High resolution single-cell chromatin 3D modeling reveals coherent chromatin aggregation with varied structures in controlling genome function stability

The genome 3D architecture is thought to be related to regulating gene expression levels in cells and can be explained by genome-wide chromatin interactions which have been explored by chromosome conformation capture based techniques, especially Hi-C. Based on single-cell Hi-C data, we developed a new method in constructing experimental consistent 3D intact genome structures for individual cells with a resolution of 10kb or higher. The modeled structures showed marked variations of 3D genome organization across different cells. However, chromosome loci marked by different proteins, such as CTCF and post-translationally modified histones, are consistently non-specifically aggregated in space. Interestingly, similar aggregations between active enhancers and active promoters were observed, especially for those separated by genomic regions of the scale of megabase or larger. Such long-range associations between active enhancers and promoters are strongly correlated with spatial aggregation of chromosome loci marked by different proteins. Through analyzing the 3D structures of intact genome, we proposed that coherent gene activation profiles among individual cells can be achieved by the consistent aggregation of protein marked loci instead of maintaining identical folded conformations.

Genome architecture that are characterized by interactions between chromosome regions separated by DNA sequence of certain length is believed to play important roles in controlling and regulating biological functions in cell nucleus[1-5]. Hi-C experiments based on genome-wide DNA fragment digestion and ligation make it possible to capture these interactions and construct contact frequencies by examining DNA sequences organized in close spatial proximity[6-11]. The contact frequency matrix generated by Hi-C experiment shows hierarchical architecture of genome from conserved topological associating domains (TAD) at hundreds of kilobase scale[12-15] to spatially segregated A/B compartments at the megabase scale[6,8,12,16,17]. Alternation of this architecture, such as deletion of TAD boundaries, can lead to dysregulation of gene expressions in cells[14,18]. Starting from Hi-C contact matrix, many efforts have been made in developing computational methods that are able to construct models consistent to Hi-C experiments[6,19-28]. In these methods, single chromosome is coarse-grained as a polymer chain and ensembles of chromosome 3D structures are generated after applying physical interactions between monomers along the chain. These modeled structures of these methods greatly enhance our understanding of chromosome folding and reveal novel principles of structural organization in regulating biological functions of chromosomes. However, recent studies have shown that genomes in single-cell level are organized in varied structures, which challenges the principles revealed by populated Hi-C based models[21,29]. To investigate genome folding in individual cells, Stevens *et al.* designed new procedures of chromosome conformation capture experiments in

single-cell level and structures were calculated to provide insight into 3D organization of intact genome in single nucleus for the first time[29]. Nonetheless, due to the relative low resolution (100kb), these structures provide limited information of interactions among chromosome loci that is only available at higher resolution; these information is critical in studying the regulation of genome folding on gene expression. Therefore, we developed a new method aimed at obtaining high-resolution structures of intact genome from data obtained in single-cell Hi-C experiment. The calculated structures not only exhibit consistent features compared to both theoretical and experimental observations, but also provide evidence for principles of genome folding via non-specific aggregation of chromosome loci marked by different proteins, as well as promoting the facts that such aggregations is important in the association between active enhancers and promoters at genome-wide scale. Based on our modeled structures, we therefore propose a new mechanism of genome folding that integrate varied folded structures and consistent biological functions of individual cells.

Result

Construction of high-resolution genome architecture from single-cell Hi-C

We developed a new method to construct three-dimensional structure of intact genome using genome-wide contact data generated by Stevens' single-cell Hi-C experiments [29]. In our method, each chromosome was presented by a polymer chain with each monomer representing one chromosome locus along the sequence. By maximizing the coherence between polymer chain conformations and single-cell Hi-C data, we constructed high-resolution structures of intact genome for 8 individual mouse embryonic stem cells.

