Folding Principle of Chromosome Emerges from Mapping of Genome Features onto its 3D Structure

Wen Jun Xie[1,2], Luming Meng[1,2], Sirui Liu[1,2], Ling Zhang[1], Xiaoxia Cai[1], Yi Qin Gao[1,*]

[1] Beijing National Laboratory for Molecular Sciences, College of Chemistry and Molecular Engineering, and Biodynamic Optical Imaging Center, Peking University, Beijing 100871, China

[2] Co-first author

* Correspondence: gaoyq@pku.edu.cn (Y.Q.G)

ABSTRACT

How chromosomes fold into 3D structures and how genome functions are affected or even controlled by their spatial organization remain challenging questions. Hi-C data are used here to construct the 3D chromatin structure and then shown to be closely connected with a plethora of genetic and epigenetic features and functions. The chromatin structures are characterized by two spatially segregated compartments, which are further dissected into two types of domains with clearly different intra contact patterns and most importantly, the size of chromatin loops. The chromatin loops segregate in the space according to their sizes, reflecting the entropic effect in chromatin structure formation as well as chromosome positioning. Taken together, these results provide clues to the folding and principles of the spatial organization of chromatin fiber, and the resulted clustering of many genome features in the 3D space.

Keywords

Hi-C, Structural Modeling, Chromosome Folding Principle, Chromatin Structure, Entropy-driven, Chromatin Compartmentalization, Chromosome Positioning, Genome Feature

INTRODUCTION

Eukaryotic genomes in the interphase are tightly packaged within cell nuclei and fold into compact 3D structures, and form chromosome territories (Cremer and Cremer, 2010; Fraser et al., 2015). The spatial organization of chromatin plays an essential role in various genome functions (Ali et al., 2016; Bickmore and van Steensel, 2013; Dekker and Mirny, 2016; Misteli, 2007; Pirrotta, 2016). However, our understanding of the folding and 3D structure of chromatin is still far from complete.

Recently, chromosome-conformation-capture (3C) based methods especially Hi-C have provided profound insights including that chromatin fiber is spatially organized into two tissue-specific compartments (A/B compartments) in the megabase scale or tissue-invariant topologically associating domains (TADs) in the kilobase scale (Dixon et al., 2012; Lieberman-Aiden et al., 2009). Identification of chromatin loops also helps to explain the role of distal genomic interactions in controlling gene expression (Rao et al., 2014). Fluorescence in situ hybridization (FISH) which can directly image the 3D chromatin also confirmed the two compartments in single chromosomes (Wang et al., 2016).

Besides experiments, modeling chromatin structure is also indispensable in our understanding of genome properties and is expected to play increasingly important roles (Dekker et al., 2013; Serra et al., 2015). In one seminal study, Lieberman-Aiden and coworkers demonstrated that the fractal instead of the

equilibrium globule is consistent with the Hi-C data at the megabase scale (Lieberman-Aiden et al., 2009), whereas, the loop-extrusion model was shown to explain better the Hi-C data in their later study (Fudenberg et al., 2016; Sanborn et al., 2015). Giorgetti et al. modeled the X inactivation center in human X chromosome and proposed that the fluctuations in chromosome conformation are coupled with transcription (Giorgetti et al., 2014). Zhang and Wolynes simulated the chromosome topology which was found to be largely free of knots (Zhang and Wolynes, 2015).

The relation between chromatin organization and genome properties is also emerging (Giorgetti et al., 2016; Pope et al., 2014; Valton and Dekker, 2016; Wang et al., 2016). But a comprehensive analysis is still lacking and there are still many gaps in the way to a comprehensive understanding of chromosomes in the highly crowding nucleus. For example, the mechanism of chromatin compartmentalization and chromosome positioning remains unknown.

Here, we further explore the spatial organization of human chromatin by physical modeling utilizing experimental Hi-C data. The modeled 3D chromatin structure provides an integrated and systematic view of a large variety of genome features and reveals the chromosome folding principle. We map a large number of genome features onto the modeled structure of an individual chromosome, revealing remarkable relations between them. The comparison of chromatin structures between human embryonic stem cell and differentiated

cells also provides information on cellular differentiation. In exploring the spatial segregation of genome features, we found that an individual chromosome can be divided into two types of domains which form the basis of higher-level chromatin compartments. The two domains have different densities of GC-content and CCCTC-binding factor (CTCF) binding site, and sizes of chromatin loops or TADs. Thus, chromatin compartmentalization can be viewed as a result of the segregation of chromatin loops or TADs according to their sizes, highlighting the importance of entropy in driving chromatin compartmentalization. We also found chromosomes position in the nucleus following this entropy-driven mechanism.

RESULTS

Structural Modeling of Human Chromosome

In Hi-C experiments, the relative frequencies of spatial contacts between genomic loci are measured. Many efforts have been made to detect specific loci interactions and topological domains in chromatin from genome-wide Hi-C data (Dekker et al., 2013). Building 3D structure of chromatin is also essential to understanding the mechanism of chromatin folding (Serra et al., 2015). Because Hi-C data represent average contacts from a cell population and homologs, an ensemble of chromosome conformations should be modeled so that the statistical properties of which are consistent with the experimental data.

Here we developed a restraint-based modeling strategy to build chromatin structure. We first coarse-grained the genome structure with a bead-spring model and each bead represents a 50-kb genomic region. The contact frequencies between loci in experimental Hi-C data were correlated with their spatial distances and thus we converted the Hi-C map into a set of restraints, which were then satisfied by optimization using molecular dynamics simulations. We obtained a population of models from optimizing random initial chromosome configurations (Methods). The experimental Hi-C data for IMR90 used here are obtained from Rao et al. (Rao et al., 2014).

The comparison between contact frequencies obtained in experiments and our modeling for chromosome 1 in IMR90 is shown in Fig. 1A. The performance of the model is demonstrated at a finer scale by zooming-in the central region into Fig. 1B. The coincidence between the experimental and calculated results clearly demonstrates the accuracy that our modeling strategy has achieved. Throughout the paper, we take chromosome 1 of IMR90 as an example, unless otherwise stated.

We next analyzed the features of the chromosome structure ensemble. Clustering based on the RMSD between any two conformations demonstrates that there isn't a dominant structure in the models (SI; Fig. S2). All the constructed models have a packed configuration. The cluster centroid is chosen as representative which is shown in Fig. 1C. This packed conformation also supports the concept of chromosome territories that each chromosome

organizes itself into a given volume in the nucleus and only contacts with its immediate neighbors (Cremer and Cremer, 2010).

