INVITED REVIEW

# Between form and function: the complexity of genome folding

A. Marieke Oudelaar[1], Lars L.P. Hanssen[1], Ross C. Hardison[2],
Mira T. Kassouf[1], Jim R. Hughes[1] and Douglas R. Higgs[1],[*]

[1]MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, OX3 9DS,
UK and [2]Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park,
PA 16802, USA

*To whom correspondence should be addressed. Tel: +44 1865 222393; Fax: +44 1865 22250; Email: doug.higgs@imm.ox.ac.uk

## Abstract

It has been known for over a century that chromatin is not randomly distributed within the nucleus. However, the question of how DNA is folded and the influence of such folding on nuclear processes remain topics of intensive current research. A longstanding, unanswered question is whether nuclear organization is simply a reflection of nuclear processes such as transcription and replication, or whether chromatin is folded by independent mechanisms and this *per se* encodes function? Evidence is emerging that both may be true. Here, using the α-globin gene cluster as an illustrative model, we provide an overview of the most recent insights into the layers of genome organization across different scales and how this relates to gene activity.

## Introduction

Nuclear organization is currently described at vastly different scales. Many nuclear structures, such as euchromatin, heterochromatin, the nuclear lamina, and non-membrane bound nuclear sub-compartments, can be identified by conventional microscopy. Technological advances have revealed finer detail of nuclear structure, including chromosome territories and the organization of telomeres and centromeres. More recently, sub-microscopic features of nuclear structure such as chromatin folding, chromatin structure, and distribution of chromatin-associated proteins have been determined, and many of these features can now be correlated with specific DNA sequences (1,2). An important aim of current research is to relate these diverse observations, at different levels of resolution, to each other and to obtain an integrated understanding of the relationship between nuclear organization and function. This is likely to provide important insights into how mutations affecting various aspects of nuclear organization may cause human genetic disease.

A key question in this field, discussed here, is whether sub-chromosomal organization is predominantly an emergent property of a wide variety of nuclear processes, or established independently, representing 'building blocks' of the genome.

## Self-Interacting Domains of Chromatin

It has previously been shown that individual eukaryotic chromosomes fold in such a way that they more frequently interact with themselves rather than with other chromosomes and occupy distinct chromosome territories in the nucleus (3). Analyses at higher resolution using chromosome conformation capture (3C) techniques (4,5) have suggested that individual chromosomes may be further organized into a variety of self-interacting domains.

One of the first classes of self-interacting domains described relatively large regions of chromatin (100–5 Mb) observed in Hi-C heat maps, which were named topologically associated domains (TADs) (6,7). TADs were originally defined by the preferential interaction of regions located within the same domain and the relative depletion of interactions between regions located in adjacent domains. TAD boundaries were initially identified using a directionality index (DI) (6), which quantifies the degree of upstream or downstream interaction bias for a genomic region. Several other algorithms for TAD identification have been developed, and it has been shown that the number of TADs called can vary depending on the algorithms used, the thresholds applied, and the resolution of the underlying data (8). Nevertheless, many TADs correlate with chromatin signatures, gene activity and replication timing (7,9,10). Based on Hi-C heat maps at relatively low resolution, it has been suggested that TADs are relatively stable across different cell types and conserved across species. At present TADs are the most frequently described self-interacting domains. Typical examples are shown in mouse embryonic stem (mES) cells in Figure 1A.

Similar Hi-C data have also been interpreted from a different perspective that describes larger structures called compartments. These are composed of transcriptionally active or inactive TAD-like structures which cluster together to form functionally distinct nuclear compartments (11,12). The active TAD-like regions (compartment A) are thought to form higher-order structures in the interior of the nucleus; the inactive TAD-like regions (compartment B) are located closer to the periphery. It has been shown that lamina-associated domains (LADs), which have been associated with gene repression, largely correspond to compartment B (2,13).

More recent analyses of nuclear compartments and TADs using a variety of 3C techniques have revealed previously unrecognized, smaller, self-interacting domains that have been referred to as contact domains (12), sub-TADs (14), insulated neighborhoods (15) and frequently interacting regions (16).

A common feature of all self-interacting domains, is that their borders are enriched for boundary elements that among many other proteins commonly bind CTCF, a highly conserved DNA-binding zinc finger protein. CTCF is strongly implicated in the formation of self-interacting domains, which often depends on the presence of pairs of CTCF-binding motifs in convergent orientations at the domain borders (12,17–21).

