level, the chromosomes tend to form their own territories (13-15). Each of these structural features act independently as well as synergistically to contribute to different biological processes, as reported by an ever-increasing number of studies. For

example, chromatin loops and TAD boundary positioning can influence enhancer-promoter interactions and gene regulation (16-18). This multi-layered organization of 3D genome is not static: the chromatin fiber is semiflexible in nature and is also subjected to forces generated by ATP driven molecular machines along with thermal fluctuations (19-21). This leads to both spatial and temporal structural heterogeneity. We often fail to consider the effect of such heterogeneity while interpreting and characterizing structural organization using “bulk” (population averaged) chromosome contact data. However, considering only average structures may can obscure the view of detailed folding mechanisms and can mask subpopulations of cells with different structures that mediate different biological outcomes (22). Therefore, it is important to consider the single-cell structural variation to gain a better understanding of the underlying structural organization.

In order to study genome organization inside the nucleus, several key techniques involving molecular biology and microscopy imaging-based approaches have been developed in the past decade (23,24). Among them, chromosome conformation capture (3C) based techniques such as Hi-C have become a very common approach for mapping the spatial organization the whole genome inside the nucleus However, when Hi-C and its derivatives are applied to millions of cells in bulk, it is challenging to study the heterogeneity of chromosome structures using these techniques. Though several polymer- and restraint- based bottom-up models have been developed to reconstruct an ensemble of structures from the bulk data, it is not clear whether such models are able to capture the variation present in true biological systems. To resolve this issue, researchers have started to develop single-cell counterparts of Hi-C and related techniques (25-27). However, these single cell contact capture techniques often suffer from low coverage, which makes it difficult to characterize the structural heterogeneity within a particular cell-type.

In contrast, oligonucleotide-based FISH imaging techniques have enabled the tracing of chromatin regions at high resolution in thousands of single cells (22,28-32). These techniques can generate distance distribution maps between imaged points along a chromosome that resemble Hi-C contact maps. By analyzing these single-cell structures at different resolutions, recent studies have started to characterize the heterogeneity associated with the structural ensemble. For example, it has been revealed that the inactive chrX exhibits not only canonically expected compact structures but also more chromosome domain inter-mingling as compared to the active chrX (28). Another study revealed that domain-like structure also exists on the inactive chrX at the single-cell level, but such structures are averaged out and not visible in bulk experiments (22). Bintu and co-authors reported the existence of TAD-like structures and sharp domain boundaries at a single cell level (29). Though the boundary positions vary from one cell to another, they show higher preference for structural protein binding sites. By analyzing the single-cell whole chromosome tracing data at a higher resolution, researchers identified a substantial number of TAD-like domains in single cells that cross compartment boundaries observed in bulk data (30).

While initial studies focused on showing that variability in chromosome structures exists, recent work has begun to turn toward the question of how and why this variability arises (33-35). One such study was performed by Cheng *et al*., where the authors found that a short chromatin segment spanning a TAD boundary exhibits a two state transition between closed and open dumbbell like conformations (33). They hypothesized that transcriptional regulation might contribute to the transition between two conformations. Another study proposed a folding coordinate that represents progression of TAD formation (34). They also reported that in the absence of structural protein cohesin, a local chromatin region may still exhibit TAD-like structures in individual snapshots, however these structures are less stable due to the entropy gain of the unfolded conformations. This effect has also been found by Nicodemi and group, where the removal of chromatin structural proteins reverses phase separated globules into random coil-like state (35).

Although these recent studies have started to shed light on the fundamental mechanisms of local chromatin folding at the single-cell level, a detailed characterization of the structural heterogeneity present in the population ensemble is still lacking. Though it is evident that chromosome structures vary from one cell to another, it is not clear how different the landscape of chromosome structure variation is between cell types or chromosomes. A recent FISH imaging study in *Caenorhabditis elegans* showed that chrV exhibits four major conformational clusters during early embryonic development, and the structure ensembles corresponding to different clusters exhibit a broad and continuous distribution rather than a handful of discrete conformational states. (31). Similarly, within a particular cell-type, the single-cell structure ensemble can be further described as a distribution of conformations. We have yet to characterize the distribution of such ensembles beyond the bulk average structure (mean). And we do not know how this fluctuation around an average structure (variance) varies in a cell-type specific manner. This question becomes particularly interesting when the bulk structural organization looks similar across different cell types or conditions, because it is possible that different underlying combinations could lead to nearly indistinguishable average structures while the variance of the ensemble are distinct. A logical follow up question is what factors might contribute to the extent of structural variation. As we know, there are various essential factors such as epigenetic marks, associated structural proteins, and nuclear lamina associations that vary in a cell-type specific manner and contribute to the organization of 3D genome (36-38). Therefore, investigating the role of these different factors in the variation of chromosome structural ensemble is of high importance.

To answer these questions about chromosome structural ensembles, we designed a statistical pipeline to analyze single-cell chromatin tracing distance data. Using this pipeline, we first characterized the landscape of single-cell structures across different cell-types. We then identified cell-type specific differences in the structural ensembles through joint analysis of structural landscapes. To further characterize these differences, we designed two different network entropy based approaches – one to quantify the degree of continuity across the chromosome structural ensemble (ensemble or landscape

entropy) and another to quantify the degree of structural variation of local genomic regions across the population (local bin entropy). Using these two approaches, we quantified the ensemble variation and local structural variation across different cell-types and different chromosome regions at different genomic resolutions. We identify overall locus compaction and domain boundary shifting as two primary but independent modes of structural fluctuation that are observed across all length scales. We observe differences in structural landscape entropy between regions that have different epigenetic state profiles, and we are able to demonstrate that interactions governed by epigenetic state can help explain these differences in the type of structure variation exhibited by the chromosome region.

MATERIAL AND METHODS

The statistical chromosome conformation analysis pipeline used in this work can be broadly divided into two parts. In the first part, the single-cell chromosome imaging data (of different genomic resolutions) is analyzed statistically to calculate pairwise similarity of the structures. Next, a network entropy-based approach is applied on that similarity data to quantify the degree of continuity present in the chromosome structural ensemble. In addition, the pairwise similarity data is used to construct a structural landscape to arrange the single-cell chromosome structures according to their individual and bulk structural properties (Fig. 1).

Construction of distance matrices from single-cell chromosome imaging data

For this work, we have used experimental single-cell chromosome imaging data at three different resolutions, each of them reflecting different levels of structural organization of the 3D genome at different genomic length scales. Among those, the single-cell chromosome imaging data obtained from Bintu *et al*. (29). has the highest genomic resolution of 30 kb. Here, the authors traced the conformations of multiple ~ 2 – 2.5 Mb spanning chromatin regions of chr21, partitioned into consecutive 30 kb segments in single-cells across different cell-types (e.g., IMR90 – human fetal lung fibroblast cells, K562 – human erythroleukemic cells, A549 – human lung adenocarcinoma cells, and HCT116 – human colon cancer cells) during different physiological and perturbed conditions. The second set of IMR90 chromatin tracing data analyzed here is obtained from Su *et al*. (30), where the single-cell conformations of chr21 non-repetitive region were partitioned into consecutive 50 kb resolution. In addition, that dataset also provides chr2 conformations captured at the same 50 kb resolution, however with an interval of 250 kb between two genomic regions, due to very large size of human chr2. Both the Bintu *et al*. and Su *et al*. data capture chromatin conformations at a comparable resolution (30 and 50 kb respectively), differing only in terms of the genomic length scale. Whereas the Bintu *et al*. data primarily represents the structural heterogeneity at the level of specific TADs and sub-TADs, the Su *et al*. data captures the heterogeneity of almost the whole chromosome (specifically the non-repetitive regions). Finally, we also analyzed chromosome structural heterogeneity of the whole active and inactive chrXs of IMR90, obtained from Wang *et al*. (28). In this dataset, TADs are imaged as the basic subunits along the chromosome, so the

resolution of this data is lower (ranging from 440 kb to 7.88 Mb) and reports on the conformations of compartments and the whole chromosome rather than the within-TAD level.

For all these different datasets, the authors reported the single-cell conformations of the chromosomes in a 3D Cartesian coordinate system where each basic genomic region (e.g., 30 or 50 kb genomic region or TAD; from here termed as bin) has X-, Y-, Z- coordinates values. However, in some single cell structures the 3D positions of some of the bins are missing due to technical difficulties. Therefore, to decrease the effect of missing data, we consider only structures with more than ~ 90% measured bins, and the values for missing bins in those structures are assigned by nearest-neighbor interpolation. In addition to that, the 3D structure of the chromosomes is also associated with rotational and translational variations which might affect underlying statistical analysis. To resolve this issue, we convert each 3D structure into a 2D symmetric distance matrix by calculating the pairwise Euclidean distance between all the genomic bins of that specific chromosome structure. For example, if a particular chromosome has m genomic bins and a total of N copies originating from N single cells, then we will have N symmetric pairwise distance matrices each of size $m \times m$, where d_{ij} represents the Euclidean distance between i^{th} and j^{th} genomic bins. Further, the spatial distance can be converted into contacts by applying a genomic resolution specific thresholding cutoff. However, for most of our analysis we use distance matrices instead of contact matrices, unless otherwise explicitly mentioned.

Generate chromosome structural landscapes using statistical analysis technique

We next make all the single-cell matrices centered around zero by subtracting the median distance matrix from each of them, $h_{ij} = d_{ij} - median(d_{ij})$. Once we obtain the median centered single-cell distance matrices, we calculate the pairwise dot product between each of those single-cell matrices, such as, $s_{\alpha\beta} = \sum_{i,j} h_{ij}^{\alpha} . h_{ij}^{\beta}$, where h^{α} and h^{β} are mean centered distance matrices obtained from single cells ($\alpha, \beta \in \{1, \ldots, \mathrm{N}\}$), respectively. The reason behind performing this step is to calculate the similarity between the structures, where two similar conformations would yield a strong positive value and vice versa. The pairwise dot product similarity information between all possible pairs of the single cells is represented in a symmetric matrix format. Next, we perform PCA on the pairwise similarity matrix and project the single-cell conformations in 2D space using the first two principal components (PCs). The largest dynamic variation along PC1 represents the overall radius of gyration (R_g) of the chromosome region, followed by shifts in structural boundaries within the region (PC2). This PC projection is termed the chromosome structural landscape. Detailed interpretation of this landscape has been provided in the RESULTS section.