We also found that our models are largely devoid of knots (SI; Fig. S4) which has been argued to be important in chromatin folding and unfolding (Lieberman-Aiden et al., 2009; Zhang and Wolynes, 2015). Our 3D structure also successfully reproduces the chromatin loop identified in Hi-C data (SI) (Rao et al., 2014). In addition, the chromatin show interesting structural features at a finer scale which can be viewed as a collection of strands (Fig. 1D). Although it has not been explicitly mentioned before, we can also find the strands from previously modeled conformation (Lesne et al., 2014). All these results strengthen the physical meaning of our modeling.

Spatial Segregation of Compartments A and B

One can normalize the Hi-C contact matrix by the observed-expected method and then obtain a corresponding correlation matrix (Fig. 2A) (Lieberman-Aiden et al., 2009). The plaid pattern in Fig. 2A suggests spatial segregation between two chromatin compartments.

Here we used spectral clustering to decompose human chromosome into A/B compartments (Fig. 2A; SI) and mapped their genomic regions on our model (Fig. 2B). It is obvious that the two compartments spatially segregate from each other despite that their sequences alternate along an individual chromosome (Fig. 2B). We also estimated the local density or compactness

around each genomic region in our models by calculating the number of segments within a certain surrounding volume (SI). The spatial density obtained from our modeling in compartment A is smaller than compartment B (SI), implying the euchromatin and heterochromatin nature of these two compartments, respectively.

For each genomic segment, the degree of compartmentalization is quantified as the logarithm ratio of normalized contacts with compartment A to normalized contacts with compartment B (SI) (Dileep et al., 2015). The high positive degree values indicate that the genomic regions locate in the interior of compartment A whereas the low negative values correspond to the inner part of compartment B (Fig. 2C). Spatially, the interior of compartment A (red color, large positive degree values) and B (blue/green color, large negative degree values) are separated by compartment boundaries (yellow color, with degree values around zero) (Fig. 2C). The degree of compartmentalization of one genomic segment roughly correlates with its local density (Fig. S5). Thus in the following we can reveal the biological significance of chromatin structure by comparing genome features with degree of compartmentalization.

Spatial Segregation of Genome Features

Chromatin compartments have been demonstrated to relate with a variety of genome features, including GC-content, replication timing, DNase I hypersensitivity and histone marks (Dekker et al., 2013; Imakaev et al., 2012;

Lieberman-Aiden et al., 2009). It is thus expected that these genome features also segregate given the spatial segregation of the two compartments.

In the following, one can find that it is instructive to map various genome features on modeled chromatin structure. The nearly quantitative consistency between various biological data and the modeled chromatin structure not only reveals their relation but also validates our modeling. The data sources for genome features analyzed here are summarized in Table S1.

Lamina-associated Domain

In the interphase, the nuclear lamina coating the inner nuclear membrane interacts with specific genomic regions named lamina-associated domains (LADs). The DNA Adenine Methyltransferase Identification (DamID) technique is used to identify LADs in a genome-wide manner (van Steensel and Henikoff, 2000). LADs have long been thought to play an important role in maintaining chromatin structure and regulating gene expression (Kind and van Steensel, 2010).

We find here that LADs can be aligned remarkably well with compartment B in the 3D space (Fig. 3A), supporting the notion that the chromatin in LADs is more compact. Given that LADs locate in the peripheral of the nucleus, our model shows that the compactness of chromatin decreases from the peripheral to central nucleus.

DNase I Hypersensitivity, RNA polymerase II and Gene Expression

By mapping the DNase I hypersensitivity data onto chromatin 3D structure

(Fig. 3A), we also see that chromatin in compartment A is more accessible, which might cause the enrichment of RNA polymerase II (RNAP II) in this compartment (Fig. 3A). These results are in line with the transcriptional activation in compartment A (Fig. 3A) and highlight the importance of chromatin structure as an emerging regulator of gene expression. Such analyses are consistent with previous analyses in that most genes in compartment B or LADs are transcriptionally inactive.

Histone Modifications

We next examined the relation between chromatin structure and various histone modifications (H3K4me3, H3K27ac, H3K36me3, H3K9me3 and H3K27me3). H3K4me3 which involves in the activation of gene expression is enriched in compartment A (Fig. 3A), which is also the case for the other two active histone marks (H3K27ac, H3K36me3) (Fig. S6). Meanwhile, repressive histone marks (H3K9me3, H3K27me3) show less obvious compartmentalization (Fig. S6), revealing a major difference between repressive and active histone marks.

Replication Timing

DNA replication occurs in a cell-type-specific temporal order known as the replication-timing (RT) program. RT reflects chromatin spatial organization in that early and late replication domains (RD) match well with compartments A and B, respectively (Rivera-Mulia and Gilbert, 2016). It is not surprising that the early and late RD would spatially separate from each other when mapped

on our chromatin model.

Strikingly, the degree of compartmentalization also coincides very well with RT as can be seen from 3D structure (Fig. 3A) or one-dimensional profile (Fig. 3B). Thus, it can be concluded from our modeling that replication timing has a close relationship with local compactness. This modeling study thus also presents a vivid illustration of the relation between chromosome replication and structure.

Spatial Segregation of DNA Methylation and Modeling of Partially Methylated Domain

DNA methylation, as the most important DNA modification, again shows spatial segregation (Fig. 4A). The widespread hypomethylation in cancer methylome resembling the partially methylated domains (PMDs) also occurs in IMR90 cell line (Berman et al., 2012). PMDs are spatially segregated from non-PMDs (genomic regions that aren't PMD) (Fig. 4B) and nearly all of them (98.8% in length) locate in compartment B, suggesting the possible pathological change of this compartment during oncogenesis.

We have inferred that PMD prefers a more compact structure than non-PMD from DNA methylation correlation and confirmed this using Hi-C data analysis, specifically, the scaling exponent of contact probability along genomic distance differs significantly between PMD and non-PMD (Zhang et al.).