The definitions of the structures discussed above are rather vague and their identification somewhat arbitrary (8). The described domains are all identified by their self-interacting features, boundary elements, and sizes. To an extent, self-interacting domains appear to be fractal structures in the genome; the definition of each unit depending on the level of resolution. It is therefore unclear if they represent fundamentally different structures, or if they are simply variants of a similar organizational process.

## The Formation of Self-Interacting Domains

It has been proposed that self-interacting domains are formed by a process of loop extrusion. This is thought to be an active process in which a loop-extruding factor, containing two DNA-binding units, associates with chromatin and travels along the fiber in opposite directions, creating a progressively larger intervening loop, until the factor is stalled at boundary elements, such as convergent CTCF-binding motifs. Hetero-dimerization between CTCF proteins or other boundary proteins then tethers the extrusion factor, defining the limits of the self-interacting domain (20,22). It can be predicted that during the dynamic formation of such loops in a population of cells, every region within a self-interacting domain, at some point, contacts every other region lying between the border elements. This is consistent with the increased interaction frequencies observed between all genomic regions within such domains either at low (Fig. 1A) or high resolution (Fig. 1B). Based on its co-localization with CTCF at domain boundaries, cohesin, a ring-shaped protein complex that contains ATPase activity and can interact with two double-stranded DNA helices, is thought to be a strong candidate for an extruding factor in the mammalian genome (23–28).

In support of the loop extrusion model, it has been shown in various studies that mutation or deletion of specific CTCF-binding sites can disrupt the border of the associated self-interacting domain, and cause a change in genome structure that allows enhancers within the domain to interact with genes in neighboring domains, resulting in mis-regulated gene expression (7,15,17,29–35). This supports the important role of CTCF/cohesin boundaries in delimiting self-interacting domains, but as these proteins are ubiquitously expressed and widely bound to chromatin, they alone cannot explain the tissue-specific structures observed in some of these studies (30,35). This is clear in the murine α-globin locus shown in Figure 1B, which is flanked by two double CTCF-binding sites, which are occupied by CTCF and cohesin in both erythroid and non-erythroid (mES) cells, but only interact in erythroid cells.

A possible explanation could be that cohesin is recruited to chromatin in a tissue-specific manner. It has been shown that cohesin is not only associated with CTCF-binding sites, but also with active promoters and enhancers and that dynamic cohesin occupancy of enhancers correlates with tissue-specific interactions between promoters and enhancers (36–38). Furthermore, it has been shown that active transcription is involved in distributing cohesin along the genome (39).

It is important to keep in mind that the self-interacting domains observed in 3C analyses represent a population average of a dynamic structure and that the spatial separation between TADs is not absolute. Insulation between domains does not completely prevent the formation of contacts across boundaries (12) and such interactions can have functional effects on gene expression (40). These long-range interactions are infrequent and therefore likely to be highly dynamic. Single-cell Hi-C studies have identified TADs in individual cells, but have also shown that the level of compaction within TADs is variable and that their higher-order organization in larger structures (e.g. compartments) is stochastic (41,42), which can be explained from the assumed polymeric behavior of chromosomes (10).

