

PERSPECTIVE

Higher-order genomic organization in pluripotent stem cells

Ping Wang*, Weiqi Zhang*, Jiping Yang, Jing Qu✉, Guang-Hui Liu✉

National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
✉ Correspondence: ghliu@ibp.ac.cn (G.-H. Liu), jqu@salk.edu (J. Qu)

Recent applications of new tools for genome-wide mapping of long-range and spatial interactions have shed light onto the fundamental mechanisms of three dimensional chromatin organizations in pluripotent stem cells and their derivatives.

Pluripotent stem cells (PSCs), including embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs), can proliferate limitlessly and differentiate into many types of somatic cells *in vitro*, providing great promise for basic research and clinical therapy. The mechanisms underlying induction and maintenance of pluripotency are, however, still a critical question in this field. Regulation of chromatin status and dynamics at the epigenetic level are believed to define the “pluripotent” identity of PSCs. During somatic reprogramming and PSC differentiation, epigenomic information (chromatin status) rather than genetic information dramatically changes. Compared to differentiated cells, PSCs exhibit unique epi-genomic features, including different DNA methylation, histone modifications, and 3D nuclear organization of the genome (Zhang et al., 2012). Investigating the 3D chromatin organization as well as genome-wide epigenetic hallmarks in PSCs could dramatically improve our understanding on the molecular principles of pluripotency, reprogramming, and lineage commitment.

Epigenetics, defined as the mechanism of inheritable phenotypes and function changes among different progeny containing identical genomes, participates in most chromatin-related biological processes. Epigenetics regulates gene expression on several levels including, DNA methylation, histone posttranscriptional modification, nucleosome positioning and higher-order chromatin arrangement, intrachromosomal and interchromosomal interaction and interplay between chromosomal loci and nuclear elements, such as

the nucleolus and nuclear envelope (Lanctôt et al., 2007). Progress has been achieved in understanding pluripotency-associated gene regulation at DNA methylation and histone modification levels. However, the 3D architecture of chromosomes in the nucleus of PSCs remains poorly understood. Recent discoveries have depended on advances in new tools for genome-wide mapping of long-range and spatial interactions. Their applications in PSCs have shed light onto the fundamental mechanisms of 3D chromatin organization, using 3D fluorescence *in situ* hybridization (3D-FISH), DNA adenine methyltransferase identification (DamID), chromosome conformation capture (3C) technology and 3C derivatives (van Steensel and Dekker, 2010). Among these, the DamID approach has been successfully used to map the interplay between chromatin and nuclear lamina, a protein meshwork underneath the nuclear envelope (van Steensel and Dekker, 2010). Previous studies indicate that interactions with nuclear lamina could cause inactivation of local gene expression, leading lamina-associated domains (LADs) to be proposed as a new term associated with the repressive chromatin microenvironment (Melcer and Meshorer, 2010). Recently, Peric-Hupkes and his colleagues constructed a map of genome-wide LADs in mouse ESCs and their differentiated derivatives (Peric-Hupkes et al., 2010). During differentiation, many lineage specific genes detached from nuclear lamina, while pluripotent genes drew closer (Peric-Hupkes et al., 2010). Lars Guelen et al. mapped about 1300 LADs in human fibroblasts, which are on average 0.1–10 megabases in size and demarcated by the CTCF, an insulator protein (Guelen et al., 2008; Phillips and Corces, 2009). In addition, studies using new methods suggest that chromosomes fold into fractal globules individually and occupy distinct territories in the nucleus, which are known as chromosome territories (Meaburn and Misteli, 2007; Lieberman-Aiden et al., 2009). The spatial location of a specific gene within a chromosome territory appears to regulate gene ex-

*These authors contributed equally to the work.

pression in a topological domain model, which is consistent with the concept of LADs. Along this line, researchers agree that topological domains may serve as units for chromosome organization and gene regulation within the nucleus (Cremer and Cremer, 2010). In particular, two recent discoveries (Dixon et al., 2012; Nora et al., 2012) have brought great progress in understanding 3D chromatin organization in PSCs. They both identified long-ranged local chromatin interaction domains as structural blocks for chromatin architecture in ESCs, which are termed as “topological domains” or “topologically associating domains” (TADs) (Dixon et al., 2012; Nora et al., 2012) (Fig. 1A).