Here we carefully examined the Hi-C patterns for PMDs and non-PMDs longer than 300 kb (Supplemental Material for review: Hi-C patterns for PMD and non-PMD). For non-PMD, the majority show the Hi-C pattern as shown in Fig. 4C which contains localized interaction domains and appears as several small squares along the diagonal of the Hi-C map. In contrast, the genomic regions inside each PMD contact nearly uniformly with each other and show a highly conserved pattern among PMDs typically as that shown in Fig. 4D.

We then construct 3D models for the two Hi-C patterns in Fig. 4C and 4D to further elucidate their structural differences. Their representative structures are presented in Figs. 4E and 4F, respectively. Considering the feasible computational cost, we used a 10-kb resolution of modeling at this domain scale in contrast to the 50-kb resolution in the modeling of the entire chromosome. Our modeling accurately captured the larger TAD numbers appeared in Fig. 4C than that in Fig. 4D. Significantly, type A is less compact than type B from our modeling (SI).

Such an observation promoted us to further cluster all the PMDs and non-PMDs into type A domain (typical as Fig. 4C) and type B domain (typical as Fig. 4D), according to their Hi-C patterns (Methods). Nearly all the PMDs (96%) are clustered into type B and the ratio for non-PMD clustered into type A is 75% (Fig. 4G).

These two types of chromatin constitute 54.4% of the entire human genome (comparable to the 67.9% coverage of annotated TADs) and have an

intimate relation with TADs and compartments. Types A and B chromatin are mainly located in compartments A and B, respectively (Fig. 4H). They also share boundaries with TADs, the same as PMD and non-PMD (Zhang et al.). Thus, types A and B chromatin are complementary to TADs and can be correspondingly regarded as building blocks for compartments A and B.

The two types of chromatin are of similar average lengths (588 kb and 744 kb for types A and B chromatin, respectively). One type A contains on average 2.34 TADs, significantly more than that of type B (0.99 TADs, p-value < 10^{-15} by Welch's unequal variances t-test). Therefore, the density of TADs also differs significantly between the two types of chromatin (3.99 TADs per Mb and 1.33 TADs per Mb for types A and B, respectively). The reason behind such large differences can be attributed to the different CTCF-binding site densities between types A and B and will be further elaborated in the next section.

Spatial Segregation of Chromatin Loop according to Loop Size

Intrigued by the different Hi-C patterns and structures of types A and B, we interrogated their underlying sequence differences. Types A and B are associated with GC rich and AT rich sequences, respectively. Most importantly, type A is 3.3 fold enriched for CTCF-binding sites (CBSs) than type B.

CBSs which are enriched at the boundaries of both TADs and chromatin loops play a crucial role in establishing chromosome structure (Guo et al., 2015). The recently proposed loop-extrusion model also highlights the

importance of CBS in that the locations of CBS alone can recapitulate most of Hi-C experimental results (Fudenberg et al., 2016; Sanborn et al., 2015). This model also unifies TADs and chromatin loops that TADs are formed by dynamic loops (Fudenberg et al., 2016). (Thus in the following we regard TAD and chromatin loop as exchangeable.) However, despite its success, the loop-extrusion model cannot explain chromatin compartmentalization (Sanborn et al., 2015).

We found that the density of CBS in compartments A and B also differs significantly (26.6 per Mb and 12.4 per Mb in compartments A and B, respectively). The higher CBS density in compartment A could result in smaller chromatin loops compared to those in compartment B, which is consistent with the different Hi-C patterns of the two chromatin types. The contrast of GC-content and CBS density in A/B compartments can also be easily seen by mapping these properties onto the chromatin model (Fig. 5A and 5B).

The spatial segregation between the two compartments implies that chromatin loops segregate according to their sizes, i.e., large loops tend to assemble with each other near nuclear periphery and small loops also spatially accumulate in the nuclear center.

Such a phenomenon strongly reflects the role of entropy in driving the segregation between two compartments due to “depletion attraction” (Asakura and Oosawa, 1958). Aggregation of large chromatin loops causes the overlap of excluded volumes surrounding them and increases the accessible volume

and entropy for small chromatin loops. The depletion force may also play a role in locating the large chromatin loops near the nuclear envelop which can be regarded as a large entity (Marenduzzo et al., 2006).

Given the highly crowding cellular environment, such non-specific entropy force might contribute to many processes in cells (Heermann, 2011; Marenduzzo et al., 2006; Snir and Kamien, 2005). One elegant example is that the replicated bacteria daughter strands spontaneously segregate due to entropic forces (Jun and Wright, 2010).

Our finding reveals that the segregation of loops of different sizes which maximizes the system entropy is likely a main driving force for the formation of the two compartments, as seen from the direct analysis of biological data and chromatin 3D structural modeling.

Loop-size Driven Chromosome Positioning

Chromosomes position in the nucleus in an activity-based way rather than according to chromosome sizes. Gene-rich chromosomes situate at the center of the nucleus, at the same time the more gene-poor chromosomes concentrate towards the nuclear periphery (Boyle et al., 2001).

The physical mechanisms underlying chromosome positioning is still unknown. It should be noted that the proteins of inner nuclear membrane may be not necessary for localizing chromosomes at the nucleus periphery (Boyle et al., 2001), suggesting the existence of non-specific driving forces.

For several chromosomes whose positions are well determined in FISH experiments, we calculated the CBS density and the median size of chromatin loops identified in Hi-C experiments (Table 1). The CBS densities of gene-rich chromosomes are much higher than gene-poor chromosomes.

Gene-rich chromosomes have much higher CBS densities and are expected to have smaller chromatin loops than gene-poor chromosomes. This is indeed the case as can be seen from the median size of chromatin loops identified from experimental Hi-C data (Table 1). The median loop size correlates well with chromosome positioning: chromosomes with smaller loop size segregate to the center of nucleus and larger loop size to nuclear periphery. Particularly, chromosomes 18 and 19 which have similar sizes show strikingly different CBS density and chromosome position.

These results are consistent with the entropy-driven chromatin compartmentalization that compartment A with smaller loop sizes locates near the nucleus center while compartment B with larger loop sizes is in the nucleus periphery.

Therefore, the size of chromatin loop plays an important role in determining chromatin structure at both intra- and inter-chromosome levels. An individual chromosome segregates into two compartments and chromosomes positions in the nucleus, both according to chromatin loop sizes.