Quantification of chromosome structural ensemble variation using network-based entropy

To quantify the variation associated with the chromosome structural ensemble, we perform a network-based entropy calculation on the pairwise similarity matrix (Fig. 1c). Similar approaches have been used

to study cellular network heterogeneity associated with cancer and cellular differentiation (39-41). In this approach, we first calculate Pearson's correlation coefficient (PCC) between each pair of rows of the similarity matrix, where c_{ij} denotes PCC between i^{th} and j^{th} rows of the pairwise similarity matrix. As PCC values ranges between -1 and 1, next we scale the PCC values in the range of 0 and 1, $w_{ij} = 0.5 * (1 + c_{ij})$. Once we have the rescaled pairwise PCC value matrix, we construct a weighted network from that where each node represents a single chromosome structure or conformation and the edges between them denotes rescaled PCC values. However, we only consider the edges having weights above a certain threshold, representing only connections between highly similar structures. This threshold is calculated by taking 50th percentile of all the edge weights of the network. After having the network constructed, we move on to calculate local Shannon-Jayne entropy of each node,

$$S_i = -\frac{1}{\log k_i} \sum_{j \in N(i)} p_{ij} \log p_{ij}$$

where $N(i)$ represents the neighborhood set of the node i and k_i denotes the cardinality of that set. p_{ij} represents the proportionality of the edge weight between nodes i and j compared to all the edges associated with node i.

$$p_{ij} = \frac{w_{ij}}{\sum_{k \in N(i)} w_{ik}}$$

This entropy S_i is bound between 0 and 1. Based on the distribution of the local Shannon-Jayne entropy of the nodes, we characterize the heterogeneity of the chromosome structural landscape. An entropy distribution with higher mean is considered as having a smooth continuous structural landscape while a distribution with relatively lower mean would have more discrete, disconnected structures. Interpretation of the entropy distribution trends have been discussed in detail in the RESULTS section.

A network-entropy based approach to quantify chromosome structure variability for each genomic bin

To characterize the variability of structure at each genomic bin along the chromosome, we use a modified version of the network-based entropy calculation approach, discussed in the previous subsection (Fig. 1d). In this step, for each single-cell structure, we first construct a network based on its distance matrix, where each node denotes the genomic bins of the structure and the edges between them represent the spatial distances. Similarly, to the previously described chromosome structural landscape entropy calculation, we then threshold the network using an experimentally satisfied distance cutoff (for 30 kb resolution 330 nm (33) and for 50 kb resolution 500 nm (30)), where edges with spatial distance higher than the cutoff are discarded from the analysis. Next, we calculate local Shannon-Jayne entropy of each of the nodes or subunits of the network using the same entropy calculation formula discussed in the previous section except here nodes are basic subunits and edge weights are spatial distance between

the nodes. Once we have the local bin entropy information, we measure the variability of each of the basic subunits across all the single cell structures by calculating the standard deviation of all the local bin entropy values specific to that region. A higher standard deviation corresponds to a highly dynamic genomic bin.

Characterization of chromatin states using epigenetic profiles

In order to characterize chromatin states of the IMR90, K562 and A549 cells, we downloaded total eleven histone marks – H3K4Me1, H3K4Me2, H3K4Me3, H3K9Ac, H3K27Ac, H4K20Me1, H3K36Me3, H3K27Me3, H3K9Me3, H2A.Z and DNase for all the three cell types from ENCODE repository (STable 1) (42). Next, we used chromHMM software (43) to annotate genomic regions of desired resolutions into five chromatin states using a combinatorial pattern of the eleven histone marks. Based on the combinatorial histone marks, we defined chromatin states as – Active state 1, Active state 2, Inactive Poised, Repressed Polycomb and Heterochromatin (Fig. S1). Briefly, if an inferred state shows strong enrichments for active marks such as H3K4Me1/2/3, H3K27Ac, H3K9Ac and DNase, we term that state as an Active states. Further, depending on the level of enrichment of those different active marks, we characterize them into Active state 1 and Active state 2. We assign Heterochromatin label to the state for which we observe strong enrichment of H3K9Me3 and for the Repressive Polycomb assignment, we look for strong H3K27Me3 and a mix of H3K9Me3 and H4K20Me1 marks. And the rest of the states are assigned Inactive Poised level.

Polymer simulation of chromatin segments based on underlying chromatin states

To test the contribution of epigenetic states to chromosome structure heterogeneity and dynamics, we perform polymer simulations of a specific chromatin segment with the interactions defined based on their epigenetic states. In recent years, polymer simulation-based studies have shown that how the compartmental profile or the underlying epigenetic state of the genomic regions give rise population-averaged structural organization of chromatin at different resolutions (9,44,45). Here, we represent a ~ 1.95 Mb spanning chromatin region as a chain of beads using a bead-spring polymer model. The model consists of total 1950 beads where beads represent genomic regions of 1 kb. Each bead is assigned a specific chromatin state as described above. In addition to that, we also model loops within the specific chromatin regions based on publicly available high-resolution Hi-C derived chromatin loop sets from Rao *et al*. (46). For each loop, we connect the loop anchors using a harmonic constraint. The potential energy function of the polymer model is described below,

$$\mathrm{U_{ref}}(r) = U_B(r) + U_P(r) + U_L(r) + U_W(r)$$

where $U_B(r)$ corresponds to bond stretching potential, modeled by finite extensible nonlinear elastic potential given by,

$$U_B(r) = -0.5\mathrm{K_S}R_0^2 \ln\left[1 - \left(\frac{r}{R_0}\right)^2\right]$$

where $\mathrm{K_S} = 30.0\frac{\epsilon}{\sigma^2}$ is spring constant and $\mathrm{R_0} = 1.6\sigma$ is the equilibrium bond length.

We use Lennard-Jones potential to model steric repulsion and attraction between the beads of different chromatin state with different parameters. For example, the potential between a bead of chromatin state A and that of B is represented below,

$$U_P(r) = 4\epsilon_{\mathrm{AB}}\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

where ϵ_{AB} denotes interaction strength between bead types A and B (STable 2) and σ refers to the size of each bead.

In case of loops, we use harmonic potential between each of the loop anchors $U_L(r)$, represented by,

$$U_L(r) = K_L(r - R_0^L)^2$$

where $\mathrm{K_L} = 300\frac{\epsilon}{\sigma^2}$ is spring constant and $\mathrm{R_0^L} = 1.2\sigma$ is the equilibrium loop bond length.

Finally, in order to model the repulsive interaction between the chromatin beads and the simulation box boundaries, we use Lennard-Jones potential

$$U_W(r) = 4\epsilon_{\mathrm{W}}\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

where $\epsilon_{\mathrm{W}} = 1.0$ denotes the interaction strength between boundary walls and chromatin beads irrespective of their types.

LAMMPS molecular dynamics package is used to perform polymer dynamics simulations (47). All the simulations are performed in terms of Lennard-Jones reduced units with τ for time, σ for length and ϵ for energy. To mimic the viscoelastic nature of the nucleoplasm, Langevin dynamics with a damping coefficient of $\gamma = 1.0\tau$ is used to maintain the temperature at $T = 1.0$. We then run the simulation for a total of 1,000,000 steps with a timestep of $dt = 0.000001\tau$. Followed by that, the timestep dt is increased to 0.01τ and a production run of a total 2,000,000 steps is performed and the conformations are saved at intervals of 1,000 steps. We repeated this process for 20 times, each time with different initial conformation and random seeds.

<u>Initial conditions and Equilibration</u> The chromatin fiber is initialized as a self-avoiding worm like chain using the python package PolymerCpp (48) and is placed in a box of size $L = 35\sigma$ with fixed repulsive boundaries. We then allow the polymer to equilibrate inside the box using four consecutive phases. In the first phase, a soft potential is used to remove overlap in the polymer for 400,000 steps with a $dt = 0.01\tau$.

$$U_S(r) = \mathrm{A}\left[1 + cos\left(\frac{\pi \mathrm{r}}{\mathrm{R}_0^S}\right)^6\right]$$

where A is pre-factor, varied gradually from 0 to 100ϵ and $\mathrm{R}_0^S = 2^{1/6}\sigma$.

For the rest three subsequent phases, we switch to potential to Lennard-Jones potential and run the simulation for 400,000, 500,000 and 500,000 steps with timesteps 0.000001τ, 0.0001τ, and 0.01τ respectively.

$$U_G(r) = 4\epsilon_{\mathrm{G}}\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6\right]$$

where $\epsilon_{\mathrm{G}} = 1.0$ denotes the interaction strength between chromatin beads irrespective of their types.

Analysis of nucleolus-associated domain (NAD) mapping data

In order to characterize nucleolar association level of the IMR90 chromosomal regions, we downloaded the NAD mapping data of the proliferating IMR90 cells from Dillinger *et al*. (49). This microarray data contains log2 normalized ratio of nucleolus-associated and genomic DNA at specific probe locations. For our analysis, we bin that log2 normalized ratio data at 10 kb resolution using multiBigwigSummary functionality of the deepTools software (50).

RESULTS

Chromosome structural landscape reveals major features of the structural ensemble

We first apply our statistical analysis technique to the single-cell chromosome structures measured for a 2 Mb region of IMR90 chr21 (28-30 Mb) and produce a chromosome structural landscape (Fig. 2b). We observe that the first major source of variation is the compactness of the chromosome structures, which can be quantified using the radius of gyration, R_g (Fig. S2a). On the left extreme of PC1 (from here on we refer to the PC1 projection as “PC1”), we have structures that are highly compact with smaller R_g values and as we move towards the right, the structures become less compact with higher R_g scores. With increasing R_g, the structural divergence also increases, with individual structure points becoming spread further along PC2 and also more spread out from each other.

Along PC2, the structure ordering represents a progressive reorganization of the domain and captures several subclusters of structures (Fig. 2c). For example, as we move from negative to positive PC2, we observe a shifting of the boundary (weakening of one boundary and gaining of another) that leads to the change in the structure of the domain (Fig. S2b). This phenomenon can be seen in different regimes of R_g, showing the existence of domain reorganization within both compact and open structures (Fig. S2b).

When applied on other chromosome and cell-type single-cell imaging data at different resolutions, we found the consistent hierarchical ordering of the dynamic modes of the structural landscape (Fig. S2c,

Fig. S2d, Fig. S3 and Fig. S4). Although the major dynamic variation remains the R_g ordering for all the different situations, the reorganization of domains along PC2 is dependent on the length of the genomic region as well as the local structural properties.