Convergent 3D structures were constructed at the genomic resolution of 10 kb for different cells (ensemble averaged rmsd of 10.5 particle radii for cell 1). These 10kb binned structures and the recalculated 100-kb structures both exhibited similar spatial organization when comparing to those in Stevens' work[29](shown in Fig.1a, averaged rmsd < 6 particle radii for 100kb resolution), which validated the reliability of both calculated structures and the strategy in constructing these structures. The contact matrices reconstructed based on modeled structures successfully reproduced the interaction patterns between genome-wide loci presented in Hi-C data (>99.8% Hi-C contact pairs reproduced, Fig.1b, Extended Data Fig.1a) and the violated loci pairs were more frequently observed to locate at sequentially close regions (Extended Data Fig.1d). Besides, loci pairs of 20 times more population were identified in modeled structures when comparing to experimental detection, suggesting the number of detected contact reads obtained by current experiment technique is far from upper limit of the possible read number in one cell nucleus. More strikingly, recalculation after removing 90% Hi-C data (with contact pair number < 10,000 after removal) still generated structures with similar contact patterns (Fig.1a, Fig.1b), suggesting the high efficiency and robustness of our method in modeling genome structures. The calculation for recombined Hi-C data of two different cells leaded to highly stretched structures with a 20-fold increase in the number of violated contact loci pairs (11.6% pairs violating experimental derived constraints, Extended Data Fig.1d) and larger degrees in violating the distance restraints presented in merged contact matrix (maximum distance of 114 particle radii for violated pairs, compared to 7.7 particle radii for separated data). In addition, we constructed 1kb-resolution conformations for each cell satisfying constraints of Hi-C contacts. Note that these

high-resolution structures were not used in further analysis due to the relative large conformational variation among replicated structures derived from limited amount of contact reads (rmsd of 13.3 particle radii). Taking together, the high consistency between calculated structures and experimental data strongly validated the reliability and robustness of our method in utilizing single-cell Hi-C data to reproduce the folded structure of intact genome at high resolution level.

High resolution 3D genome organization under high resolution

The modeled chromatin 3D structures demonstrated similar folding conformations displayed by those presented in Stevens' work[29](Fig.1a, Fig.1c). For example, *Rabl* configuration, and discrete chromosome territories were observed in all cells. Constitutive lamina associated domains (LADs) are consistently organized at nuclear membranes or nucleolar periphery, while early replicated regions and CpG enriched loci are located between them. Besides, directed mapping of ChIP-seq data on 3D structures shows that chromosome segments marked by CTCF, H3K27me3, H3K27ac, H3K36me3, H3K4me1, H3K4me3, active enhancers and active genes are coherently enriched in the spatial regions with distance of 0.3-0.7 nuclear radius to the center(shown in Fig.2a, Extended Data Fig.3c). This suggests that euchromatin prefer to locate away from nucleolus and membranes where heterochromatin prefers to reside. In the calculated structures, each chromosome in cell nucleus is organized in a fractal globule manner and domains of varied genomic size can be observed in separation index matrix plot. These domains exhibit hierarchical structures of chromosome organization from hundreds of kilobases to tens of megabases (Fig.2b). Considering the calculations are performed based on Hi-C contact matrices with varied interaction patterns among genome-wide loci, such consistent distribution of chromosome segments among individual cells suggests that both experimental data and modeled structures successfully capture the coherent features of genome organization.

Varied 3D genome structures in individual cells

Consistent with substantially varied interaction patterns in Hi-C contact maps, chromosomes adopt alternative structures in organizing their consecutive segments in different cells, resulting in significant variation in 3D folding at different genomic scales. Such large structural variation is not due to the lack of contact constraints, since similar conformations are observed for structures calculated from the same Hi-C data. One example of such variation is that *chr17* adopts compact structure with polarized organization for A/B compartment in cell 3, while exhibits snake-like structure in cell 1 in a narrow but long territory confined by other chromosomes (shown in Fig.2c, Extended data Fig.2c). The varied organization of chromosome segments results in substantially different distance between intra-chromosome loci, which can be visualized in the plot of separation index matrices (shown in Fig.2c). These matrices show that regions of the same compartment type are not always aggregated in single chromosome, indicating the principles of genome folding derived from populated Hi-C may not always follow in single-cell level. These findings suggest that one chromosome can adopt different structure in the genome folding process, which to some extend contradict with the well accepted understanding about the determination of folded genome structure on the biological functions. Therefore, it is urgent to integrate these two scenes into a consistent picture that can unravel the underlying regulation mechanism of chromosome spatial organization in gene expression.