Since loop-extrusion model can be used to address the formation of chromatin loops (or TADs) (Fudenberg et al., 2016; Sanborn et al., 2015), our

proposed mechanism of entropy-driven compartmentalization and positioning of chromosome complements this model and combining them offers a unified understanding of chromatin folding principle. Genomic regions with higher (lower) densities of GC-content and CTCF binding site tend to form smaller (larger) chromatin loops. The chromatin loops spatially segregate according to loop sizes due to an entropy-driven mechanism which result into the chromatin compartmentalization and also chromosome positioning.

DISCUSSION

Using Hi-C experimental data and a restraint-based modeling strategy, we built chromatin spatial structures. The correlated spatial segregation of compartments and various genome features were observed. Most importantly, this work provides new insight on the folding principle of chromosome. The fact that the spatial organization of chromatin and genome features are well correlated suggest the great importance of studies on the structure-function relationship for chromatins.

Explanation of the Power-law Scaling in Hi-C Contact Probability

The power-law scaling of Hi-C contact frequency along the genomic distance in the first Hi-C experiment was previously shown to be consistent with a fractal-globule polymer model instead of an equilibrium polymer at the megabase scale (Lieberman-Aiden et al., 2009). However, the scaling exponents for different chromosomes were found not to be conserved in later

studies which promotes people to reconsider the underlying fundamental principles of chromatin folding (Barbieri et al., 2013).

One schematic picture is the “strings and binders switch” model in which chromatin conformations are established through attachment of diffusible factors (binders, such as CTCF) to binding sites (Barbieri et al., 2012). The binder concentration regulates the compaction of chromatin and causes its heterogeneous character. Our study provides a more detailed confirmation of this scenario and shows that the mixing of A/B compartments (or type A/B chromatin) with different compaction accounts for the non-conserving scaling exponents in different chromosomes. The fraction of compartment A or B varies in different chromosomes and the scaling exponent of the contact frequency in the overall chromosome differs accordingly.

Most importantly, we highlight that the heterogeneous character of chromatin may result mainly from the distribution of CTCF-binding sites in two different kinds of compartments. Such results are also in accordance with the loop-extrusion model in that the locations of CTCF-binding alone can recapitulate most of Hi-C experimental results (Fudenberg et al., 2016; Sanborn et al., 2015). Both our results and loop-extrusion model emphasize the importance of CTCF-binding sites in chromatin organization which has been verified in genome editing experiments (Guo et al., 2015; Sanborn et al., 2015).

Structural Change of Chromatin during Cellular Differentiation

Cellular differentiation is closely related with the organization of chromatin (Dixon et al., 2015). The tissue-specific compartment changes tremendously in different cell types. Here we compare chromosome structures between IMR90 and H1 human ES (hES) or K562 cells to elucidate the structural change of chromatin during cellular differentiation. The Hi-C data for K562 is also obtained from Rao et al. (Rao et al., 2014), and for hES we used the data from Dixon et al. (Dixon et al., 2012). The comparisons between degree of compartmentalization and RT in these two cell lines are presented in Fig. S7.

The spatial structure and compartmentalization of hES and K562 cells are presented in Fig. 6A and 6B, respectively. Their overall organization and compartmentalization are obviously different from IMR90. Interestingly, in contrast to IMR90, there are long consecutive genomic regions assigned to the same compartment (A or B) in these two cells, consistent with extensive compartment switching in lineage commitment. About 25% of compartment alters between IMR90 and hES or K562. We also identified the spatial distribution of facultative and constitutive chromatin (Fig. 6C and 6D). It is found that constitutive regions lie in the interior of compartment A or B with facultative regions locate at the boundary between two compartments, coinciding with the fact that differences in replication timing in different cell types (developmental domains) are significantly less compartmentalized than constitutive domains (Dileep et al., 2015).

METHODS

Structural Modeling

We developed a restraint-based method to construct an ensemble of 3D chromosome models which were verified by the reproduction of experimental Hi-C contact frequencies. In our method, chromosome was coarse-grained as a polymer chain consisting of a string of beads. The polymer structure was optimized according to distance restraints derived from Hi-C data. To achieve this, we first converted the contact frequency matrix measured by the Hi-C experiment to a distance matrix which provides the spatial restraints for the coarse-grained beads. Then we performed molecular dynamics simulations starting from randomly generated initial conformations using biased potentials to generate an ensemble of conformations based on the restraint distance matrix. Further details and statistical analysis on these conformations are presented in SI.

Clustering for Type A/B

We clustered all the PMDs and non-PMDs longer than 300-kb into type A or B according to their Hi-C patterns. As mentioned in the main text, the genomic regions within an individual PMD possess a uniform contact pattern, and their contact variance decays differently along the genome compared with typical non-PMDs. Therefore, we chose the contact variance as the parameter for segment type classification. We labeled long segments by k-means clustering

and used them to train and validate the linear discriminant analysis classification model. The resulted model was then applied to all the PMDs and non-PMDs to classify them into type A/B.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found with this article online.

ACKNOWLEDGEMENTS

The work was supported by National Science Foundation of China (21233002, 21573006 and 91427304) and Ministry of Science and Technology of China (2012CB917304).

AUTHOR CONTRIBUTIONS

W.J.X. and Y.Q.G. conceived of and designed the study. W.J.X. proposed the folding principle and led data analyses. L.M. performed chromatin modeling and structural analyses. S.L. processed experimental data and performed Hi-C clustering. L.Z. and X.C. contributed to data analyses. Y.Q.G. supervised the study. W.J.X. wrote the manuscript with the inputs from L.M. and S.L. All authors contributed to the final version of the manuscript.

REFERENCES

Ali, T., Renkawitz, R., and Bartkuhn, M. (2016). Insulators and domains of gene expression. Curr. Opin. Genet. Dev. *37*, 17-26.

Asakura, S., and Oosawa, F. (1958). Interaction between particles suspended in solutions of macromolecules. J. Polym. Sci. *33*, 183-192.

Barbieri, M., Chotalia, M., Fraser, J., Lavitas, L.M., Dostie, J., Pombo, A., and Nicodemi, M. (2012). Complexity of chromatin folding is captured by the strings and binders switch model. Proc. Natl. Acad.

Sci. USA *109*, 16173-16178.

Barbieri, M., Scialdone, A., Gamba, A., Pombo, A., and Nicodemi, M. (2013). Polymer physics, scaling and heterogeneity in the spatial organisation of chromosomes in the cell nucleus. Soft Matter *9*, 8631-8635.