Network entropy captures cell-type- and chromosome-specific features of the chromosome structural ensemble

Once we were able to reference individual conformations on the structural landscape spanning by collective coordinates expressed using PC projections, we first inspect cell-type-specific variations associated with chromosome structural landscape. In order to do that, we performed a joint analysis by constructing structural landscape using the 28-30 Mb chr21 region data from three cell types – IMR90 (lung fibroblast), K562 (leukemia) and A549 (lung adenocarcinoma) collectively. From that resultant joint structural landscape, we can clearly visualize the cell-type-specific variation present in structural ensemble. To further characterize that quantitatively, we divide the landscape in 10 equally sized regimes of R_g along PC1 and calculated the proportion of different cell types within each regime. We observe that K562 shows a higher proportion of chromosome structures in lower R_g regimes and has an overall more compact structural ensemble compared to A549 and IMR90 (Fig. 3a and Fig. 3b).

While the relative R_g shift between cell types is straightforward to observe and quantify, other features of the chromosome structural landscape are more difficult to interpret visually. Therefore, we devised a network-entropy based measurement to quantitatively compare the structural landscapes, as discussed in the **Methods** section. According to this approach, a network with a higher number of edges will have a higher entropy value and vice-versa. In terms of the chromosome structural ensemble, a set of structures that smoothly transit across the entire conformational space will have a highly connected structural landscape and thus high landscape entropy (Fig. 1a, blue line). On the other hand, if structures coalesce around discrete states (Fig. 1a, red line), the discontinuity in the structural ensemble will be higher, the resulting network will have fewer edges, and a low entropy score will be produced.

We first applied the network entropy approach to structures of 2 Mb regions of chr21 (30 Kb resolution) in different cell-types and conditions obtained from Bintu *et al*. (29). There is variation in the sample size (i.e., the number of conformations) obtained for different experiments, so we relied on a resampling approach to make the comparison consistent. Based on this approach, when we compared the 28-30 Mb genomic region from three different cell-types analyzed above, we find that IMR90 has the highest landscape entropy, followed by K562 and then A549 with the lowest entropy (Fig. 3c). Thus, this region in IMR90 cells has the highest continuity across different structures while the corresponding distribution in A549 cells exists as a relatively discrete set of structures. In a similar fashion, when we compared two different genomic regions of IMR90 chr21, we found that although both are from same cell-type and same chromosome, the 28-30 Mb region has lower entropy values compared to 18-20 Mb region (Fig. 3c). This difference indicates that the latter region has a higher degree of continuity in its structural ensemble. This

18-20 Mb region lies entirely in the B compartment (putative inactive, gene poor regions) while the 28-30 Mb region contains a majority of A compartment regions mixed with B compartment regions. This led us to investigate the possibility that when a chromatin segment resides completely within a compact, inactive region, the single cell structures exhibit a more continuous structural landscape and higher ensemble entropy level. To test this connection, we quantified the degree of continuity present in whole active and inactive chrX structural ensembles (440 Kb to 7.8 Mb) using the network entropy approach (Fig. 3d). We can observe that indeed, the chrXi has higher ensemble entropy values compared to its active counterpart. This chromosome which has a highly compact global structure has a higher landscape entropy than a specific, relatively open structure. This reiterates that the open structure is not that of an equilibrated random coil ensemble but instead exists in more discrete structures (Fig. S5a, Fig. S5b and Fig. S5c) (51).

Furthermore, intrigued by whole chromosomal level observation for chrX, we analyzed the structural variation of the entire chr2 and chr21 q-arm obtained from Su *et al*. (30) at higher resolution of 50 Kb (Fig. 3e). We observe that chr21 has a higher structural ensemble entropy compared to chr2. To test whether the difference in length between chr2 and chr21 contributed to their differences in entropy, we calculated the ensemble entropy of regions of different lengths selected from chr2. We observed that as the length increases the ensemble entropy decreases. (Fig. S5d). This effect likely stems from the increased number of possible conformations that a longer polymer can adopt, which corresponds to a poorer coverage of conformational space. We also found that ensemble entropy is lower in S/G2 phase compared to G1 phase for chr21. This result is also consistent with the entropies of 28-30 Mb IMR90 chr21 region structural ensembles from different cell-cycle stages, where the G1 phase has highest ensemble entropy followed by S and G2 phases, as shown in Fig. 3a. One major difference between the chromosomes in the phases of the cell cycle is the addition of tethers of chromosomal regions to sister chromatids in S and G2 compared to G1 (52). This again suggests that increased constraints on the chromosome structure would correspond to highly distinct structures and lower structural ensemble entropy.

Contribution of epigenetic properties to the cell-type-specific variation of chromosome structural ensemble

Building on the above observation that regions with different A/B spatial compartment identity had very different structural landscape entropy, we next examined the influence of local epigenetic state on the resulting structural landscape. Recent studies have shown the association of epigenetic states with the ensemble genome organization (53-55). Figure 4a shows the chromatin states of the 28-30 Mb spanning genomic region in the three different cell types – IMR90, K562 and A549 (detailed steps in the **Methods** section). We observe that the region in IMR90 mostly has active marks while in K562 the region has a higher amount of repressive polycomb and heterochromatin marks. The A549 region shows a mixture of active and repressive marks. By qualitatively comparing this chromatin state data with the ensemble

entropy distributions for these three cell types, these results appear to contradict our earlier hypothesis that inactive states lead to higher entropy, since here, the two cell types with more inactive states have a lower entropy. Therefore, these results help us refine our understanding of the relationship between overall structural ensemble entropy and epigenetic state: the more uniform epigenetic states (whether mostly active or mostly inactive) will generate smoother structural landscapes while regions with mixed epigenetic states exhibit more constrained structures with lower entropy.

In addition to the overall dynamic properties of the region, we sought to quantitatively compare local genomic bin epigenetic state with the degree of structure fluctuation experienced by that bin. We first calculated the local bin entropy standard deviation for each cell type (**Methods**) which shows the variability of a particular genomic region across the population of cells. However, as our chromatin state information is categorical and has five categories, we also digitized the standard deviation values into five groups (Fig. 4b). We do not find a significant correlation between these properties at the bin level (30 kb). But, when we compare the percentage of different chromatin states to the percentage of digitized local bin entropy standard deviation level for each cell type, we see a stronger association (Fig. 4c and d). Again, this result supports the idea that it is the overall epigenetic state diversity of the region that associates with structural fluctuation rather than dynamics of each isolated bin being determined by its local epigenetic state. Throughout this comparison, it is important to remember that we are comparing a population averaged epigenetic state with single cell structural fluctuation. It is possible that additional relationships could become evident if single cell variation in epigenetic state were considered.

To test our hypothesis that epigenetic state differences contribute to cell-type-specific variation in the chromosome structural ensemble, we performed a set of polymer simulations of the desired genomic region of 28-29.5 Mb at a resolution of 1 kb, as described in the **Methods** section. We generated the models for IMR90 and K562 separately by assigning interactions between the polymer beads based on their specific chromatin states (Fig. S6a). Once we obtained the 3D conformations from the simulations, we constructed pairwise Euclidean distance matrices for each conformation and further quantified the structural ensemble using our network entropy approach. For the resampling purpose, we combined all the snapshots from 20 independent simulations and then selected 500 conformations randomly from this set 1000 times. When we compared the simulated IMR90 entropy distribution with simulated K562 data, we found that the K562 structural ensemble shows a lower entropy, which is the same trend observed in our analysis of the Bintu *et al*. imaging data (Fig. 3c and Fig. 4f). Since the only difference between cell types in these simulations was the encoded epigenetic state of the polymer beads, this result corroborates the idea that the epigenetic state is one factor that contributes to the variation in chromosome structural ensemble.

Analysis of the genomic regions extracted from the whole chromosome imaging data validates the effect of epigenetic and other genomic and structural properties

We further evaluated the contribution of epigenetic factors to structural dynamics by analyzing single-cell structures of six genomic regions, each 2 Mb in length, obtained from IMR90 chr21 whole q-arm imaging data (Fig. 5a) (30). For simplicity, the regions will be denoted as S1 to S6 based on their starting genomic coordinates. The regions S1 (chr21:18-20 Mb) and S2 (chr21:28-30 Mb) were selected in order to match the two IMR90 genomic regions data analyzed in Bintu *et al*. Next, the three regions S3, S5 and S6 were selected based on their epigenetic state similarity with epigenetic state of the chr21:28-30 Mb region in K562 cells (Fig. S6b). To evaluate epigenetic state similarity, we designed a sliding window-based approach where we move a window of size 40 (2 Mb genomic region with 50 kb resolution) over the whole q-arm of chr21 and for each window compare the IMR90 epigenetic state with the K562 28-30 Mb region epigenetic state using binary comparison (Fig. S6c). However, as the resolution of the Bintu *et al*. data (65 bins per 2 Mb region) is different from the Su *et al*. data (40 bins per 2 Mb region), we take the least common multiple of the two bin numbers and expanded each data to 250 bins by repeating each bin element required number of times. Here, a higher comparison value shows higher similarity with the K562 epigenetic state. Among the three regions, S3 has the highest score, S6 has the second highest and S5 has the third highest one (Fig. S5c). The region S4 was selected randomly.

When we quantified the variation in the structural ensemble of these six regions, we first observed that the S1 and S2 regions have consistent entropy distribution with the Bintu *et al*. data (Fig. 5b). Next, we found that the S3 region has the lowest entropy distribution among all the six regions which is also consistent with the low entropy of the Bintu *et al*. K562 region. In addition, the ordering of S6 and S5 are also consistent with their similarity with K562 epigenetic state. Finally, the S4 region, which has a very low similarity score with K562 state, exhibits a very high ensemble entropy distribution. Again, this analysis supports the idea that epigenetic patterns contribute to the degree of variation present in the chromosome structural ensemble.

In addition to the epigenetic states, we also called domain boundaries in the whole q-arm of chr21 using the HiCRatio approach (with a 300 kb window size) (56). Since this boundary calling software is designed for contact data, we first converted the single-cell chromosome structural distance data into contact data using a cutoff of 500 nm, as reported here. We found that regions with a higher number of strong domain boundaries exhibit lower ensemble entropy and vice versa (Fig. 5c). For example, the S3 region, which has the lowest ensemble entropy, has three strong domain boundaries and one weak. On the other hand, regions like S1 and S4 which exhibit higher ensemble entropies have fewer or weaker domain boundaries. This suggests that the more constraints (here domain boundaries) a genomic region has, the more specific distinct structures the ensemble exhibits, which eventually leads to a lower structural ensemble entropy.