Aggregation of chromosome regions with different features

As mentioned above, chromosome loci marked by different proteins, such as CTCF and H3K27me3, tend to spatially co-enriched with compartment A (Fig.1c, Extended Data Fig.3c), and show consistent aggregation in different cells (Fig.3a). Further analyses indicate that such aggregations of these loci are positively correlated with the increased genomic distance, regardless whether the marked proteins are same or not (Fig.3d, Extended Data Fig.4). This suggests the significance of proteins in regulating 3D genome structure. Strikingly, clustering analysis on loci marked by same protein shows that the aggregation are relatively randomly formed in different cells, without specific recognition on the genomic locations of loci in close spatial proximity (Fig.3b). Interestingly, H3K9me3 marked loci show relative weak aggregation with themselves but significant strong aggregation with those marked by other proteins (such as H3K27me3, H3K36me1, H3K4me1 and H3K4me3) (Fig.3a, Fig.3d). For loci marked by different proteins, coherent spatial aggregations are observed in different cells; it not only suggests the wide distribution of such aggregation in genome folding, but also further validates the reliability of our calculated structures. These consistent organizations suggest that the non-specific aggregation of protein marked loci may play important roles in maintaining coherent biological functions of individual cells of the same type.

In addition to the spatial aggregation of loci marked by different proteins, the average protein densities smoothed over loci in close spatial proximity show strong correlation with the average densities smoothed over genomic sequence (Extended Data Fig.5a). Similar results are obtained after excluding loci within 1Mb regions when selecting spatially close loci (Fig.3c), proving that such correlation is not caused by the coherence between spatial proximity and sequential proximity. Considering protein density is obtained based on ChIP-seq data of populated cells, such strong correlation suggests this property is robust among individual cells. Also, significant strong correlations of spatially averaged protein densities are observed between different cells (for CTCF average correlation value is 0.31, compared to 0.19 for random structures) (Fig.3c), suggesting that each locus is surrounded by chromosomes of similar biological features. The varied 3D organization of genome and consistent local biological environment of chromosome loci suggests that it is the biological features of the surrounded chromosome regions, instead of the exact conformation of folded chromosome, that maintain the similar biological functions in same type of cells. We therefore hypothesis that it is the non-specific aggregations between loci marked by different proteins may contribute to achieving such consistent biological environment.

Association between active enhancers and active promoters with protein marked loci

The spatial association between enhancers and promoters, which has been validated by 3C experiments[30,31], is thought to regulate the gene expression levels. In our modeled structures, an average of 2.57 active enhancers are located in close spatial proximity (within cutoff distance of 4 particle radii) from one active promoter, whereas only 0.04 contact pairs on average can be detected by current single cell Hi-C experiments. The spatial aggregation between active promoters and active enhancers is supported by experimentally observed structures such as transcription factories[32,33], and active chromatin hub[3,34,35], as well as recently proposed phase separation model[36-38], which independently validate our observation that is from the

computational perspective. Different from experimentally identified proximate pairs between active enhancers and promoters, which are mostly located within genomic distance of 1Mb, the spatial association in calculated structures can be formed with genomic distance up to hundreds of megabases (Fig.4a). Compared to random structures, the number of spatially associated active enhancer-promoter pairs with genomic distance greater than 1Mb (including inter-chromosome) has increased by 110%, suggesting remarkable promotion of such associations in genome folding (Fig.4a). For calculated cells, the associated pairs with genomic distance larger than 1Mb cover 43.9% active genes, suggesting long range enhancer-promoter association can also play an important role in regulating gene expression in cells. Moreover, the calculated structures show that the associated enhancer-promoter pairs vary substantially from cell to cell (Fig.4b) (an average of 0.05% common enhancer-promoter pairs with genomic distance larger than 1Mb are observed in two cells), which agrees with the features of super-enhancers[39,40] that can regulate the expression of different genes in populated cells. The high percentages of active promoters that are non-specifically associated with active enhancers indicate varied folded genome structure can achieve similar gene activation profiles in individual cells.