Berman, B.P., Weisenberger, D.J., Aman, J.F., Hinoue, T., Ramjan, Z., Liu, Y., Noushmehr, H., Lange, C.P., van Dijk, C.M., Tollenaar, R.A.*, et al.* (2012). Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. Nat. Genet. *44*, 40-46.

Bickmore, W.A., and van Steensel, B. (2013). Genome architecture: Domain organization of interphase chromosomes. Cell *152*, 1270-1284.

Boyle, S., Gilchrist, S., Bridger, J.M., Mahy, N.L., Ellis, J.A., and Bickmore, W.A. (2001). The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. Hum. Mol. Genet. *10*, 211-219.

Cremer, T., and Cremer, M. (2010). Chromosome territories. CSH Perspect. Biol. *2*, a003889.

Dekker, J., Marti-Renom, M.A., and Mirny, L.A. (2013). Exploring the three-dimensional organization of genomes: Interpreting chromatin interaction data. Nat. Rev. Genet. *14*, 390-403.

Dekker, J., and Mirny, L. (2016). The 3D genome as moderator of chromosomal communication. Cell *164*, 1110-1121.

Dileep, V., Ay, F., Sima, J., Vera, D.L., Noble, W.S., and Gilbert, D.M. (2015). Topologically associating domains and their long-range contacts are established during early G1 coincident with the establishment of the replication-timing program. Genome Res. *25*, 1104-1113.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W.*, et al.* (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331-336.

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.

Fraser, J., Williamson, I., Bickmore, W.A., and Dostie, J. (2015). An overview of genome organization and how we got there: From FISH to Hi-C. Microbiol. Mol. Biol. Rev. *79*, 347-372.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of chromosomal domains by loop extrusion. Cell Rep. *15*, 2038-2049.

Giorgetti, L., Galupa, R., Nora, E.P., Piolot, T., Lam, F., Dekker, J., Tiana, G., and Heard, E. (2014). Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. Cell *157*, 950-963.

Giorgetti, L., Lajoie, B.R., Carter, A.C., Attia, M., Zhan, Y., Xu, J., Chen, C.J., Kaplan, N., Chang, H.Y., Heard, E.*, et al.* (2016). Structural organization of the inactive X chromosome in the mouse. Nature *535*, 575-579.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y.*, et al.* (2015). CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. Cell *162*, 900-910.

Heermann, D.W. (2011). Physical nuclear organization: Loops and entropy. Curr. Opin. Cell Biol. *23*, 332-337.

Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat.

Methods *9*, 999-1003.
Jun, S., and Wright, A. (2010). Entropy as the driver of chromosome segregation. Nat. Rev. Microbiol. *8*, 600-607.
Kind, J., and van Steensel, B. (2010). Genome-nuclear lamina interactions and gene regulation. Curr. Opin. Cell Biol. *22*, 320-325.
Lesne, A., Riposo, J., Roger, P., Cournac, A., and Mozziconacci, J. (2014). 3D genome reconstruction from chromosomal contacts. Nat. Methods *11*, 1141-1143.
Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O.*, et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.
Marenduzzo, D., Finan, K., and Cook, P.R. (2006). The depletion attraction: An underappreciated force driving cellular organization. J. Cell Biol. *175*, 681-686.
Misteli, T. (2007). Beyond the sequence: Cellular organization of genome function. Cell *128*, 787-800.
Pirrotta, V. (2016). The necessity of chromatin: A view in perspective. CSH Perspect. Biol. *8*, a019547.
Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W.S., Denas, O., Vera, D.L., Wang, Y.L., Hansen, R.S., Canfield, T.K.*, et al.* (2014). Topologically associating domains are stable units of replication-timing regulation. Nature *515*, 402-405.
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., Lander, E.S.*, et al.* (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665-1680.
Rivera-Mulia, J.C., and Gilbert, D.M. (2016). Replication timing and transcriptional control: Beyond cause and effect - part iii. Curr. Opin. Cell Biol. *40*, 168-178.
Sanborn, A.L., Rao, S.S.P., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J.*, et al.* (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. USA *112*, E6456-E6465.
Serra, F., Di Stefano, M., Spill, Y.G., Cuartero, Y., Goodstadt, M., Bau, D., and Marti-Renom, M.A. (2015). Restraint-based three-dimensional modeling of genomes and genomic domains. FEBS Lett. *589*, 2987-2995.
Snir, Y., and Kamien, R.D. (2005). Entropically driven helix formation. Science *307*, 1067-1067.
Valton, A.L., and Dekker, J. (2016). Tad disruption as oncogenic driver. Curr. Opin. Genet. Dev. *36*, 34-40.
van Steensel, B., and Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nat. Biotechnol. *18*, 424-428.
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.
Zhang, B., and Wolynes, P.G. (2015). Topology, structures, and energy landscapes of human chromosomes. Proc. Natl. Acad. Sci. USA *112*, 6062-6067.
Zhang, L., Xie, W.J., Liu, S., Meng, L., and Gao, Y.Q. DNA methylation landscape reflects the spatial organization of chromatin in different cells. bioRxiv *doi: http://dx.doi.org/10.1101/080531*.