Finally, we asked whether chromosome association with the nucleolus could contribute to the variation of chromosome structural ensemble entropy (Fig. 5d). The NAD data source and processing steps are described in the **Methods** section. We observe that S1 and S4, the regions with the highest structural

ensemble entropy, are the most uniform in their NAD state: S1 is completely associated with the nucleolus, while S4 has almost no association with the nucleolus. As seen with epigenetic state, this suggests that uniformity of association leads to a higher continuity level in the structural ensemble. In contrast, S3 has the lowest entropy and also the alternating levels of nucleolus association in the first half of the genomic region, followed by an unconstrained region. This variation in constraints across the region likely leads to specific distinct structures and thus a lower ensemble entropy. Overall, all these analyses suggest that as the number of different states and constraints increases, the ensemble becomes more highly distinct and discontinuous and vice versa.

The effect of cohesin depletion on structural dynamics depends on the underlying constraints on the chromosome region

In order to study how the chromosome structural landscape and ensemble varies due to removal of cohesin, a key architectural protein implicated in loop extrusion and TAD formation (57), we analyzed chr21 single-cell structural data from HCT116 cells in wildtype (WT) and cohesin depleted (+Auxin) conditions (29). For this analysis, we looked into two different genomic regions (chr21:28-30 Mb and chr21:34-37 Mb) with different underlying structural properties. The jointly constructed chromosome structural landscape for the first genomic region shows that the WT and cohesin-depleted structure ensembles mostly overlap with a slight bias for the +Auxin structures in the higher R_g regime (Fig. 6c). This suggests that removal of cohesin makes the structures less compact to some extent (58). On the other hand, for the second region, the jointly constructed landscape shows a shift upwards along PC2 as well as a rightward shift towards the large R_g regime for cohesin-depleted structures (Fig. 6d). In addition, we found that after the cohesin removal both the genomic regions become more variable as shown by local bin entropy standard deviation (Fig. S6d and Fig. S6e). When we examine the structural landscape entropy of these two regions, we see apparently conflicting effects of cohesin removal: the entropy is decreased in the 28-30 Mb region but mostly unchanged in the 34-37 Mb region (Fig 6e). Note that the initial entropy of these regions in the WT follows the expected pattern for regions of different lengths. We hypothesize that, for the first region, the change with cohesin removal relates to the fact that compartmentalization and loop extrusion can act in opposition to each other (59). In the first region, in the presence of cohesin, the entire region has a chance of being pulled into the same loop, increasing the uniformity of the chromatin state, and creating a highly continuous landscape of possible structures (high entropy). But, when cohesin is absent, a subdomain can form more strongly within this region based on epigenetic state, as has been seen to happen throughout the genome in the absence of cohesin (60). This alternating state now creates more discrete structures that exhibit less continuity without cohesin to work against this compartmentalization. In contrast, the 34-37 Mb region, which lies in an active compartment, exhibits a mostly unchanged entropy distribution before and after cohesin removal. This shows that the structural ensemble maintains a similar degree of continuity after the elimination of TADs via cohesin removal. As before, cohesin removal is expected to reduce the active movement of

chromatin, but in this case, the region is entirely in the same compartment, so there are no additional constraints that would lead to the formation of discrete states in the cohesin-depleted structural landscape. These results overall show that upon cohesin removal the chromosome structural ensemble changes depend on the underlying epigenetic state.

DISCUSSION

As the number and quality of single cell chromosome structural datasets continues to increase, a diverse toolbox of methods to capture properties of chromosome structure variation is needed. Our statistical analysis of single cell chromosome structures allows us to characterize both the major modes of chromosome structure fluctuation and the properties of the structure fluctuation landscape. While some approaches focus on classifying single cell structures into discrete clusters, other recent work has shown that chromosome structures can exist along a continuum (22,31,33-35). Our analysis does not require a decision to be made about how many different clusters of structures exist, but instead can capture the overall types of structure variations and can measure the degree of continuity between the structures.

For all observed length scales of chromosome structures, from local TAD scale to whole chromosome arms, we find that the primary type of single cell structure variation is the overall compaction of the region. For each structure assayed, there exist some cells in which the region is very compact and others in which it is much more extended. This overall size fluctuation is true of both active and inactive regions, though inactive regions and the inactive X chromosome show a shift toward smaller sizes overall. However, we find that this overall region compaction is largely independent of a secondary fluctuation mode: the reorganization of domains. No matter how compact or extended the overall region is, we observe a similar progressive reorganization of the domain boundaries as a secondary type of structure variation.

By quantifying the network entropy-based properties of the structural ensemble, we can measure how discrete vs. continuous the structures adopted by a certain chromosome region are. We find that the entropy of a structural ensemble is influenced by the constraints put on the region by epigenetic state, domain boundaries formed by cohesin loop extrusion, and association with the nucleolus. Overall, constraints on the chromosome structure and boundaries between different types of regions decrease structural entropy and result in more discrete structures while regions with more uniform linear chromatin states exhibit a more continuous structural landscape with higher entropy.

It might seem likely that the active or inactive chromatin status of a region would be the primary contribution of epigenetics to the structural fluctuation of a region. That is, one might expect an inactive region to have high structural entropy and an active region low entropy or vice versa. However, we find that this is not the case. Instead, the uniformity of the epigenetic state appears to be a key factor, such that a uniformly active region shares structural landscape properties with a uniformly inactive region (both

have high entropy) while a region divided between alternating states behaves differently (low entropy). This leads us to the idea that alternating epigenetic states place constraints on the possible conformations and thus create a more disjointed, rather than continuous, structural ensemble.

The idea that constraints are key to shaping dynamic structural landscapes is echoed again by the results we observe in cohesin depletion. When cohesin is removed, regions consisting of multiple epigenetic states are more likely to adopt more discrete structures, resulting in a decrease in entropy compared to the state where cohesin is active. In contrast, when a region exists in a uniform epigenetic state, the removal of cohesin does not result in a reduction of entropy. This observation is consistent with the idea that cohesin loop extrusion acts to increase chromosome conformation dynamics, promoting distal contacts and countering the tendency of domains to settle into more discrete sub-structures (59,60).

The analysis of cohesin depletion results also reiterate the value of considering both the structural landscape and the entropy of that landscape. The structural landscape of a region can change (adopting a different range of radius of gyration or domain structures) while the continuity (entropy) stays the same. Alternately, the entropy can change while the structural landscape is overall conserved, indicating cases where the range and types of structures a region can adopt are the same, but the region is more often in certain discrete structures within that range rather than continuously sampling the range.

AVAILABILITY

Example analysis code is provided on the ChromosomeStructuralEnsembleCharacterization GitHub repository (https://github.com/rpmccordlab/ChromosomeStructuralEnsembleCharacterization).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Rebeca San Martin for valuable suggestions and insightful discussions in the development of the research project.

FUNDING

This work was supported by the National Institutes of Health [NIGMS grant R35GM133557 to R.P.M]. Funding for open access charge: National Institutes of Health/ R35GM133557.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

1. McCord, R.P., Kaplan, N. and Giorgetti, L. (2020) Chromosome conformation capture and beyond: toward an integrative view of chromosome structure and function. *Molecular cell*, **77**, 688-708.
2. Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N. and Mirny, L.A. (2016) Formation of chromosomal domains by loop extrusion. *Cell reports*, **15**, 2038-2049.
3. Nichols, M.H. and Corces, V.G. (2015) A CTCF code for 3D genome architecture. *Cell*, **162**, 703-705.
4. Sanborn, A.L., Rao, S.S., Huang, S.-C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A. and Li, J. (2015) Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences*, **112**, E6456-E6465.
5. Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A. and Cavalli, G. (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, **148**, 458-472.
6. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., Van Berkum, N.L., Meisig, J. and Sedat, J. (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, **485**, 381-385.
7. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S. and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376-380.
8. Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J. and Dorschner, M.O. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science*, **326**, 289-293.
9. Falk, M., Feodorova, Y., Naumova, N., Imakaev, M., Lajoie, B.R., Leonhardt, H., Joffe, B., Dekker, J., Fudenberg, G. and Solovei, I. (2019) Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature*, **570**, 395-399.
10. Sanulli, S., Justin, N., Teissandier, A., Ancelin, K., Portoso, M., Caron, M., Michaud, A., Lombard, B., da Rocha, S.T. and Offer, J. (2015) Jarid2 methylation via the PRC2 complex regulates H3K27me3 deposition during cell differentiation. *Molecular cell*, **57**, 769-783.
11. Wang, L., Gao, Y., Zheng, X., Liu, C., Dong, S., Li, R., Zhang, G., Wei, Y., Qu, H. and Li, Y. (2019) Histone modifications regulate chromatin compartmentalization by contributing to a phase separation mechanism. *Molecular cell*, **76**, 646-659. e646.
12. Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X. and Karpen, G.H. (2017) Phase separation drives heterochromatin domain formation. *Nature*, **547**, 241-245.
13. Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., Müller, S., Eils, R., Cremer, C. and Speicher, M.R. (2005) Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS biology*, **3**, e157.