The Ren group analyzed genome-wide topological domains in mouse and human ESCs as well as human fibroblasts by Hi-C experiments (Dixon et al., 2012). As a derived technology of 3C, Hi-C can probe higher-order chromatin interactions of entire genomes by linking proximity-based ligation with high-throughput sequencing (van Steensel and

Dekker, 2010; Yaffe and Tanay, 2011). The Dekker group constructed the first Hi-C maps with a resolution of 1 megabase in the human genome, which uncovered several folding principles of chromatin conformation (Lieberman-Aiden et al., 2009). They showed that intrachromosomal and interchromosomal interactions comprehensively existed in human genomes to form fractal globule chromosome territories (Lieberman-Aiden et al., 2009). And the chromosome territories could be segregated to form open and closed chromatin neighborhoods, which are called A- and B-type compartments, respectively (Lieberman-Aiden et al., 2009). Ren’s group followed previous work and improved the resolution of Hi-C maps to less than 100 kilobases (kb). Computational modeling of Hi-C data with more than 1.7-billion read pairs helped them identify interaction patterns of the mammalian genome (Dixon et al., 2012). Firstly, they figured that highly self-interacting chromatin segments emerge with sizes less than 100 kb, which is defined as a topological domain.

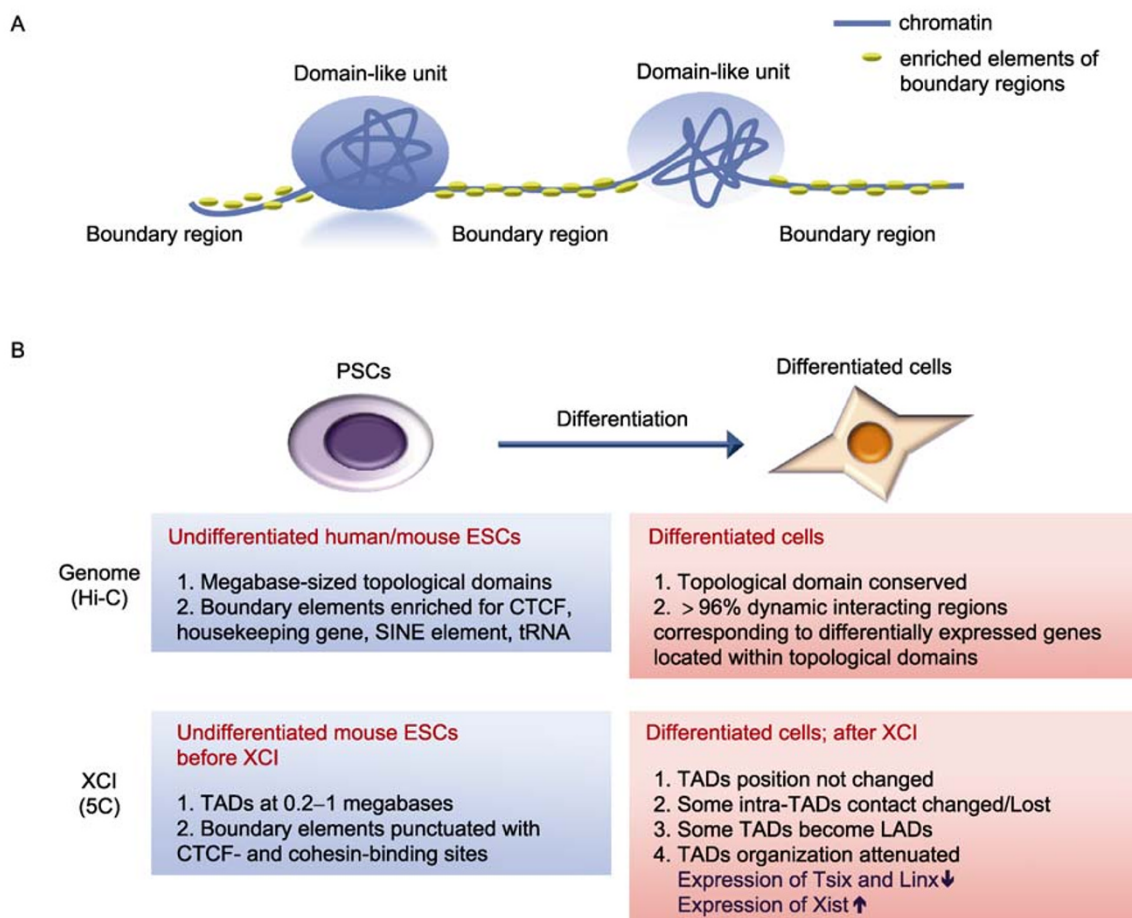


Figure 1. Schematic illustration of genomic organization as topological domains and flanking boundary elements in PSCs (A) and the alterations in properties of these domains upon differentiation (B). PSCs, pluripotent stem cells; XCI, X-chromosome inactivation; TAD, topological associated domain; LAD, lamina-associated domain.