## The Conservation of Self-Interacting Domains Across Species, Tissues and Developmental Stages

Comparisons between mouse and human Hi-C maps from ES and B-lymphoblast cells have shown that approximately half of the identified large self-interacting domains in orthologous regions are conserved (9,12). A similar conservation of chromosome structure was observed in a study in which Hi-C maps from hepatocytes isolated from four mammalian species were examined (18). Furthermore, these self-interacting domains have been shown to be similar in different tissues. Comparisons between cell types have shown that approximately 50–60% of TADs (6,12,16,43) and approximately 40% of the functionally
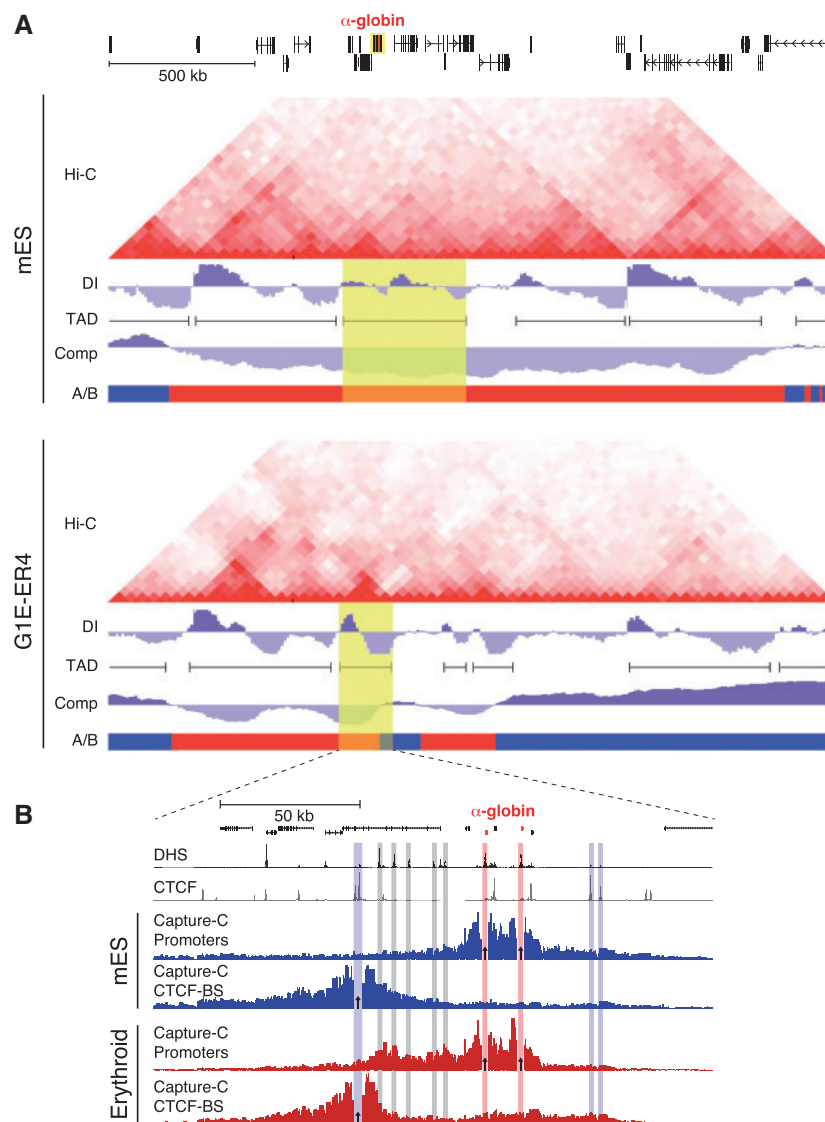
**Figure 1.** The structural organization of the murine α-globin cluster in its inactive and active conformation. (**A**) Structural features of the α-globin cluster (2.5 Mb) defined by Hi-C in inactive mES cells and G1E-ER4 erythroid progenitor cells with basal α-globin transcription. The adult α-globin genes (*Hba-1* and *Hba-2*; shown in red) are located in a gene-rich region of the genome shown at the top. The Hi-C heat maps underneath display interaction frequencies in mES cells and G1E-ER4 cells at 40 kb resolution (64,65). The tracks below each heat map display the directionality index (DI), from which the TAD calls shown are derived. Compartment scores (Comp) and the separation into compartment A (active; shown in red) or B (inactive; shown in blue) are displayed at the bottom. The TADs containing the α-globin cluster are highlighted in yellow. The Hi-C heat map and derived features were generated using resources from the VISION project (http://www.bx.psu.edu/~giardine/vision/), including the 3D browser (66). (**B**) Detailed interaction data describing the α-globin locus (200 kb) at high-resolution using Capture-C data in inactive mES cells and active primary erythroid cells. The α-globin genes are shown in red at the top, followed by tracks displaying DNaseI Hypersensitivity Sites (DHS) and CTCF occupancy in primary erythroid cells. Capture-C interaction profiles generated from the viewpoint (indicated by arrows) of the duplicated α-globin promoters and an upstream CTCF-binding site (CTCF-BS) that forms one of the domain boundaries, are shown below (35,54).

distinct A/B compartments (16) are invariant. In some areas of the murine genome, self-interacting domains are thought to encompass regulatory regions (enhancers and promoters) to form structures, early in development, that are found in all tissues and which are subsequently modified in a cell type-specific manner (44). A similar picture has been observed in Drosophila, where many long-range interactions are similar across tissues and developmental stages (45). Importantly, rearrangements in mouse and human leading to duplication or deletion of entire self-interacting domains at the *Sox9* locus have been shown to be less disruptive to gene expression and cause less extensive changes in phenotypes than those which partially disrupt a self-interacting domain (34).