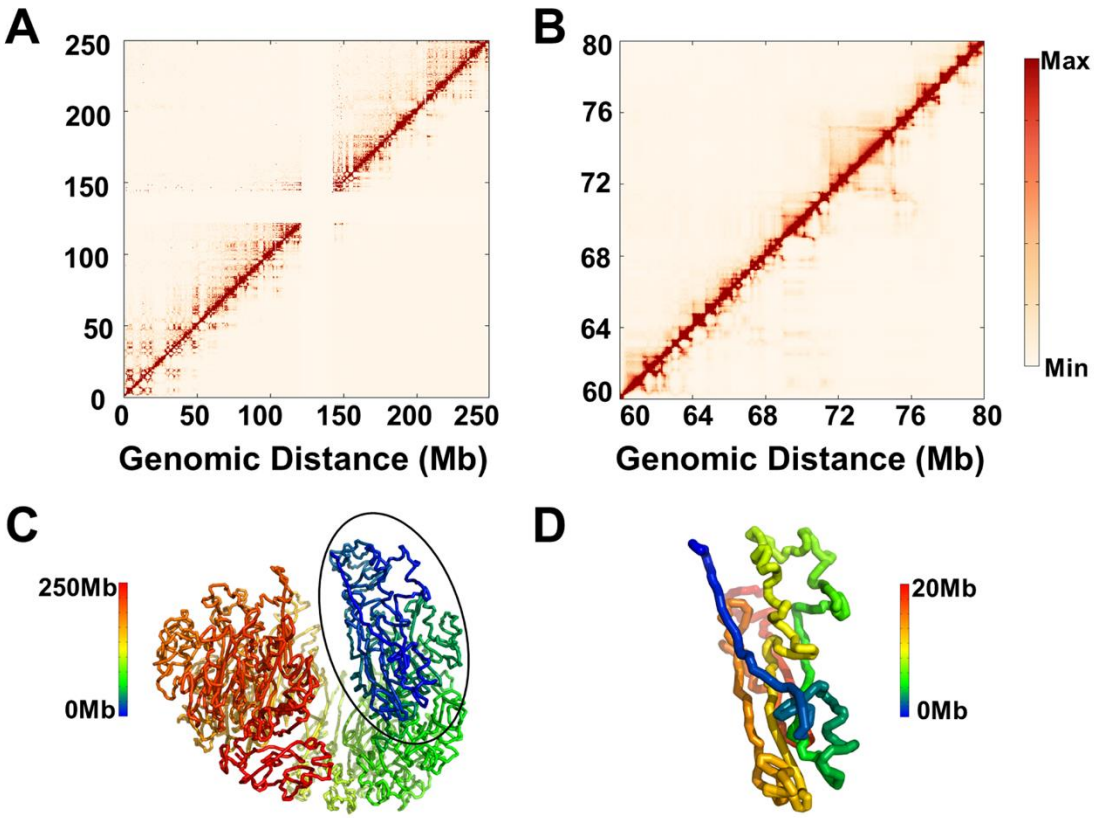

Figure 1. Validation of Modeling and Structural Characterization of Human Chromosomes.

(A) Comparison between experimental (upper triangle) and modeled (lower triangle) contact frequencies. (B) Zoomed-in version of A in a 20 Mb region. (C) Representative modeled structure. (D) Part of C (20 Mb) with enlarged view.

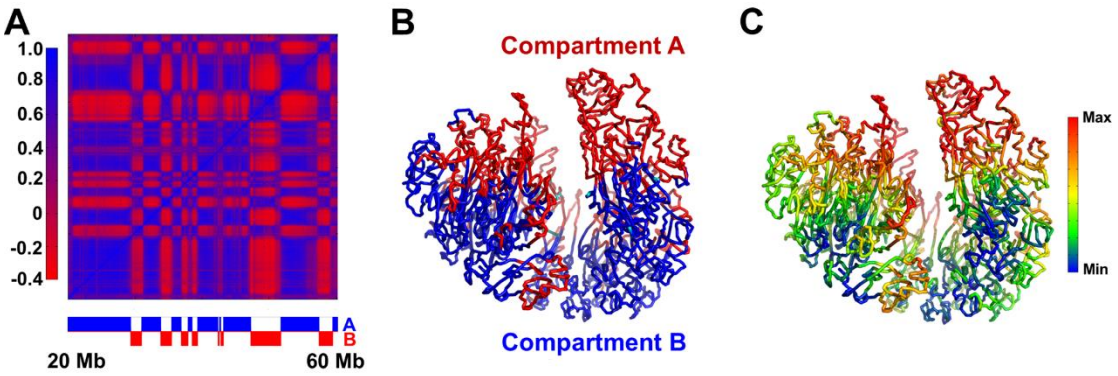

Figure 2. Characterization of A/B Compartments.

(A) Pearson correlation matrix. (B) Spatial segregation of compartment A (red color) and compartment B (blue color). (C) Projection of degree of compartmentalization onto representative conformation.