The existence of topological domains was also supported by the fact that intrachromosomal interactions are dramatically more frequent than interchromosomal interactions. They located about 2200 topological domains in total with a median size of 880 kb, which cover almost 91% of the genome. They also labeled the genomic regions on the periphery of the topological domains as “topological boundary regions.” Surprisingly, they found most of the boundary regions are unchanged upon differentiation (Fig. 1B), which is contradictory to previous work showing chromatin configuration differently between cell types (Noordermeer et al., 2011). Detailed investigation revealed that the changed interacting regions were responding to lineage-specific gene expression and located within the same topological domain, which may explain the consistency of genome-wide topological architecture (Dixon et al., 2012). Secondly, topological domains were compared with previously reported domain-like structures of the genome, such as A- and B-type compartments, and LADs (Lieberman-Aiden et al., 2009; Melcer and Meshorer, 2010). The results showed that these domains are indeed correlated, but are not identical. Of note, the topological domains identified by Hi-C may correspond to domains of open or closed chromatin (Dixon et al., 2012). Thirdly, they explored what factors are enriched in the boundary regions and may contribute to the segregation of topological domains. CTCF and housekeeping genes are strongly enriched in topological boundaries, and account for nearly one-third of all segregation elements in the genome. Furthermore, tRNA genes, transcription start sites, global run on sequencing and retrotransposons are also enriched in boundary regions (Dixon et al., 2012). Taken together, their findings indicate a model for the spatial conformation of genomic DNA consisting of topological domains linked by topological boundary regions. The model seems well conserved in evolution, as mouse and human ESCs exhibit a similar pattern of topological domains in genome organization.

Using chromosome conformation capture carbon-copy (5C) technology, Heard and his colleagues provided additional findings indicating that topological segments may be fundamental domains for genome organization (Nora et al., 2012). The 5C approach, adapted from 3C, could detect all interactions among selected panels of loci (Dostie et al., 2006; van Steensel and Dekker, 2010). In their study, the authors focused on the biological process of X-chromosome inactivation (XCI) in mouse ESCs, which is affected by *Xist* activity. *Xist* activity in turn is modulated by the antisense transcript *Tsix* and by regulators such as *Xite*, *DXPas34* and *Tsx*. After simultaneously investigating almost 250,000 possible chromosome contacts in undifferentiated mouse ESCs in an unbiased manner, they firstly identified a set of discrete genomic blocks with a size range from 0.2 to 1 megabase, which preferentially form long-range interactions over 50 kb. Further experiments confirmed that folding domains spanning the sub-megabase scale are not random, and are termed TADs.

Of note, these domains are smaller than the topological domains identified by Hi-C. Thus TADs may represent a more subtle 3D organization of chromatin. Similar to Ren's work, Nora et al. found discrete boundary elements flanking the TADs, which contain CTCF- and cohesion-binding sites (Nora et al., 2012). Secondly, the researchers also found *Xist* and *Tsix* are within neighboring TADs. To address the function of boundary regions in the formation of TADs, the authors knocked out the sequence between the *Xist* and *Tsix* TADs in mouse ESCs. The deletion of the boundary between the *Xist* and *Tsix* TADs led to the ectopic contact of neighboring TADs and an altered organization of boundary elements (Nora et al., 2012). Thirdly, the pattern of TADs is retained during differentiation or XCI. When comparing 5C data from mouse embryonic fibroblasts and neuronal progenitor cells to mouse ESCs, no dramatic position changes of TADs were found. However, attenuated interactions within TADs were observed and some TADs were found to turn into LADs upon differentiation (Fig. 1B). These results indicate that XCI may largely be regulated by internal organization of LADs rather than organization of TADs (Nora et al., 2012). Fourthly, this work probed the relationship between TAD organization and dynamic gene expression changes in specific X chromosome loci during early development. Physical proximity of gene clustering within the same TADs may coordinate with long-range transcriptional control of gene expression. Deletion of TAD boundaries leads to transcriptional misregulation to a great extent. Thus, distant interactions within TADs are necessary for correct regulation of *Tsix* during development (Nora et al., 2012). Lastly, they identified a novel long-ranged regulatory element within the *Tsix* TAD, which produces *Linx*, a large intervening non-coding RNA. The latter plays a role in the long-range transcriptional regulation of *Tsix* (Nora et al., 2012). Collectively, the analysis of *Xist/Tsix* loci also makes it clear that the spatial compartmentalization of TADs divides X-inactivation centre (*Xic*) loci into two opposite regulatory domains, which facilitates the dynamic and temporal regulation of XCI. Therefore, the three-dimensional architecture of *Xic* regulates the long-range transcriptional activity of *Xist/Tsix*.