These observations at specific loci provide preliminary evidence for a mechanism which organizes the genome early in development independently of gene expression and chromatin modification and yet plays an important role in regulating subsequent gene expression. The possibility that these structures form before gene activation would predict that specific factors, elements or processes are capable of generating a chromatin environment that promotes the formation of interactions between regulatory *cis*-elements (46). Mutations affecting such factors would alter chromatin structure and therefore gene regulation, while other mutations may only affect the ability of a *cis*-element to influence a promoter, without substantially changing the structure. Although CTCF, cohesin and their

binding sites are strong candidates, *direct* experimental evidence supporting their role in forming such constitutive structures imposing a regulatory role in gene expression is limited, but under intense investigation.

Functional experiments have shed further light on the role of CTCF and cohesin in the formation of self-interacting domains and their effects on gene expression. A recent study showed that complete depletion of CTCF resulted in a loss of TAD structure, but left active and inactive genome compartments properly segregated (47). Surprisingly, the effect on gene expression appeared relatively modest. In contrast to the strong effects of CTCF on genome architecture, early studies of cohesin depletion showed only moderate changes (48–50). However, depletion in these studies was not complete, which probably confounded these conclusions. A more recent study used deletion of the cohesin loading factor *Nipbl* to efficiently displace cohesin from chromatin. This had a greater effect on gene expression and caused dramatic genome reorganization in which TADs vanished globally. The disappearance of TADs unmasked a finer compartment structure that reflected transcriptional activity and chromatin state and these domains remained segregated into compartments A and B (51).

## The Interplay Between Self-Interacting Domains, Nuclear Processes and Gene Activity

The observations described above suggest that CTCF and cohesin play a role at some, but by no means all, CTCF-binding *cis*-elements in establishing and/or maintaining self-interacting domains. However, the observation that some aspects of self-interacting domains persist in the absence of CTCF or cohesin, shows that these structures are more complex than first thought. This has also been suggested by computer simulations of polymer models (52,53). A recent study showed that models based on CTCF-mediated architecture alone could not predict *in vivo* genome organization. However, when interactions between other *cis*-elements were included in the models, the observed self-interacting structures could be better explained (52).

It seems therefore possible that there are at least two (largely) independent mechanisms underlying self-interacting domains. One might involve a process of loop extrusion that depends on CTCF, cohesin and/or other boundary associated proteins. This sets up constitutive architectural domains that are largely tissue-invariant and separate some regions of the genome into domains that restrict interactions between regulatory elements.

In addition, nuclear processes such as changes in chromatin accessibility, the binding of key transcription factors to *cis*-elements, increased enhancer-promoter interactions and active transcription may further shape the structure of these domains. Cohesin may also play a role in these interactions (37). The domains formed by these mechanisms are highly tissue-specific (54) and cluster together to form compartments with similar chromatin states and activities.

Although some of the processes driving constitutive and cell type-specific structures may be independent, they may nevertheless influence each other. For example, the constitutive structures thought to be set up by cohesin/CTCF may constrain interactions between regulatory elements and therefore contribute to their specificity (55–58). Furthermore, the recruitment and localization of cohesin in tissue-specific domains might be influenced by transcription factors, co-activators and the active process of transcription (37,39).

The structures related to nuclear processes were initially described as chromatin 'loops' formed specifically between active promoters, enhancers and boundaries (59). However, high-resolution, quantitative 3C data show that rather than specific peaks of interaction over these *cis*-elements, broad, self-interacting domains may appear over relatively large genomic regions upon activation of such elements in specific cell types. Within this region, increased interactions over certain elements can be identified (54,60) (Fig. 1). This is consistent with a mechanism such as loop extrusion promoting a general increase in self-interaction over a given genomic region delimited by boundary elements. Increased interactions over certain *cis*-elements suggest that when such elements, bound by multi-protein complexes, are brought together, they form more stable or frequent interactions than the rest of the self-interacting domain.