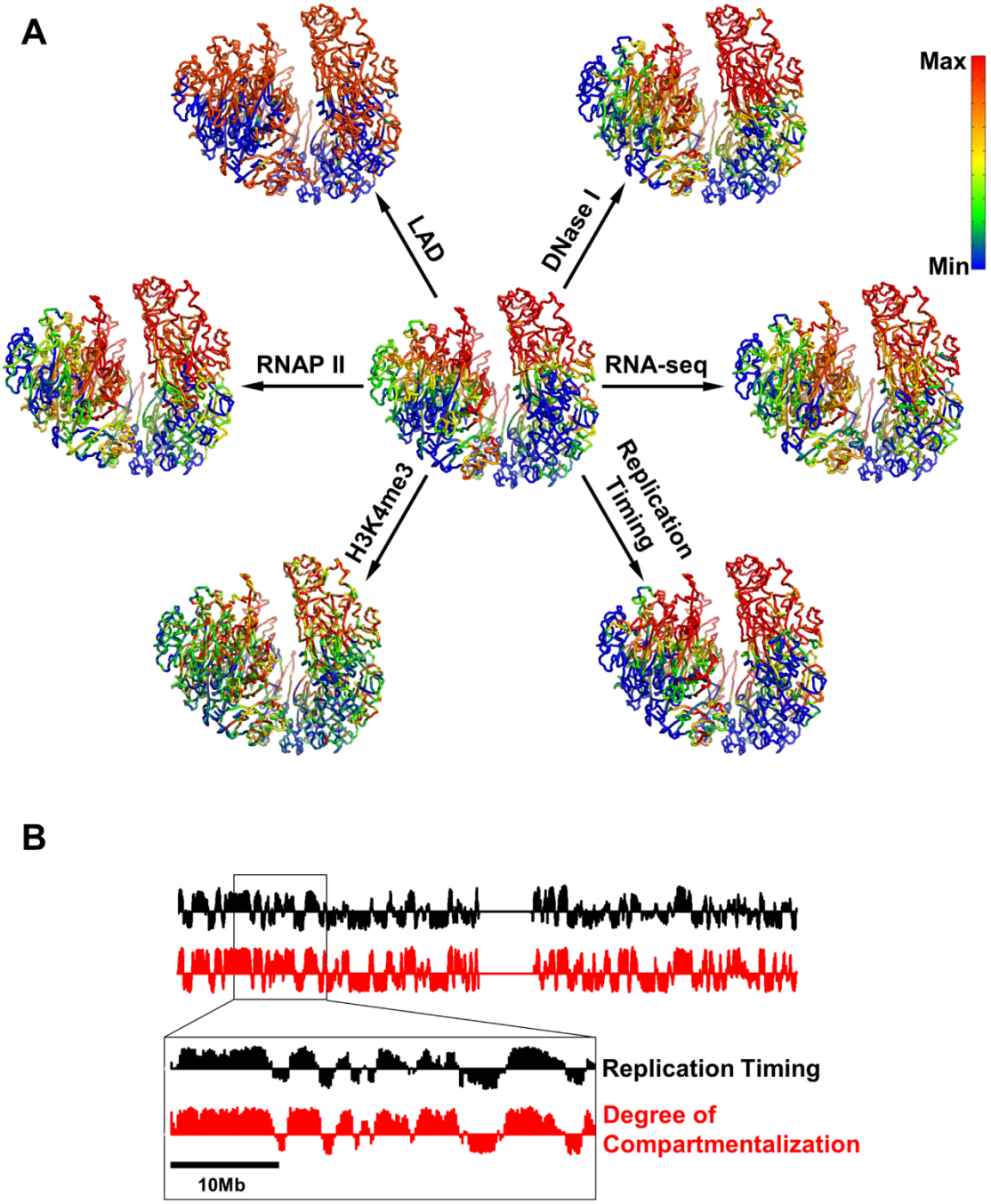

Figure 3. Spatial Segregation of Genome Features.

(A) Mapping of genome features onto the 3D chromatin structure. The degree of compartmentalization is shown in the center for comparison. Genomic features are presented as smoothed experimental data over different window sizes: 50-kb for LAD and H3K4me3, 200-kb for replication timing, 400-kb for

DNase I, 2-Mb for RNA-seq and RNAP II. The maximum/minimum of color bar stands for the value of average+/-standard deviation. The red color represents deficiency of lamina-associated domain (LAD), enrichment of DNase I hypersensitive sites (DNase I), enrichment of RNA polymerase II (RNAP II), active gene expression (RNA-seq), enrichment of active histone modifications (H3K4me3) and early replication timing (Replication Timing). (B) Comparison between replication timing and degree of compartmentalization along the chromosome.

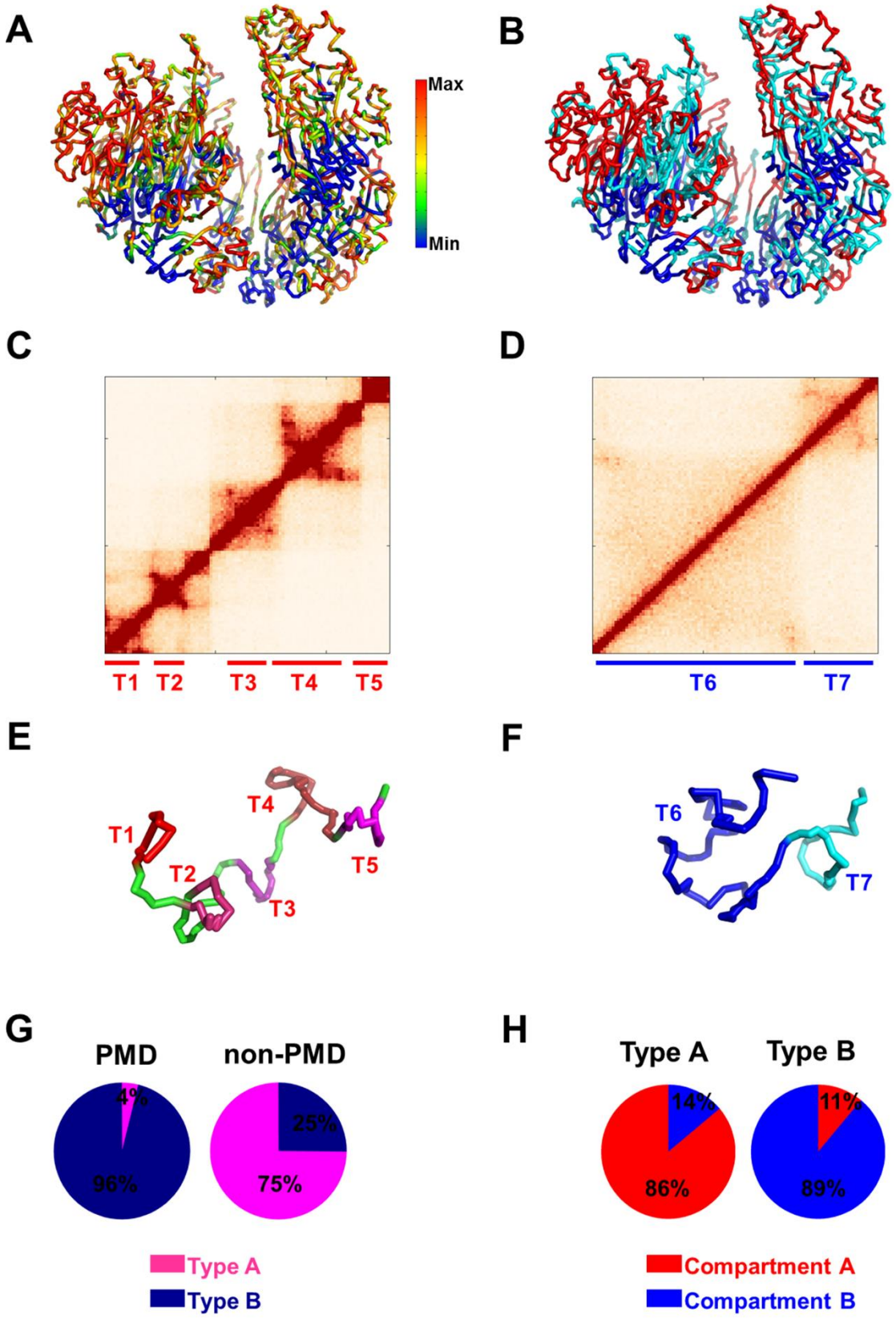

Figure 4. Characterization of Type A and B Chromatin.