To explore the influence of aggregation of protein marked loci on the association between active enhancers and promoters, we defined a promotion factor that quantifies the improvement of enhancer-promoter association under the condition that two loci must be marked by the analyzed proteins. The result shows that the spatial aggregation between loci marked by proteins such as CTCF, H3K27ac, H3K27me3, H3K4me1 and H3K4me3 significantly promotes the active enhancer-promoter association and the promotion effect grows with the increase of genomic distance (Fig.4d, Fig.4e). Interestingly, H3K9me3 inhibits these associations in genomic distance 1Mb to 10Mb, while exhibit promotion effect for distance greater than 10Mb. The varied promotion patterns exhibit different roles of protein in regulating enhancer-promoter association. Similarly, aggregation of loci marked by different proteins show inhibition effect for distance in 100kb-1Mb region while promote active enhancer-promoter association for distance larger than 10Mb with different degrees (Fig.4e). These results suggest that the association between active enhancers and promoters are correlated with the aggregation network among loci marked by different proteins.

Conclusion

In this work, we developed a new method to model experimental consistent high-resolution (10kb resolution) three-dimensional structures for intact genome based on single-cell Hi-C data of 100-kb resolution. The generated chromatin 3D structure results allow us to explore the relationship between chromosome folding and gene regulation at the single nucleus level. The modeling results demonstrated both cell-specific folding conformations of the genome and consistent organization of chromosome regions labeled by different features. Based on these structures, we discovered that chromosome loci marked by different proteins, including CTCF and modified histones are observed to randomly aggregate in nuclear space to ensure each locus is spatially surrounded by loci of similar biological features in individual cells. Moreover, such

random aggregation of these loci is also correlated with regulating the association between active promoters and active enhancers at different genomic scales, suggesting that consistent spatial aggregation between chromosome loci marked by proteins in varied 3D genome structures plays an important role in maintaining similar gene expression levels of individual cells of the same type. Through this work, we would like to point out that theoretical analysis on 3D intact genome structures modeled from single-cell Hi-C data can effectively facilitate our investigation and understanding of chromosome organization, as well as the impact on biological functions of nucleus in single-cell level.
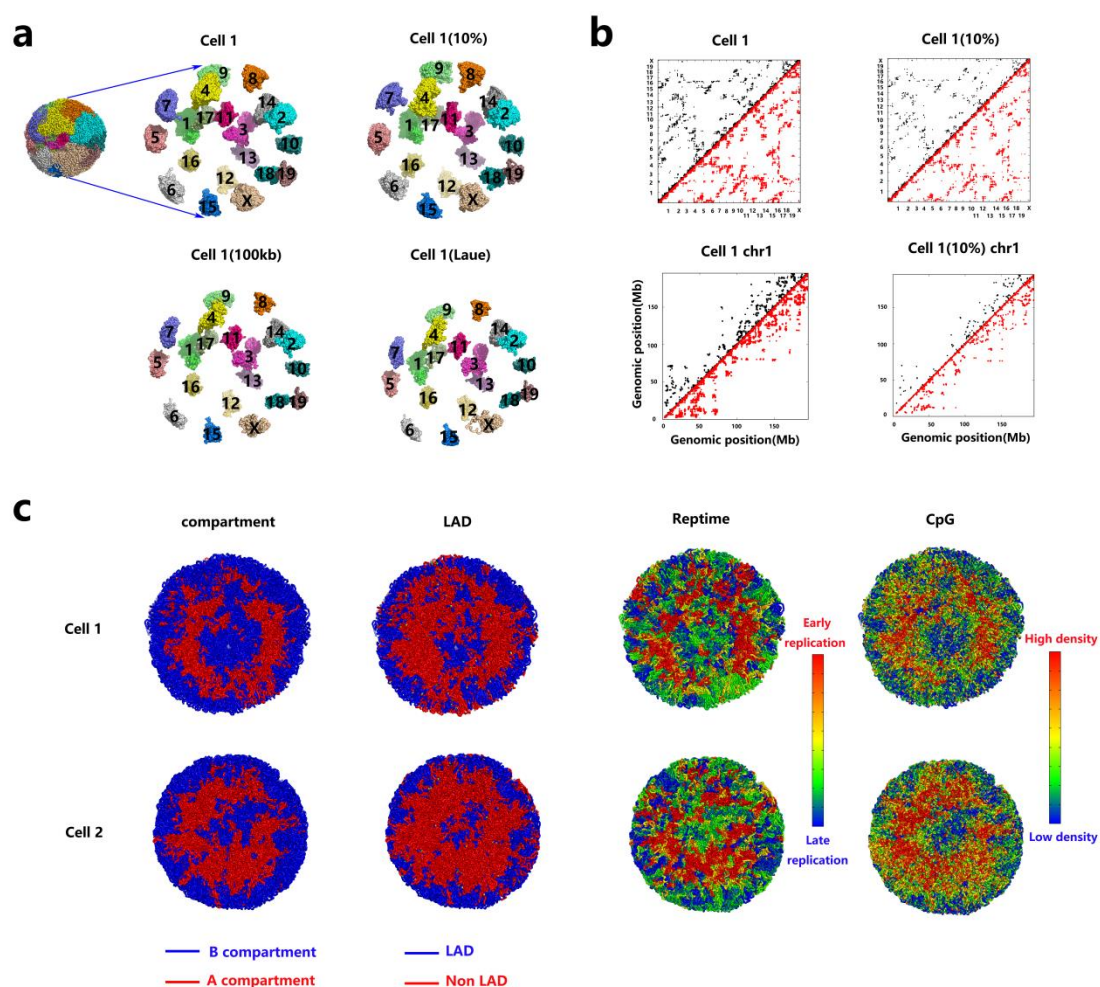
Figure 1. Validation of calculated structures. a. Expanded view of spatial territories for chromosomes in calculated structures of cell 1 constructed with different methods: 10kb resolution with all Hi-C data(upper left), 10kb resolution with 10% Hi-C data(upper right), 100kb resolution with all data(lower left) and 100kb resolution structures presented in Stevens'[29] work(lower right). **b**. Comparison between experimental Hi-C contacts (upper triangle, in black) and reconstructed contacts (lower triangle, in red) among genome-wide loci for structures constructed from different datasets: cell 1 with all Hi-C data(upper left) and only 10% data