Though nuclear processes related to gene expression likely influence genome structure, current evidence argues against a role for transcription *per se* in establishing self-interacting domains: a recent study of chromatin conformation during Drosophila embryogenesis showed that the appearance of TADs occurs independently of transcription (61). However, this study also showed that although higher-order chromatin conformation and intra-TAD contacts are still present, albeit reduced, in embryos in which transcription was abolished, the lack of transcription nevertheless affects the properties of TAD organization and resulted in a significant loss of inter-TAD insulation.

Overall, these studies suggest that the structures observed in contact matrices result from a mixture of different mechanisms and reflect an intricate interplay between form and function (Fig. 2).

## Analysis of Self-Interacting Domains Containing and Surrounding the α-Globin Cluster

The globin loci have been used extensively to establish general principles of mammalian genome organization and gene regulation and relate perturbations in these processes to human genetic disease. Principles first established at these loci have usually been found to be true of many other loci (62,63).

To examine the concepts discussed here, we analyzed 3C interaction data for the well-characterized murine α-globin cluster. Some self-interacting domains in this gene-dense 2.5 Mb region appear as relatively discrete structures in mES cells (Fig. 1A, top panel). However, other areas form poorly defined structures, and appear to be organized in a fractal pattern. This suggests that even though some areas of the genome are organized into well-defined, preformed, self-interacting domains, this by no means applies to the entire genome.

Figure 1 also shows a comparison between mES cells, in which the α-globin cluster is not expressed, and G1E-ER4 cells, in which α-globin is expressed at basal levels. This specific example clearly shows that the self-interacting domains, which in this case correspond to TADs, are dynamic, and quite different between these two cell types. Activation of the α-globin genes in G1E-ER4 cells is accompanied by increased interactions between promoters and enhancers, the appearance of smaller, more specific self-interacting domains, and a shift of regions adjacent to the active α-globin cluster from compartment A (active) to compartment B (inactive). This shows that not all self-interacting domains are conserved across tissues.

Comparison between primary erythroid and mES tissue at higher resolution (Fig. 1B) shows that activation of the α-globin genes and their *cis*-acting regulatory elements causes the formation of a strong self-interacting domain, in which contacts
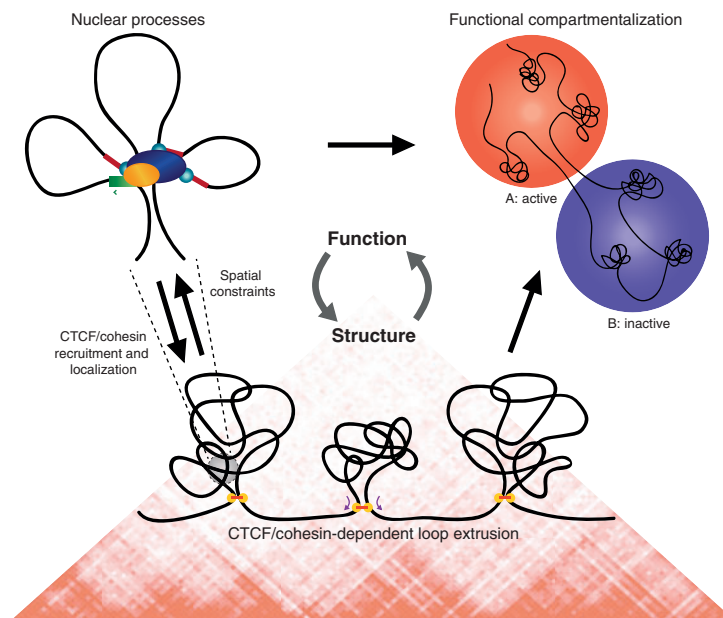
**Figure 2.** Proposed model for the processes contributing to chromatin structure and their interplay with genome function. TAD-like structures observed in Hi-C maps may be formed by a process of CTCF/cohesin-dependent loop extrusion (bottom). These domains restrain the connectivity between genes and regulatory elements, but are simultaneously influenced by the interactions between, and activity of, the elements they contain (left). The activity of these domains determines the higher-order organization of TAD-like structures into functionally distinct compartments inside the nucleus (right).

between the α-globin promoters and enhancers are significantly increased. The boundaries of this domain are formed by predominantly convergent CTCF-binding sites. These become closely apposed specifically in erythroid cells, while no interactions between convergent CTCF sites are observed in mES cells. This shows that this self-interacting domain is not preformed, but becomes established in a cell type-specific manner. It has recently been shown that these CTCF/cohesin boundaries constrain the activity of the α-globin enhancers to the self-interacting domain (35).