(A) Spatial distribution of the DNA methylation level. Red color represents high

DNA methylation level. (B) Spatial segregation between PMD (blue color) and non-PMD (red color). Domains with size smaller than 300 kb are not grouped into type A or B (cyan color). (C) Typical Hi-C of type A chromatin (chr1: 15.46-16.54 Mb) (D) Typical Hi-C of type B chromatin (chr1: 233.47-234.50 Mb). (E) Modeled structure of type A chromatin, corresponding to Hi-C data in (C). (F) Modeled structure of type B chromatin, corresponding to Hi-C data in (D). TADs are annotated in Hi-C data (C,D) and correspondingly colored in blocks in (E,F). (G) Composition of type A/B chromatin in PMD/non-PMD. (H) Composition of compartments A/B in type A/B chromatins.

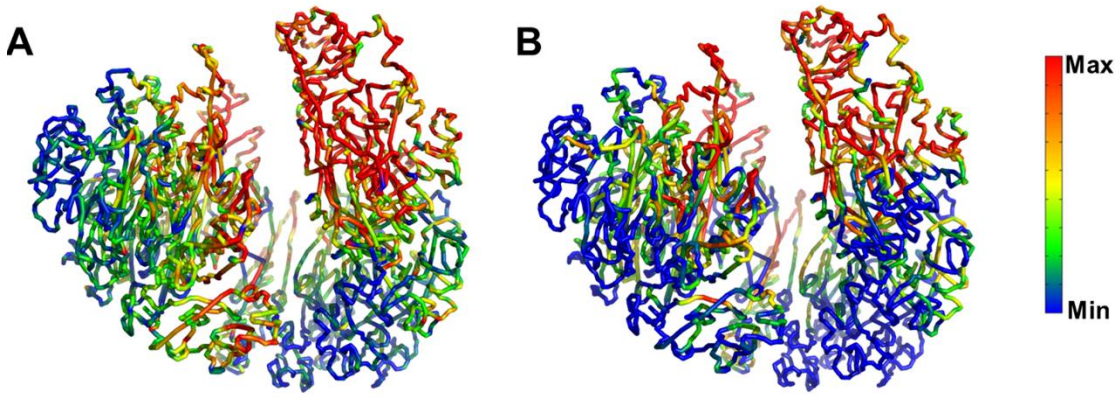

Figure 5. Mapping Sequence Properties onto the 3D Chromatin Structure.

(A) CpG density (Colored as the count of CpG in a 50-kb chromosome region).

(B) Density of CTCF binding site (Colored as the count of CTCF binding site in a 400-kb chromosome region).

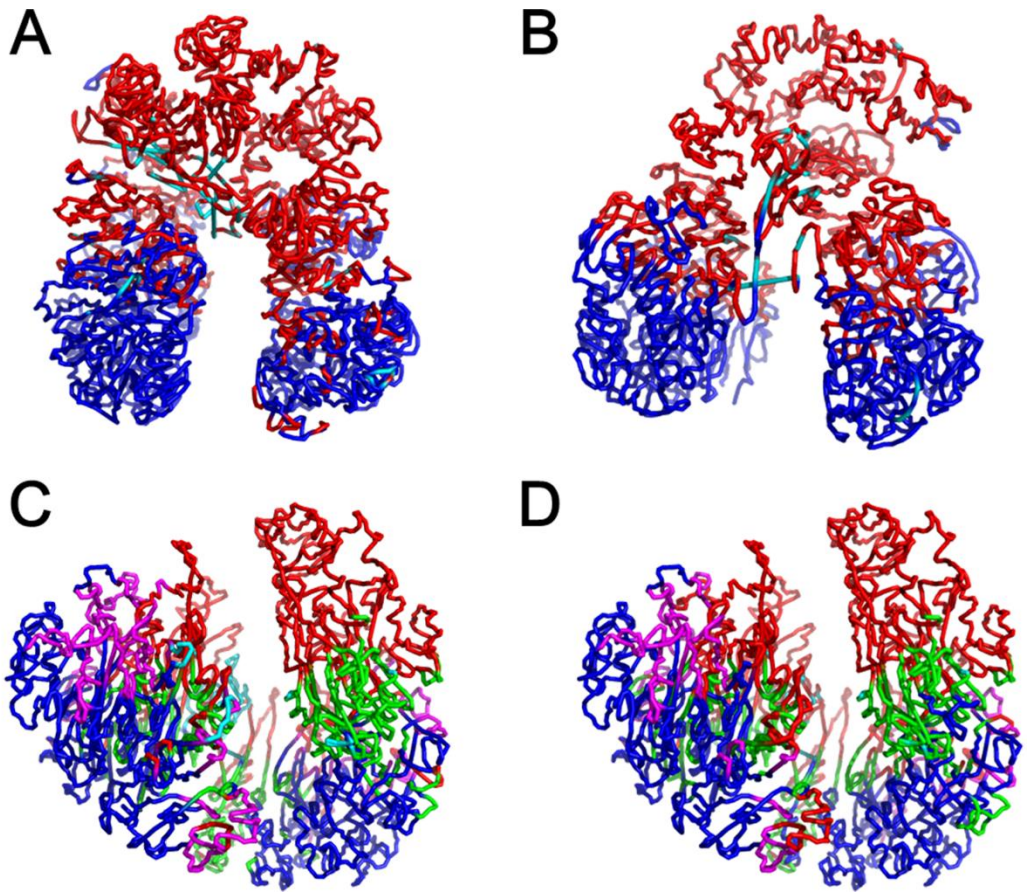

Figure 6. Structural Changes of Chromatin during Cellular Differentiation. (A,B) Structure and compartmentalization in hES (A) and K562 (B). Compartment A in red and compartment B in blue, the same as Fig.2B for IMR90. (C,D) Compartment switch between IMR90 and hES (C) or K562 (D) mapped in IMR90. Constitutive compartment A and B are colored in red and blue, respectively. Compartment A and B in IMR90 while changed in hES or K562 in magenta and green, respectively. Regions without Hi-C data are colored in cyan.

Table 1. Density of CTCF-binding Site and Median Size of Chromatin Loops in Different Chromosomes.

Chromosome	CBS density (Mb^{-1})	Loop median size (kb)
1	17.5	200
16	16.7	180
17	26.6	150
19	30.1	110
22	17.9	160
2	14.6	240
13	9.7	303
18	12.0	240

*Chromosome 1, 16, 17, 19, 22 are gene-rich and close to the nuclear center while chromosome 2, 13, 18 are gene-poor and close to the nuclear periphery for fibroblast cell (Boyle et al., 2001).