14. Cremer, T. and Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature reviews genetics*, **2**, 292-301.
15. Stack, S.M., Brown, D.B. and Dewey, W. (1977) Visualization of interphase chromosomes. *Journal of cell science*, **26**, 281-299.
16. Kadauke, S. and Blobel, G.A. (2009) Chromatin loops in gene regulation. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, **1789**, 17-25.
17. Braccioli, L. and de Wit, E. (2019) CTCF: a Swiss-army knife for genome organization and transcription regulation. *Essays in biochemistry*, **63**, 157-165.
18. Zuin, J., Roth, G., Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., Mach, P., Kryzhanovska, M., Tihanyi, G., Kohler, H. *et al.* (2021) Nonlinear control of transcription through enhancer-promoter interactions. *bioRxiv*, 2021.2004.2022.440891.
19. Clapier, C.R., Iwasa, J., Cairns, B.R. and Peterson, C.L. (2017) Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature reviews Molecular cell biology*, **18**, 407-422.
20. Kim, Y. and Yu, H. (2020) Shaping of the 3D genome by the ATPase machine cohesin. *Experimental & Molecular Medicine*, **52**, 1891-1897.
21. Kim, J.M., Visanpattanasin, P., Jou, V., Liu, S., Tang, X., Zheng, Q., Li, K.Y., Snedeker, J., Lavis, L.D. and Lionnet, T. (2021) Single-molecule imaging of chromatin remodelers reveals role of ATPase in promoting fast kinetics of target search and dissociation from chromatin. *bioRxiv*.
22. Takei, Y., Zheng, S., Yun, J., Shah, S., Pierson, N., White, J., Schindler, S., Tischbirek, C.H., Yuan, G.-C. and Cai, L. (2021) Single-cell nuclear architecture across cell types in the mouse brain. *Science*, eabj1966.
23. Kempfer, R. and Pombo, A. (2020) Methods for mapping 3D chromosome architecture. *Nature Reviews Genetics*, **21**, 207-226.
24. Jerkovic, I. and Cavalli, G. (2021) Understanding 3D genome organization by multidisciplinary methods. *Nature Reviews Molecular Cell Biology*, 1-18.
25. Ramani, V., Deng, X., Qiu, R., Gunderson, K.L., Steemers, F.J., Disteche, C.M., Noble, W.S., Duan, Z. and Shendure, J. (2017) Massively multiplex single-cell Hi-C. *Nature methods*, **14**, 263-266.
26. Tan, L., Ma, W., Wu, H., Zheng, Y., Xing, D., Chen, R., Li, X., Daley, N., Deisseroth, K. and Xie, X.S. (2021) Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. *Cell*, **184**, 741-758. e717.
27. Arrastia, M.V., Jachowicz, J.W., Ollikainen, N., Curtis, M.S., Lai, C., Quinodoz, S.A., Selck, D.A., Ismagilov, R.F. and Guttman, M. (2021) Single-cell measurement of higher-order 3D genome organization with scSPRITE. *Nature Biotechnology*, 1-10.
28. Wang, S., Su, J.-H., Beliveau, B.J., Bintu, B., Moffitt, J.R., Wu, C.-t. and Zhuang, X. (2016) Spatial organization of chromatin domains and compartments in single chromosomes. *Science*, **353**, 598-602.
29. Bintu, B., Mateo, L.J., Su, J.-H., Sinnott-Armstrong, N.A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A.N. and Zhuang, X. (2018) Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, **362**, eaau1783.
30. Su, J.-H., Zheng, P., Kinrot, S.S., Bintu, B. and Zhuang, X. (2020) Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. *Cell*, **182**, 1641-1659. e1626.
31. Sawh, A.N., Shafer, M.E., Su, J.-H., Zhuang, X., Wang, S. and Mango, S.E. (2020) Lamina-dependent stretching and unconventional chromosome compartments in early C. elegans embryos. *Molecular cell*, **78**, 96-111. e116.
32. Mateo, L.J., Murphy, S.E., Hafner, A., Cinquini, I.S., Walker, C.A. and Boettiger, A.N. (2019) Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature*, **568**, 49-54.

33. Cheng, R.R., Contessoto, V.G., Aiden, E.L., Wolynes, P.G., Di Pierro, M. and Onuchic, J.N. (2020) Exploring chromosomal structural heterogeneity across multiple cell lines. *Elife*, **9**, e60312.
34. Xie, W.J., Qi, Y. and Zhang, B. (2020) Characterizing chromatin folding coordinate and landscape with deep learning. *PLoS computational biology*, **16**, e1008262.
35. Conte, M., Fiorillo, L., Bianco, S., Chiariello, A.M., Esposito, A. and Nicodemi, M. (2020) Polymer physics indicates chromatin folding variability across single-cells results from state degeneracy in phase separation. *Nature communications*, **11**, 1-13.
36. Zhu, C., Zhang, Y., Li, Y.E., Lucero, J., Behrens, M.M. and Ren, B. (2021) Joint profiling of histone modifications and transcriptome in single cells from mouse brain. *Nature Methods*, **18**, 283-292.
37. Chen, H., Tian, Y., Shu, W., Bo, X. and Wang, S. (2012) Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PloS one*, **7**, e41374.
38. Bersaglieri, C., Kresoja-Rakic, J., Gupta, S., Bär, D., Kuzyakiv, R. and Santoro, R. (2020) Genome-wide maps of nucleolus interactions reveal distinct layers of repressive chromatin domains. *bioRxiv*, 2020.2011.2017.386797.
39. Cheng, F., Liu, C., Shen, B. and Zhao, Z. (2016) Investigating cellular network heterogeneity and modularity in cancer: a network entropy and unbalanced motif approach. *BMC systems biology*, **10**, 301-311.
40. Banerji, C.R., Miranda-Saavedra, D., Severini, S., Widschwendter, M., Enver, T., Zhou, J.X. and Teschendorff, A.E. (2013) Cellular network entropy as the energy potential in Waddington's differentiation landscape. *Scientific reports*, **3**, 1-7.
41. Menichetti, G., Bianconi, G., Castellani, G., Giampieri, E. and Remondini, D. (2015) Multiscale characterization of ageing and cancer progression by a novel network entropy measure. *Molecular BioSystems*, **11**, 1824-1831.
42. Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K. and Narayanan, A.K. (2018) The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic acids research*, **46**, D794-D801.
43. Ernst, J. and Kellis, M. (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nature methods*, **9**, 215-216.
44. Chiang, M., Michieletto, D., Brackley, C.A., Rattanavirotkul, N., Mohammed, H., Marenduzzo, D. and Chandra, T. (2019) Polymer modeling predicts chromosome reorganization in senescence. *Cell reports*, **28**, 3212-3223. e3216.
45. Shi, G., Liu, L., Hyeon, C. and Thirumalai, D. (2018) Interphase human chromosome exhibits out of equilibrium glassy dynamics. *Nature communications*, **9**, 1-13.
46. Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D. and Lander, E.S. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665-1680.
47. Plimpton, S. (1995) Fast parallel algorithms for short-range molecular dynamics. *Journal of computational physics*, **117**, 1-19.
48. Stefko, M. and Douglass, K.M. (2017).
49. Dillinger, S., Straub, T. and Nemeth, A. (2017) Nucleolus association of chromosomal domains is largely maintained in cellular senescence despite massive nuclear reorganisation. *PLoS One*, **12**, e0178821.
50. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A. and Manke, T. (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic acids research*, **42**, W187-W191.
51. Cheng, Y., Liu, M., Hu, M. and Wang, S. (2021) TAD-like single-cell domain structures exist on both active and inactive X chromosomes and persist under epigenetic perturbations. *bioRxiv*, 2021.2005.2012.443887.

52. Mitter, M., Gasser, C., Takacs, Z., Langer, C.C., Tang, W., Jessberger, G., Beales, C.T., Neuner, E., Ameres, S.L. and Peters, J.-M. (2020) Conformation of sister chromatids in the replicated human genome. *Nature*, **586**, 139-144.
53. Szabo, Q., Bantignies, F. and Cavalli, G. (2019) Principles of genome folding into topologically associating domains. *Science advances*, **5**, eaaw1668.
54. Di Pierro, M., Cheng, R.R., Aiden, E.L., Wolynes, P.G. and Onuchic, J.N. (2017) De novo prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture. *Proceedings of the National Academy of Sciences*, **114**, 12126-12131.
55. Di Pierro, M., Potoyan, D.A., Wolynes, P.G. and Onuchic, J.N. (2018) Anomalous diffusion, spatial coherence, and viscoelasticity from the energy landscape of human chromosomes. *Proceedings of the National Academy of Sciences*, **115**, 7753-7758.
56. Lazaris, C., Kelly, S., Ntziachristos, P., Aifantis, I. and Tsirigos, A. (2017) HiC-bench: comprehensive and reproducible Hi-C data analysis designed for parameter exploration and benchmarking. *BMC genomics*, **18**, 1-16.
57. Davidson, I.F. and Peters, J.-M. (2021) Genome folding through loop extrusion by SMC complexes. *Nature Reviews Molecular Cell Biology*, **22**, 445-464.
58. Hao, X., Parmar, J.J., Lelandais, B., Aristov, A., Ouyang, W., Weber, C. and Zimmer, C. (2021) Super-resolution visualization and modeling of human chromosomal regions reveals cohesin-dependent loop structures. *Genome biology*, **22**, 1-20.
59. Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. and Mirny, L.A. (2018) Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proceedings of the National Academy of Sciences*, **115**, E6697-E6706.
60. Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C.H. and Mirny, L. (2017) Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, **551**, 51-56.