## Conclusions

Early studies of nuclear organization of the genome at relatively low resolution showed the existence of self-interacting domains, such as TADs, which were initially interpreted to be evolutionarily conserved and invariant between cell types. More recent analysis at higher resolution suggests that these domains may represent a fractal structure, with many domains contained within previously defined TADs and compartments. In addition, some of these domains appear to be tissue-specific.

Current data are consistent with at least two processes contributing to the observed nuclear organization. In some areas of the genome, preformed self-interacting domains may exist, which are formed by processes that are not fully understood, but likely depend on CTCF and cohesin. Processes related to gene expression within these domains impose further organization on the genome in a cell type-specific manner (Fig. 2).

If the current model is correct, there should be more examples of human genetic diseases that result from perturbations of the formation of self-interacting domains rather than from mutations of genes and their promoters or enhancers. Once the *cis-* and *trans*-acting factors responsible for establishing these

domains have been identified, it should be possible to mutate these precisely and rigorously to demonstrate their role in nuclear organization and gene regulation.

To further understand the origins and effects of self-interacting domains, we require more high-resolution analyses, of individual examples, such as the α-globin cluster discussed here, where interactions between the constituent elements can be related to processes leading to gene expression.

## References

1. Bickmore, W.A. and van Steensel, B. (2013) Genome architecture: domain organization of interphase chromosomes. *Cell*, **152**, 1270–1284.
2. Gibcus, J.H. and Dekker, J. (2013) The hierarchy of the 3D genome. *Molecular Cell*, **49**, 773–782.

3. Cremer, T. and Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.*, **2**, 292–301.

4. Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002) Capturing chromosome conformation. *Science*, **295**, 1306–1311.

5. Davies, J.O.J., Oudelaar, A.M., Higgs, D.R. and Hughes, J.R. (2017) How best to identify chromosomal interactions: a comparison of approaches. *Nature Methods*, **14**, 125–134.

6. 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.

7. 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*, **485**, 381–385.

8. Forcato, M., Nicoletti, C., Pal, K., Livi, C.M., Ferrari, F. and Bicciato, S. (2017) Comparison of computational methods for Hi-C data analysis. *Nature Methods*, **14**, 679–685.

9. Dixon, J.R., Gorkin, D.U. and Ren, B. (2016) Chromatin domains: the unit of chromosome organization. *Molecular Cell*, **62**, 668–680.

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

11. 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.

12. Rao, S.S.P., 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.

13. Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W.M., 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.

14. Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y. *et al.* (2013) Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*, **153**, 1281–1295.

15. Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schujiers, J., Lee, T.I., Zhao, K. *et al.* (2014) Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*, **159**, 374–387.

16. Schmitt, A.D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C.L., Li, Y., Lin, S., Lin, Y., Barr, C.L. *et al.* (2016) A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.*, **17**, 2042–2059.

17. Gómez-Marín, C., Tena, J.J., Acemel, R.D., López-Mayorga, M., Naranjo, S., de la Calle-Mustienes, E., Maeso, I., Beccari, L., Aneas, I., Vielmas, E. *et al.* (2015) Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proc. Natl. Acad. Sci. U.S.A*, **112**, 7542–7547.

18. Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A. and Hadjur, S. (2015) Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.*, **10**, 1297–1309.

19. 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.

20. 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. U.S.A*, **112**, E6456–E6465.

21. de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Verstegen, M.J.A.M., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H.L. and de Laat, W. (2015) CTCF binding polarity determines chromatin looping. *Mol. Cell*, **60**, 676–684.

22. 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.

23. Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T. *et al.* (2008) Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell*, **132**, 422–433.

24. Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T. *et al.* (2008) Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*, **451**, 796–801.