separately(upper right). Same with two above matrices with enlarged view for chromosome 1. **c**. Cross view of 3D structures of intact genome of cell 1 and cell 2 with chromosome loci colored based on the corresponding features: loci belonging to compartment A(red) or B(blue); loci that are identified as LAD (blue) or not (red); the replication timing of each locus calculated based on ChIP-seq data; and CpG density of each locus.
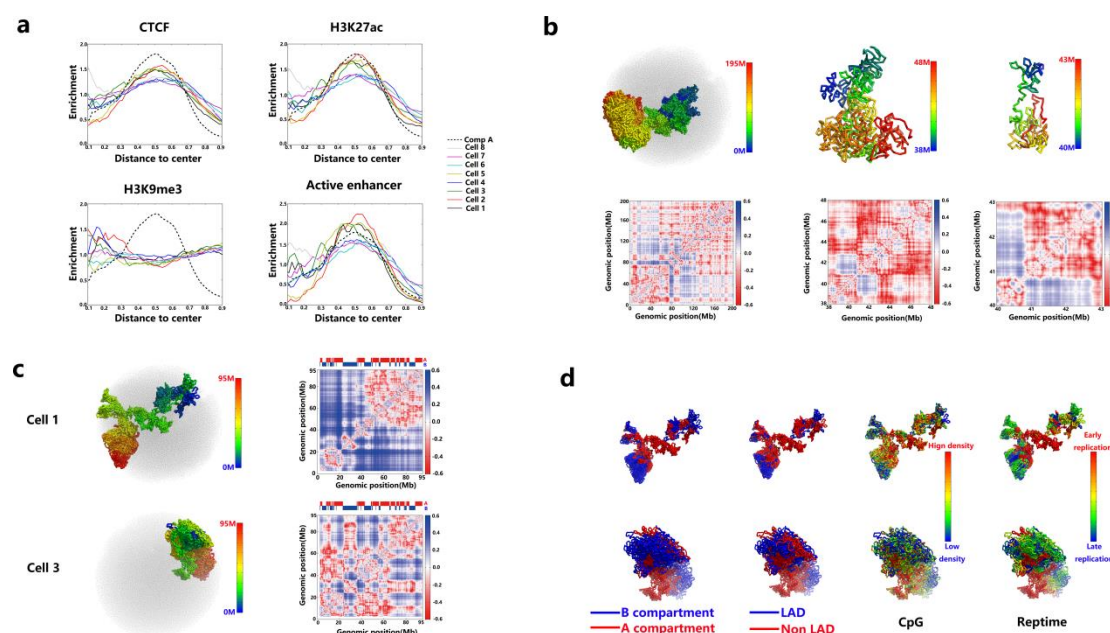
Figure 2.Consistent 3D organization of chromosome with cell dependent structures. a. Plot of enrichment (calculation presented in Supporting Information) of chromosome loci marked by CTCF (upper left), H3K27ac (upper right), H3K9me3 (lower left) and active enhancers (lower right)with respect to scaled radial distance to nuclear center in different cells. As comparison, enrichment of compartment A loci is also presented in dashed lines in each plot. **b**. Plot of 3D structures and the corresponding separation index matrices (bottom) for different chromosome regions in cell 1: intact chromosome 1 (left), 38M-48M in chromosome 1 (middle) and 40M-43M in chromosome 1(right). In 3D structure plot, chromosome loci are colored based on the genomic positions along the sequence.   **c**. Same with B for chromosome 17 in calculated structures for cell 1 and cell 3, respectively. **d.** 3D projection of different chromosome features on varied structures of chromosome 17 from cell 1 and cell 3. The coloring of chromosome loci is performed with the same format shown in in Fig.1c.