TABLE AND FIGURES LEGENDS

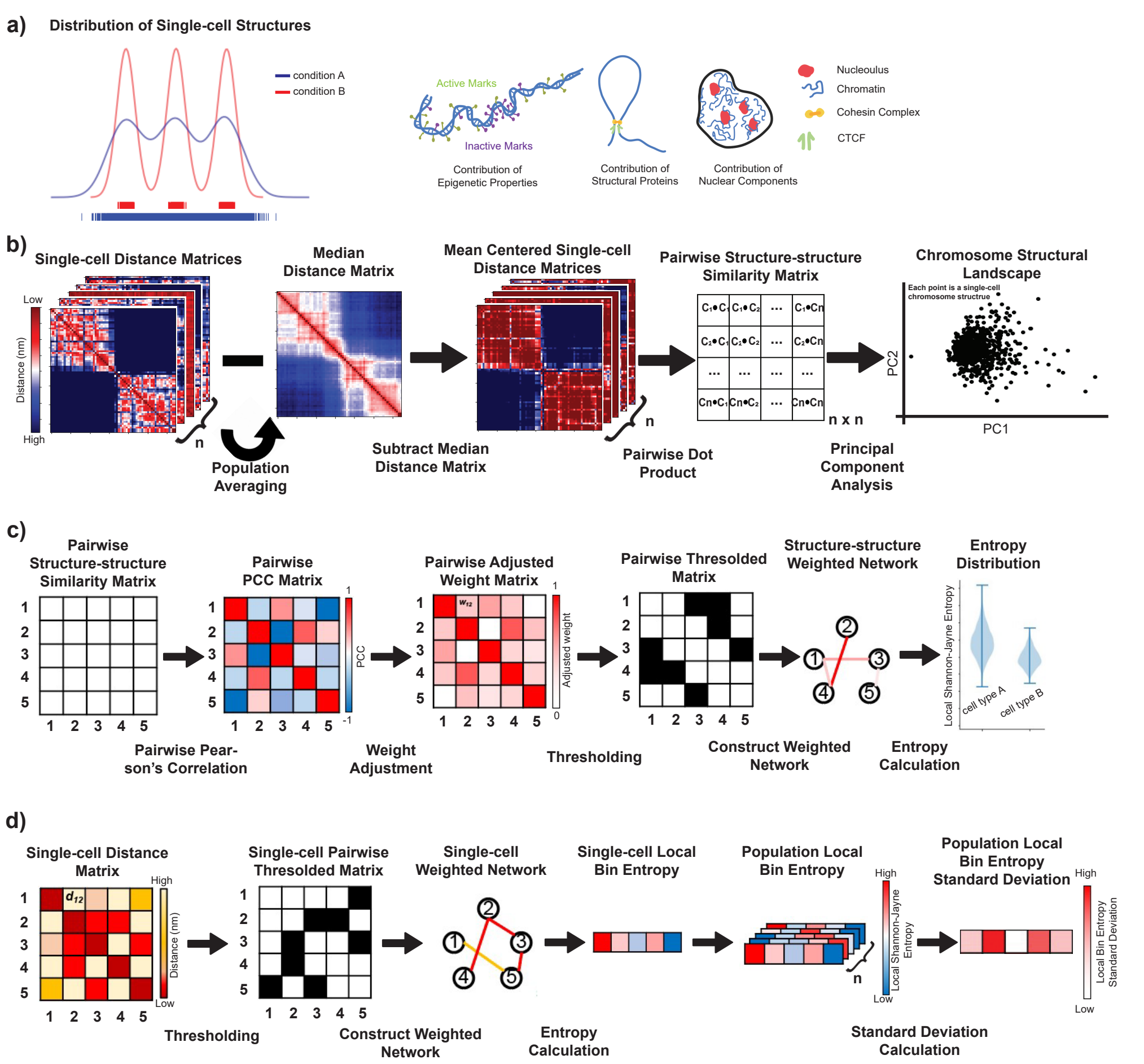
a)
Distribution of Single-cell Structures
condition A
condition B
Active Marks
Inactive Marks
Contribution of Epigenetic Properties
Contribution of Structural Proteins
Contribution of Nuclear Components
Nucleoulus
Chromatin
Cohesin Complex
CTCF
b)
Single-cell Distance Matrices
Low
High
Distance (nm)
n
Median Distance Matrix
Population Averaging
Subtract Median Distance Matrix
Mean Centered Single-cell Distance Matrices
n
Pairwise Dot Product
Pairwise Structure-structure Similarity Matrix
n x n
Principal Component Analysis
Chromosome Structural Landscape
Each point is a single-cell chromosome structrue
PC2
PC1
c)
Pairwise Structure-structure Similarity Matrix
Pairwise Pear-son's Correlation
Pairwise PCC Matrix
PCC
Weight Adjustment
Pairwise Adjusted Weight Matrix
Adjusted weight
Thresholding
Pairwise Thresolded Matrix
Construct Weighted Network
Structure-structure Weighted Network
Entropy Calculation
Entropy Distribution
Local Shannon-Jayne Entropy
cell type A
cell type B
d)
Single-cell Distance Matrix
High
Low
Distance (nm)
Thresholding
Single-cell Pairwise Thresolded Matrix
Construct Weighted Network
Single-cell Weighted Network
Entropy Calculation
Single-cell Local Bin Entropy
Population Local Bin Entropy
n
Local Shannon-Jayne Entropy
Standard Deviation Calculation
Population Local Bin Entropy Standard Deviation
Local Bin Entropy Standard Deviation

Figure 1. Schematic representation of the statistical pipeline used to quantify conformational diversity presented in single-cell distance matrices. **a)** Cartoon representation of chromosome structure ensemble diversity associated with different conditions and factors that may influence this chromosome structure landscape. **b)** A pairwise similarity score is calculated between each pair of median-centered distance matrices. PCA is performed on the pairwise structure-structure similarity matrix and transformed data is projected in 2D space using PC1 and PC2 to create the "chromosome structural landscape". **c)** To calculate chromosome structural ensemble entropy, we construct a Pearson's correlation matrix from the pairwise structure-structure similarity matrix. This correlation matrix is then scaled to range between 0 and 1 and thresholding is performed. The thresholded matrix is converted into a weighted network and entropy values of each of the nodes are calculated using local Shannon-Jayne entropy calculation. **d)** To calculate local bin entropy standard deviation, for each single-cell, we convert the distance matrix into a thresholded matrix using a cutoff and then construct a network from that matrix. We calculate local Shannon-Jayne entropy of each node to represent the local bin entropy. These steps are repeated for all the single-cells across the population and the standard deviation of local bin entropy across population is calculated.

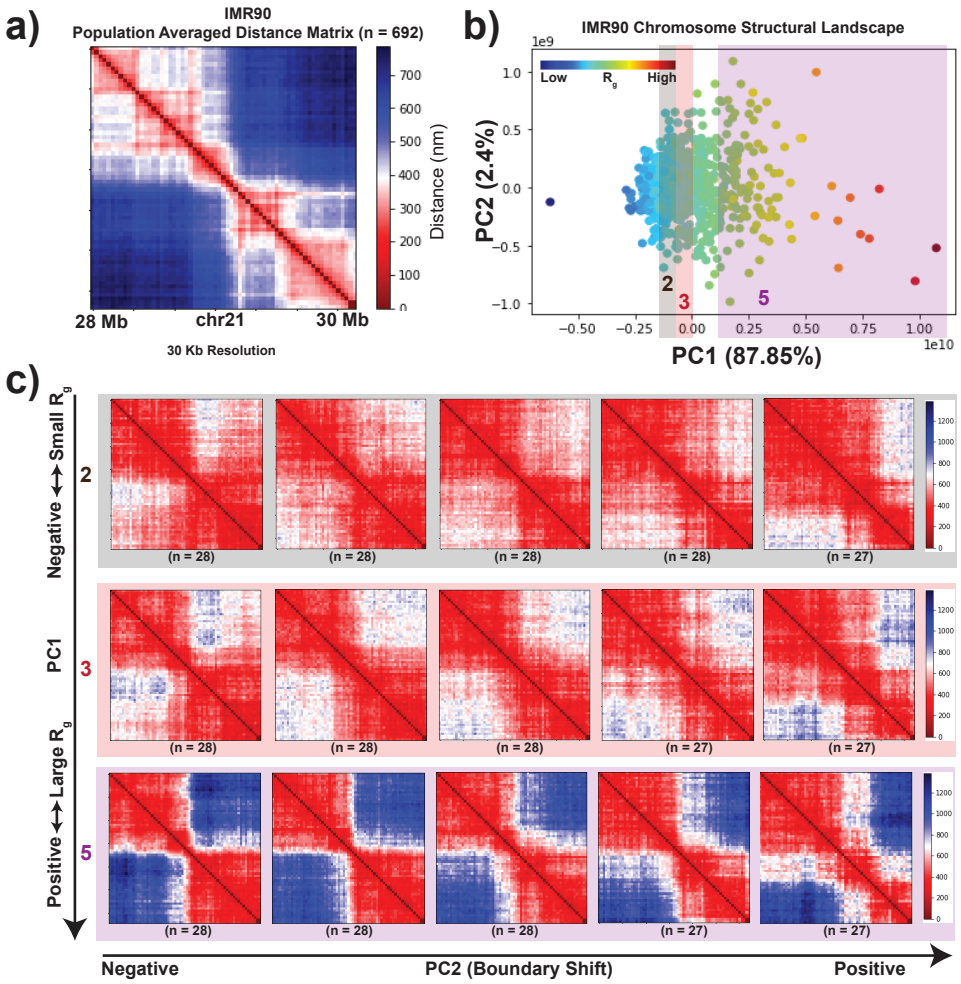
a)
IMR90
Population Averaged Distance Matrix (n = 692)
Distance (nm)
28 Mb
chr21
30 Mb
30 Kb Resolution
b)
IMR90 Chromosome Structural Landscape
Low
R_g
High
PC2 (2.4%)
PC1 (87.85%)
c)
Negative ← → Small R_g
PC1
Positive ← → Large R_g
(n = 28)
(n = 27)
Negative
PC2 (Boundary Shift)
Positive

Figure 2. The chromosome structural landscape orders single-cell structures on a continuum based on their radius of gyration followed by domain reorganization. **a)** Population-averaged distance map of IMR90 chr21:28-30 Mb at 30 kb resolution. **b)** Chromosome structural landscape of IMR90 chr21:28-30 Mb chromatin region, generated as described in the **Methods** section. Each point represents a single-cell structure. Points are colored according to the radius of gyration (R_g) of the corresponding structures (dark blue – smaller R_g and dark red – larger R_g). **c)** Structures are first divided into five groups based on the PC1 ordering (which correlates highly with R_g). Structures belonging to groups 2, 3 and 5 are shown. Within each group structures are further divided into five subgroups based on their PC2 ordering. Average structures within each subgroup are shown.