Thus, both reports revealed that the human genome could be divided into large and demarcated topological domains, flanked by boundary elements. The borders are punctuated with specific regulatory elements, such as CTCF-binding sites (Dixon et al., 2012; Nora et al., 2012). The discovery of topological domains as architectural units of chromatin organization is truly exciting. However, further experiments are needed to investigate (1) how spatial organization of genomes is established in different cell types; (2) the molecular mechanisms by which boundary elements may contribute to the 3D structure of chromatin; (3) the effect of chromatin architecture on gene expression as well as other layers of epigenetic modifications; (4) the role of subnuclear chromatin organization in reprogramming, differentiation, and direct

lineage conversions; and (5) how spatial conformation of chromosomes responds to physiological and pathological stimulations and contributes to human aging and diseases (Liu et al., 2011a). With several novel genome-wide detection approaches on hand, we are able to construct an integrated view of chromosome architecture, which could greatly improve our understanding of 3D chromatin organization. It would also be helpful to couple high-throughput single-cell sequencing technologies with existing technologies to realize molecular mapping of chromatin architecture for single cells. On the other hand, the combination of gene-editing tools (Li et al., 2011; Liu et al., 2011a, b, 2012), with novel readouts for higher-order chromatin conformation could provide a powerful and promising platform to reveal the mechanisms underlying 3D chromosome architecture and gene regulation.

ACKNOWLEDGEMENTS

G.-H.L. was supported by the National Laboratory of Biomacromolecules, the 100 Talents Program of Chinese Academy of Sciences, a CIRM grant (No. TG2-01158), and a Glenn Foundation grant. J.Q. was supported by the AFAR/Ellison Medical Foundation.

REFERENCES

- Cremer, T., and Cremer, M. (2010). Chromosome territories. *Cold Spring Harb Perspect Biol* 2, a003889.
- 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* 2012 Apr 11. doi: 10.1038/nature11082. [Epub ahead of print]
- Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res* 16, 1299–1309.
- Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453, 948–951.
- Lancôt, C., Cheutin, T., Cremer, M., Cavalli, G., and Cremer, T. (2007). Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat Rev Genet* 8, 104–115.
- Li, M., Suzuki, K., Qu, J., Saini, P., Dubova, I., Yi, F., Lee, J., Sancho-Martinez, I., Liu, G.H., and Izpisua Belmonte, J.C. (2011). Efficient correction of hemoglobinopathy-causing mutations by homologous recombination in integration-free patient iPSCs. *Cell Res* 21, 1740–1744.
- 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.
- Liu, G.H., Barkho, B.Z., Ruiz, S., Diep, D., Qu, J., Yang, S.L., Panopoulos, A.D., Suzuki, K., Kurian, L., Walsh, C., et al. (2011a). Recapitulation of premature ageing with iPSCs from Hutchinson-Gilford progeria syndrome. *Nature* 472, 221–225.
- Liu, G.H., Sancho-Martinez, I., and Izpisua Belmonte, J.C. (2012). Cut and paste: restoring cellular function by gene correction. *Cell Res* 22, 283–284.
- Liu, G.H., Suzuki, K., Qu, J., Sancho-Martinez, I., Yi, F., Li, M., Kumar, S., Nivet, E., Kim, J., Soligalla, R.D., et al. (2011b). Targeted gene correction of laminopathy-associated LMNA mutations in patient-specific iPSCs. *Cell Stem Cell* 8, 688–694.
- Meaburn, K.J., and Misteli, T. (2007). Cell biology: chromosome territories. *Nature* 445, 379–781.
- Melcer, S., and Meshorer, E. (2010). The silence of the LADs: dynamic genome-lamina interactions during ESC differentiation. *Cell Stem Cell* 6, 495–497.
- Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., De Laat, W., and Duboule, D. (2011). The dynamic architecture of Hox gene clusters. *Science* 334, 222–225.
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 2012 Apr 11. doi: 10.1038/nature11049. [Epub ahead of print]
- Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., et al. (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell* 38, 603–613.
- Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. *Cell* 137, 1194–1211.
- van Steensel, B., and Dekker, J. (2010). Genomics tools for unraveling chromosome architecture. *Nat Biotechnol* 28, 1089–1095.
- Yaffe, E., and Tanay, A. (2011). Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat Genet* 43, 1059–1065.
- Zhang, W., Ding, Z., and Liu, G.H. (2012). Evolution of iPSC disease models. *Protein Cell* 3, 1–4.