Figure 3. Non-specific aggregation of chromosome loci marked by proteins. a. Violin plot of the distribution of local density of chromosome loci marked by different proteins in cell 1. Relative high local density in modeled structure indicates these loci are aggregated in folded genome. **b**. Mutual information between spatial clusters formed by spatially aggregated loci marked by the same protein in different cells. As comparison, similar calculation is performed for random structures. Similar mutual information values suggest spatial clusters of aggregated proteins are randomly formed in modeled structures. **c**. Correlation between spatially averaged protein density (loci within 1Mb excluded in calculation) and sequentially averaged protein density (smoothed over 100kb) (left); Correlation between spatial averaged protein density (loci within 1Mb excluded in calculation) from different cells (right) **d**. Heatmap plot of aggregation degree matrix for chromosome loci marked by different proteins with respect to varied genomic distances. All calculations are performed based on modeled structures of all cells except panel (a).

Figure 4. Spatial aggregation between active enhancers and promoters correlate with

aggregation of chromosome loci marked by proteins. **a**. Comparison between aggregated active promoter-enhancer pair number for calculated structures and random structures with respect to different genomic distance. Calculation was performed on all cells. **b**. Heatmap plot of overlapping ratio between aggregated active promoter-enhancer pairs from different structures: structure replicas of the same cell (top row), random structures for different cells(median row) and modeled structures from different cells (bottom row). **c**. Box plot of aggregation degree between active promoters and active enhancers based on calculated structures for 8 cells with respect to different genomic distances. **d.** Heatmap plot of promotion factor of active enhancer-promoter association with respect to spatial aggregation of loci marked by the same proteins. **e**. Plot of promotion factor for chromosome loci marked by different proteins with respect to different genomic distances.

Reference

1   Bickmore, W. A. & van Steensel, B. Genome Architecture: Domain Organization of Interphase Chromosomes. *Cell* **152**, 1270-1284, doi:10.1016/j.cell.2013.02.001 (2013).

2   Sexton, T. & Cavalli, G. The Role of Chromosome Domains in Shaping the Functional Genome. *Cell* **160**, 1049-1059, doi:10.1016/j.cell.2015.02.040 (2015).

3   Pombo, A. & Dillon, N. Three-dimensional genome architecture: players and mechanisms (vol 16, pg 245, 2015). *Nat Rev Mol Cell Bio* **16** (2015).

4   Therizols, P. *et al.* Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science* **346**, 1238-1242, doi:10.1126/science.1259587 (2014).

5   Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116-120, doi:10.1038/nature11243 (2012).

6   Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289-293, doi:10.1126/science.1181369 (2009).

7   Jin, F. L. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290-294 (2013).

8   Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping (vol 159, pg 1665, 2014). *Cell* **162**, 687-688, doi:10.1016/j.cell.2015.07.024 (2015).

9   Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306-1311, doi:DOI 10.1126/science.1067799 (2002).

10   Ma, W. X. *et al.* Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. *Nat Methods* **12**, 71-78, doi:10.1038/Nmeth.3205 (2015).

11   Martin, P. *et al.* Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. *Nat Commun* **6**, doi:ARTN 1006910.1038/ncomms10069 (2015).