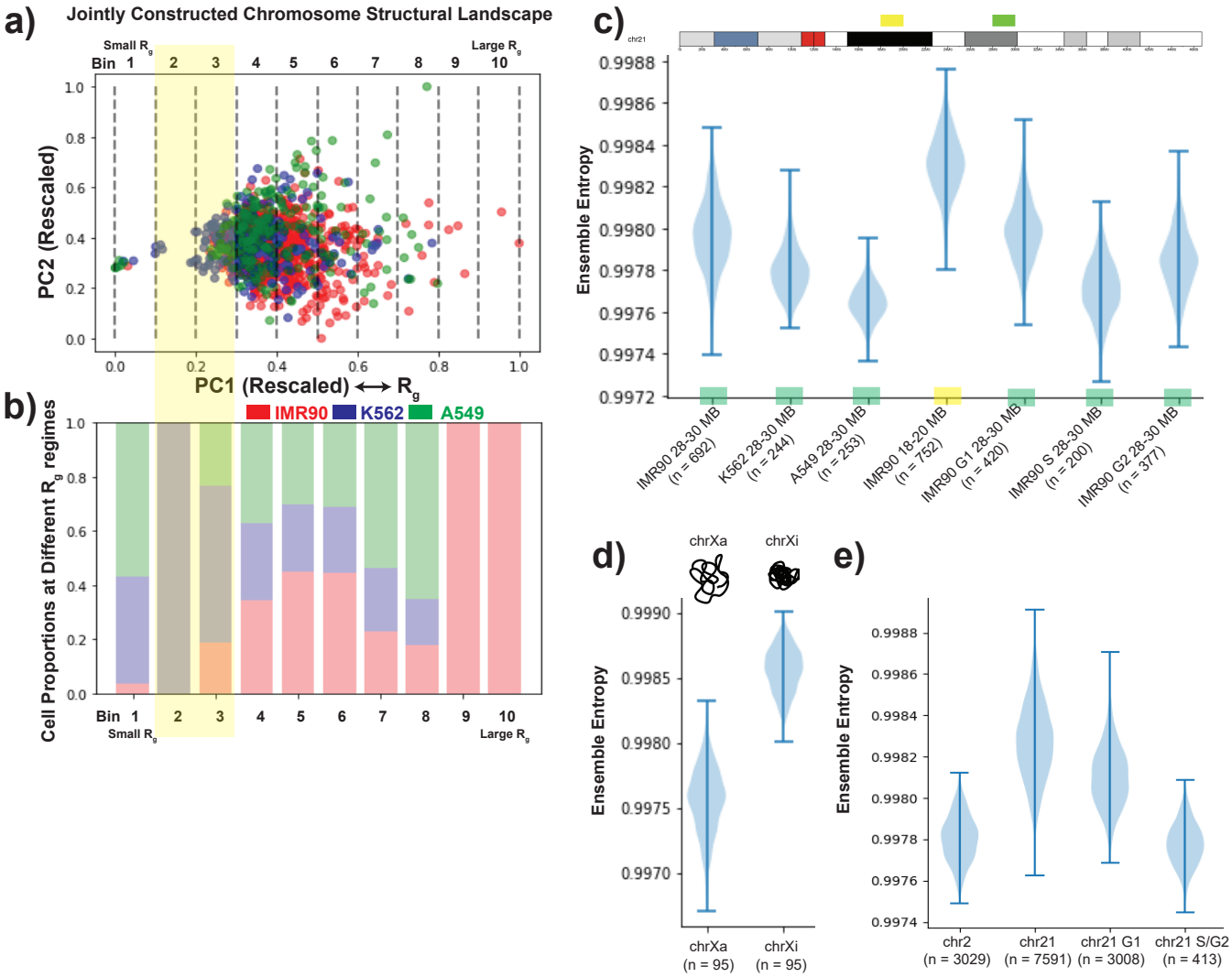
a)
Jointly Constructed Chromosome Structural Landscape
Small R_g
Large R_g
Bin 1 2 3 4 5 6 7 8 9 10
PC2 (Rescaled)
PC1 (Rescaled) ⟷ R_g
b)
IMR90 K562 A549
Cell Proportions at Different R_g regimes
c)
chr21
Ensemble Entropy
IMR90 28-30 MB (n = 692)
K562 28-30 MB (n = 244)
A549 28-30 MB (n = 253)
IMR90 18-20 MB (n = 752)
IMR90 G1 28-30 MB (n = 420)
IMR90 S 28-30 MB (n = 200)
IMR90 G2 28-30 MB (n = 377)
d)
chrXa
chrXi
Ensemble Entropy
chrXa (n = 95)
chrXi (n = 95)
e)
Ensemble Entropy
chr2 (n = 3029)
chr21 (n = 7591)
chr21 G1 (n = 3008)
chr21 S/G2 (n = 413)

Figure 3. Network based entropy quantifies cell-type- and chromosome-specific variability present in chromosome structural ensemble at different resolutions. **a)** Jointly constructed chromosome structural landscape of genomic region chr21:28-30 Mb for cell types IMR90, K562 and A549. The only difference between individual and joint landscape is that all the different cell-type structures are considered together while calculating pairwise structure-structure similarity matrix. Points are colored based on the cell types. Further, the landscape is divided into ten equal sized bins (or regions) based on PC1 values, where the bins on the left represent lower R_g regions and bins on the right are with higher R_g. **b)** Relative proportion of IMR90, K562 and A549 single-cell structures present in each bin represented in a). **c)** Chromosome structural ensemble entropy distributions for different 2 Mb chromosome regions from different cell-types and conditions, obtained from Bintu *et al*. (29). The entropy value of the structural ensemble is calculated using a network entropy based approach. For each structural ensemble, 200 single-cell structures are selected randomly (except IMR90 S 28-30 Mb where we select 150 structures for resampling), and entropy values are calculated. This step is repeated 1000 times and values are plotted as a distribution. **d)** Chromosome structural ensemble entropy distributions for active and inactive IMR90 chrX at TAD resolution, obtained from Wang *et al*. (28). 50 single-cell structures are selected randomly each time for resampling. **e)** Chromosome structural ensemble entropy distributions for different IMR90 chromosomes and conditions at higher resolution, obtained from Su *et al.* (30). Here, for resampling 300 single-cell structures are selected randomly each time.

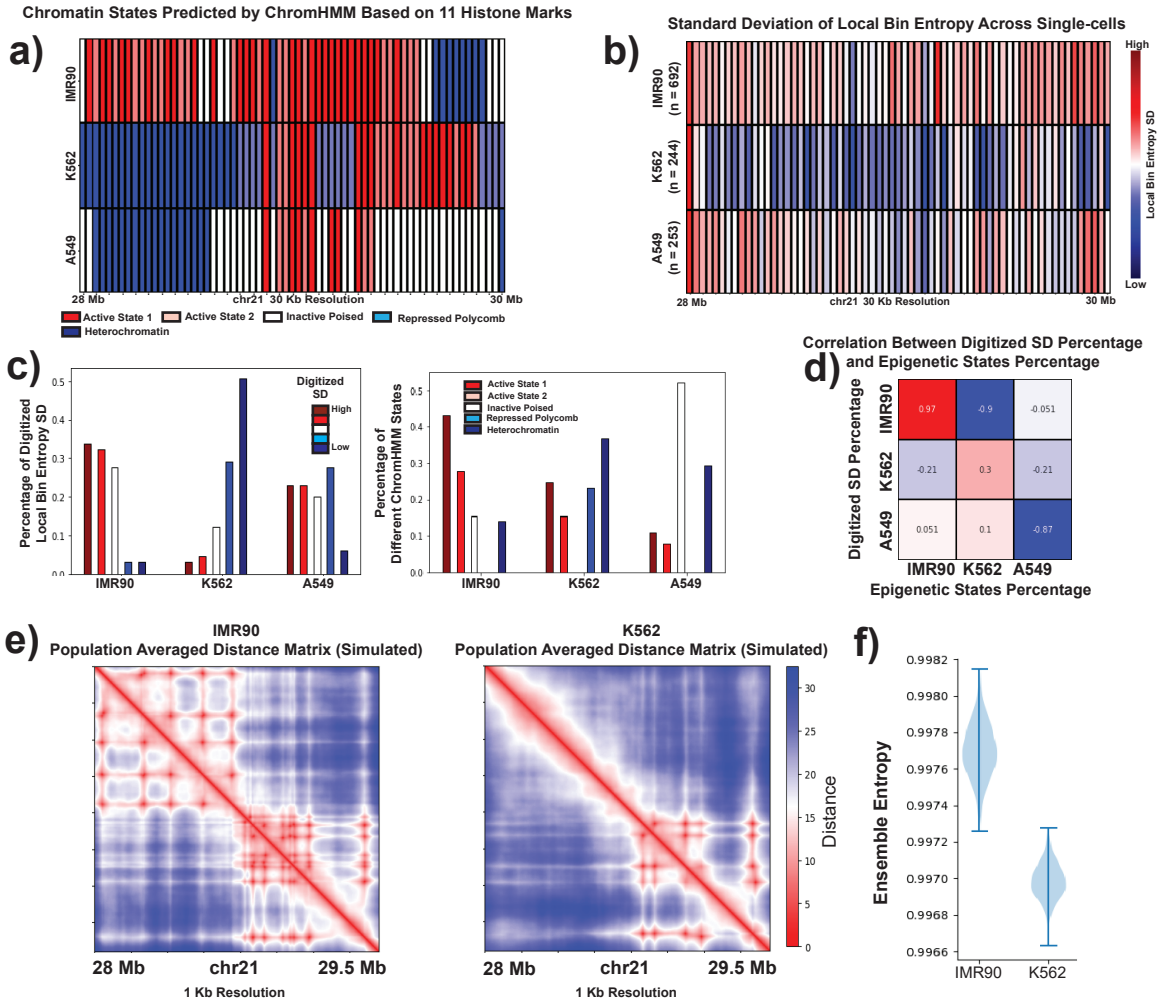
Chromatin States Predicted by ChromHMM Based on 11 Histone Marks
a)
IMR90
K562
A549
28 Mb
chr21 30 Kb Resolution
30 Mb
Active State 1
Active State 2
Inactive Poised
Repressed Polycomb
Heterochromatin
Standard Deviation of Local Bin Entropy Across Single-cells
b)
IMR90 (n = 692)
K562 (n = 244)
A549 (n = 253)
High
Low
Local Bin Entropy SD
c)
Percentage of Digitized Local Bin Entropy SD
Digitized SD
High
Low
Percentage of Different ChromHMM States
Correlation Between Digitized SD Percentage and Epigenetic States Percentage
d)
Digitized SD Percentage
0.97
-0.9
-0.051
-0.21
0.3
-0.21
0.051
0.1
-0.87
Epigenetic States Percentage
e)
IMR90
Population Averaged Distance Matrix (Simulated)
K562
Population Averaged Distance Matrix (Simulated)
Distance
28 Mb
chr21
29.5 Mb
1 Kb Resolution
f)
Ensemble Entropy
IMR90
K562

Figure 4. Epigenetic properties (histone modifications) contribute to the cell-type-specific variation in the chromosome structural ensemble. **a)** Chromatin states of chr21:28-30 Mb genomic region from three different cell-types – IMR90, K562 and A549 at 30 kb resolution. The chromatin states are inferred based on 11 different histone marks for each cell-type by ChromHMM tool. **b)** Local bin entropy standard deviation of the chr21:28-30 Mb region from IMR90, K562 and A549, where a higher value represents higher degree of structural variation across the population and vice versa. **c)** Left panel: Percentage of different digitized local bin entropy standard deviation levels for different cell-types. Right panel: Percentage of different chromatin states for different cell-types. **d)** Pairwise Spearman's correlation between percentage of digitized local bin entropy standard deviation levels and chromatin states for different cell-types. **e)** Population-averaged simulated distance maps of IMR90 and K562 chr21:28-29.5 Mb genomic region at 1 kb resolution. **f)** Chromosome structural ensemble entropy distributions for chr21:28-29.5 Mb from simulated IMR90 and K562. For resampling, 500 single-cell structures are selected randomly each time.

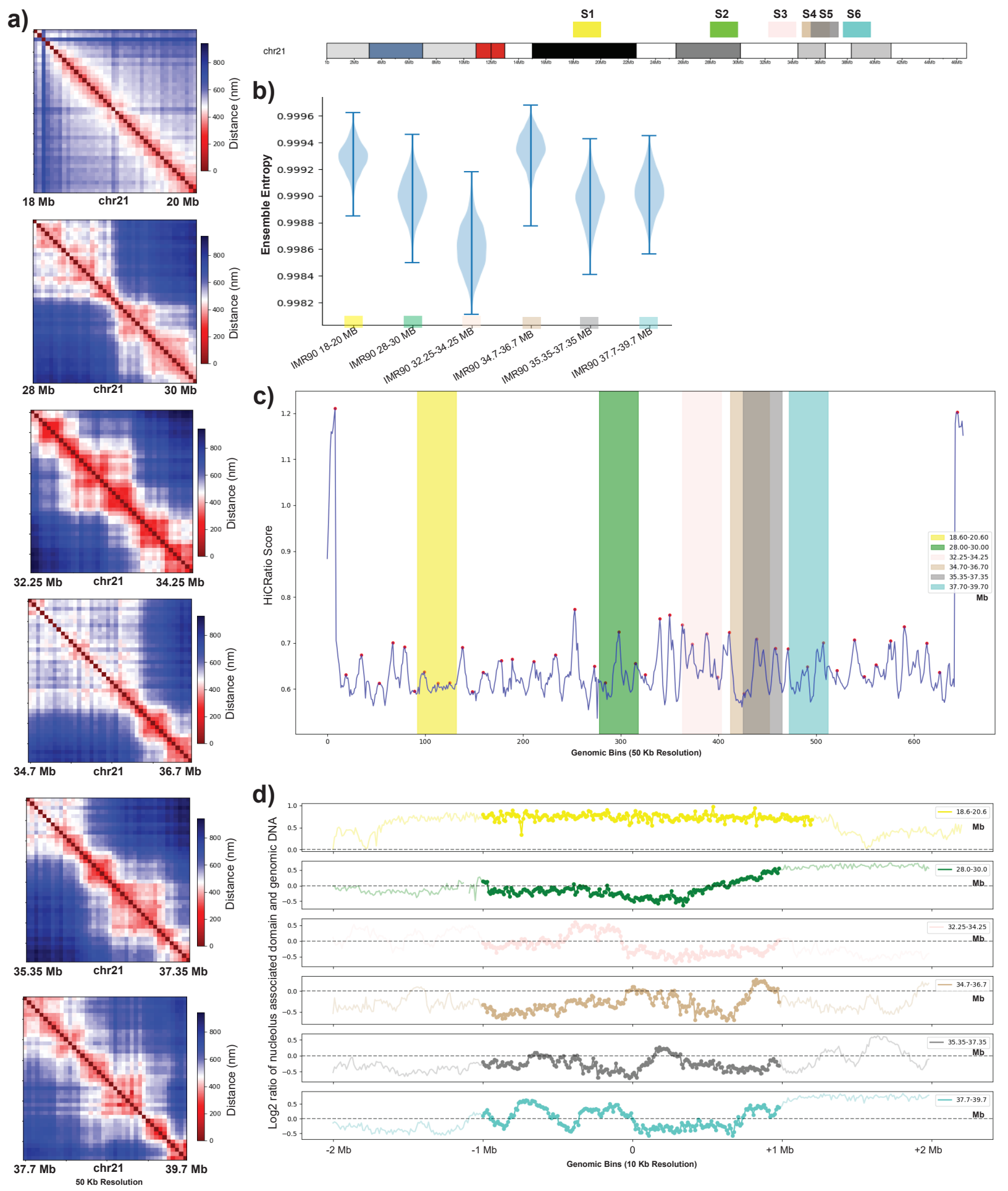
a)
S1
S2
S3
S4 S5
S6
chr21
Distance (nm)
18 Mb
chr21
20 Mb
28 Mb
30 Mb
32.25 Mb
34.25 Mb
34.7 Mb
36.7 Mb
35.35 Mb
37.35 Mb
37.7 Mb
39.7 Mb
50 Kb Resolution
b)
Ensemble Entropy
IMR90 18-20 MB
IMR90 28-30 MB
IMR90 32.25-34.25 MB
IMR90 34.7-36.7 MB
IMR90 35.35-37.35 MB
IMR90 37.7-39.7 MB
c)
HiCRatio Score
Genomic Bins (50 Kb Resolution)
18.60-20.60
28.00-30.00
32.25-34.25
34.70-36.70
35.35-37.35
37.70-39.70
Mb
d)
Log2 ratio of nucleolus associated domain and genomic DNA
18.6-20.6
28.0-30.0
32.25-34.25
34.7-36.7
35.35-37.35
37.7-39.7
-2 Mb
-1 Mb
0
+1 Mb
+2 Mb
Genomic Bins (10 Kb Resolution)

Figure 5. Evaluation of structural variation of regions with different epigenetic, genomic, and structural properties. **a)** Population-averaged distance maps of 6 different genomic regions of IMR90 chr21 at 50 kb resolution obtained from Su *et al.* (30). Locations of these 6 regions are indicated on the top right chromosome ideogram and labeled with colors used throughout the figure. **b)** Chromosome structural ensemble entropy distributions for the same 6 regions of IMR90 chr21. For resampling, 300 single-cell structures are selected randomly each time. **c)** HiCRatio score of each genomic bin of chr21 q-arm at 50 kb resolution (calculated with a 300 kb window size). The red points represent the TAD boundary positions inferred by HiCRatio. **d)** Nucleolus association level of the same 6 regions of IMR90 chr21 (49). For each genomic region, the 1 Mb upstream and downstream regions are also shown. A value greater than 0 represents higher levels of nucleolus association and vice versa.

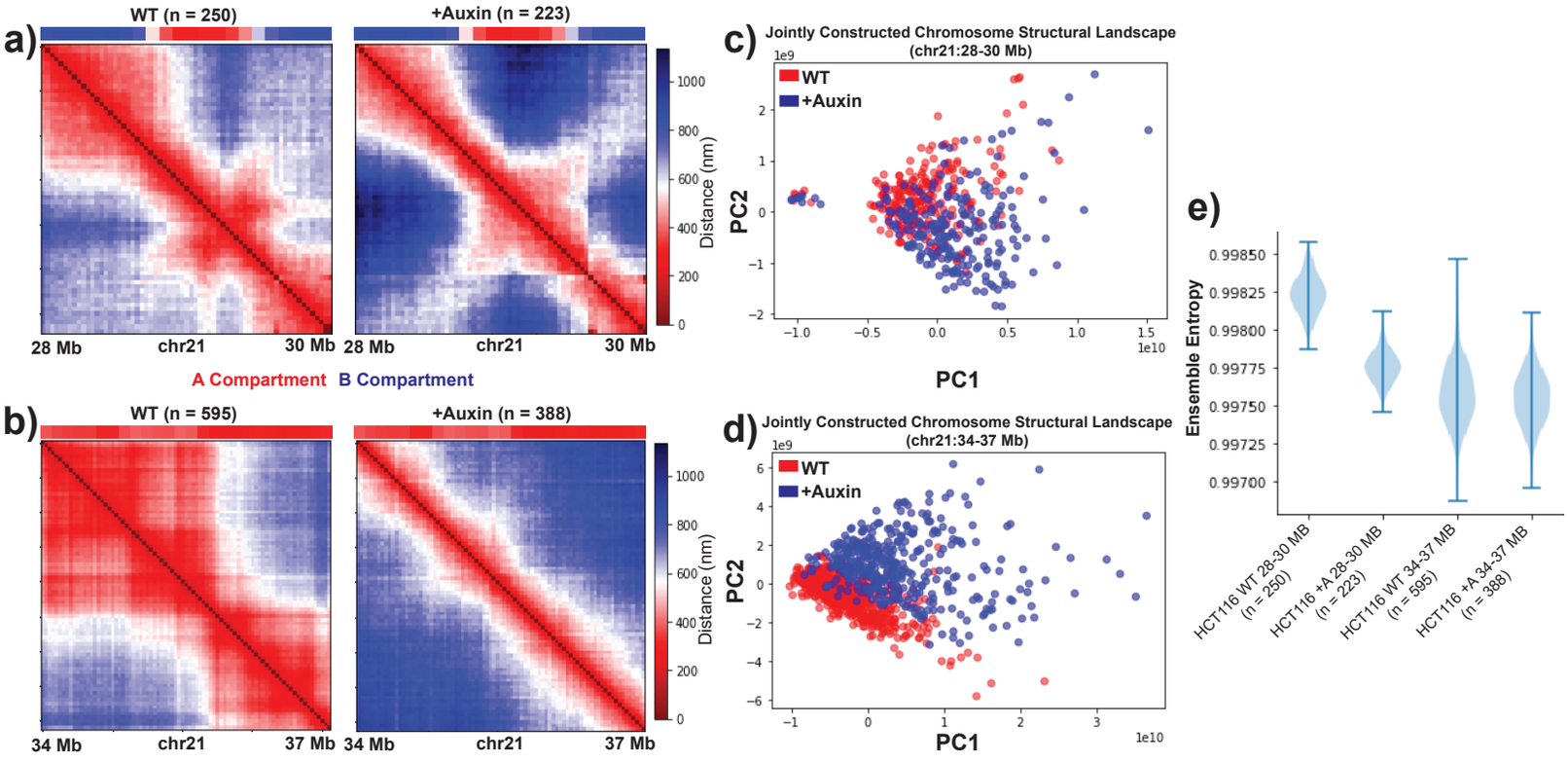
a)
WT (n = 250)
+Auxin (n = 223)
28 Mb
chr21
30 Mb
Distance (nm)
A Compartment
B Compartment
b)
WT (n = 595)
+Auxin (n = 388)
34 Mb
chr21
37 Mb
c)
Jointly Constructed Chromosome Structural Landscape
(chr21:28-30 Mb)
WT
+Auxin
PC1
PC2
d)
Jointly Constructed Chromosome Structural Landscape
(chr21:34-37 Mb)
e)
Ensemble Entropy
HCT116 WT 28-30 MB (n = 250)
HCT116 +A 28-30 MB (n = 223)
HCT116 WT 34-37 MB (n = 595)
HCT116 +A 34-37 MB (n = 388)

Figure 6. Cohesin depletion alters chromosome structural dynamics depending on other underlying constraints. **a,b)** Population-averaged distance maps of HCT116 chr21:28-30 Mb (a) or chr21:34-37 Mb (b) at 30 kb resolution for wild-type (WT) and cohesion-depleted (+Auxin) conditions based on data obtained from Bintu *et al*. (29). Compartment status from Hi-C data at these locations shown above each matrix. **c,d)** Jointly constructed chromosome structural landscapes of chr21:28-30 Mb (c) or chr21:34-37 Mb (d) of HCT116 for WT and +Auxin conditions. **e)** Chromosome structural ensemble entropy distributions for different local chromosome regions of HCT116 for WT and +Auxin conditions. 200 single-cell structures are selected randomly each time for resampling.