12   Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).

13   Sexton, T. *et al.* Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell* **148**, 458-472, doi:10.1016/j.cell.2012.01.010 (2012).

14   Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385, doi:10.1038/nature11049 (2012).

15   Le, T. B. K., Imakaev, M. V., Mirny, L. A. & Laub, M. T. High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome. *Science* **342**, 731-734, doi:10.1126/science.1242059 (2013).

16   Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat Rev Genet* **17**, 661-678, doi:10.1038/nrg.2016.112 (2016).

17   Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38**, 1348-1354, doi:10.1038/ng1896 (2006).

18   Lupianez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161**, 1012-1025, doi:10.1016/j.cell.2015.04.004 (2015).

19   W.J., X. *et al.* Structural Modeling of Chromatin Integrates Genome Features and Reveals Chromosome Folding principle. *Sci Rep-Uk* **7** (2017).

20   Lesne, A., Riposo, J., Roger, P., Cournac, A. & Mozziconacci, J. 3D genome reconstruction from chromosomal contacts. *Nat Methods* **11**, 1141-1143, doi:10.1038/Nmeth.3104 (2014).

21   Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59-+, doi:10.1038/nature12593 (2013).

22   Nicodemi, M. & Prisco, A. Thermodynamic Pathways to Genome Spatial Organization in the Cell Nucleus. *Biophys J* **96**, 2168-2177, doi:10.1016/j.bpj.2008.12.3919 (2009).

23   Chiariello, A. M., Annunziatella, C., Bianco, S., Esposito, A. & Nicodemi, M. Polymer physics of chromosome large-scale 3D organisation. *Sci Rep-Uk* **6**, doi:ARTN 2977510.1038/srep29775 (2016).

24   Zhang, B. & Wolynes, P. G. Topology, structures, and energy landscapes of human chromosomes. *P Natl Acad Sci USA* **112**, 6062-6067, doi:10.1073/pnas.1506257112 (2015).

25   Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* **15**, 2038-2049, doi:10.1016/j.celrep.2016.04.085 (2016).

26   Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *P Natl Acad Sci USA* **112**, E6456-E6465, doi:10.1073/pnas.1518552112 (2015).

27   Naumova, N. *et al.* Organization of the Mitotic Chromosome. *Science* **342**, 948-953, doi:10.1126/science.1236083 (2013).

28   Tjong, H. *et al.* Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. *P Natl Acad Sci USA* **113**, E1663-E1672, doi:10.1073/pnas.1512577113 (2016).

29   Stevens, T. J. *et al.* 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59-+, doi:10.1038/nature21429 (2017).

30   Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *P Natl Acad Sci USA* **111**, 996-1001, doi:10.1073/pnas.1317788111 (2014).

31   Phillips-Cremins, J. E. *et al.* Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell* **153**, 1281-1295, doi:10.1016/j.cell.2013.04.053 (2013).

32   Sutherland, H. & Bickmore, W. A. Transcription factories: gene expression in unions? *Nat Rev Genet* **10**, 457-466, doi:10.1038/nrg2592 (2009).

33   Schoenfelder, S. *et al.* Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet* **42**, 53-U71, doi:10.1038/ng.496 (2010).

34   Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* **10**, 1453-1465, doi:Doi 10.1016/S1097-2765(02)00781-5 (2002).

35   de Laat, W. & Grosveld, F. Spatial organization of gene expression: the active chromatin hub. *Chromosome Res* **11**, 447-459, doi:Doi 10.1023/A:1024922626726 (2003).

36   Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13-23, doi:10.1016/j.cell.2017.02.007 (2017).

37   Larson, A. G. *et al.* Liquid droplet formation by HP1 alpha suggests a role for phase separation in heterochromatin. *Nature* **547**, 236-+, doi:10.1038/nature22822 (2017).

38   Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547**, 241-+, doi:10.1038/nature22989 (2017).

39   Pott, S. & Lieb, J. D. What are super-enhancers? *Nat Genet* **47**, 8-12, doi:10.1038/ng.3167 (2015).

40   Hnisz, D. *et al.* Super-Enhancers in the Control of Cell Identity and Disease. *Cell* **155**, 934-947, doi:10.1016/j.cell.2013.09.053 (2013).