25. Rubio, E.D., Reiss, D.J., Welcsh, P.L., Disteche, C.M., Filippova, G.N., Baliga, N.S., Aebersold, R., Ranish, J.A. and Krumm, A. (2008) CTCF physically links cohesin to chromatin. *Proc. Natl Acad. Sci. USA*, **105**, 8309–8314.

26. Stedman, W., Kang, H., Lin, S., Kissil, J.L., Bartolomei, M.S. and Lieberman, P.M. (2008) Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *embo J*, **27**, 654–666.

27. Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.G. and Merkenschlager, M. (2009) Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*, **74**, 595–413.

28. Xiao, T., Wallace, J. and Felsenfeld, G. (2011) Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol. Cell. Biol.*, **31**, 2174–2183.

29. Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O. and Reinberg, D. (2015) CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science*, **347**, 1017–1021.

30. Narendra, V., Bulajić, M., Dekker, J., Mazzoni, E.O. and Reinberg, D. (2016) CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes Dev.*, **30**, 2657–2662.

31. Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R. *et al.* (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, **161**, 1012–1025.

32. Lee, H.K., Willi, M., Wang, C., Yang, C.M., Smith, H.E., Liu, C. and Hennighausen, L. (2017) Functional assessment of CTCF sites at cytokine-sensing mammary enhancers using CRISPR/Cas9 gene editing in mice. *Nucleic Acids Res.*, **45**, 4606–4618.

33. Symmons, O., Pan, L., Remeseiro, S., Aktas, T., Klein, F., Huber, W. and Spitz, F. (2016) The Shh topological domain facilitates the action of remote enhancers by reducing the effects of genomic distances. *Dev. Cell*, **39**, 529–543.

34. Franke, M., Ibrahim, D.M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., Kraft, K., Kempfer, R., Jerković, I., Chan, W.-L. *et al.* (2016) Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature*, **538**, 265–269.

35. Hanssen, L.L.P., Kassouf, M.T., Oudelaar, A.M., Biggs, D., Preece, C., Downes, D.J., Gosden, M., Sharpe, J.A., Sloane-Stanley, J.A., Hughes, J.R. *et al.* (2017) Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nat. Cell Biol.*, **13**, 74.

36. Faure, A.J., Schmidt, D., Watt, S., Schwalie, P.C., Wilson, M.D., Xu, H., Ramsay, R.G., Odom, D.T. and Flicek, P. (2012) Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. *Genome Res.*, **22**, 2163–2175.

37. Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S. *et al.* (2010) Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, **467**, 430–435.

38. Schmidt, D., Schwalie, P.C., Ross-Innes, C.S., Hurtado, A., Brown, G.D., Carroll, J.S., Flicek, P. and Odom, D.T. (2010) A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.*, **20**, 578–588.

39. Busslinger, G.A., Stocsits, R.R., van der Lelij, P., Axelsson, E., Tedeschi, A., Galjart, N. and Peters, J.-M. (2017) Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature*, **544**, 503–507.

40. Lower, K.M., Hughes, J.R., De Gobbi, M., Henderson, S., Viprakasit, V., Fisher, C., Goriely, A., Ayyub, H., Sloane-Stanley, J., Vernimmen, D. *et al.* (2009) Adventitious changes in long-range gene expression caused by polymorphic structural variation and promoter competition. *Proc. Natl Acad. Sci. USA*, **106**, 21771–21776.

41. Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A. and Fraser, P. (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, **3**, 502, 59–64.

42. Stevens, T.J., Lando, D., Basu, S., Atkinson, L.P., Cao, Y., Lee, S.F., Leeb, M., Wohlfahrt, K.J., Boucher, W., O'Shaughnessy-Kirwan, A. *et al.* (2017) 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature*, **544**, 59–64.

43. 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.

44. Lonfat, N., Montavon, T., Darbellay, F., Gitto, S. and Duboule, D. (2014) Convergent evolution of complex regulatory landscapes and pleiotropy at Hox loci. *Science*, **346**, 1004–1006.

45. Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W. and Furlong, E.E.M. (2014) Enhancer loops appear stable during development and are associated with paused polymerase. *Nature*, **513**, 89–100.

46. Anguita, E., Hughes, J., Heyworth, C., Blobel, G.A., Wood, W.G. and Higgs, D.R. (2004) Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *embo J.*, **23**, 2841–2852.

47. Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A. and Bruneau, B.G. (2017) Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell*, **169**, 930–944.e22.

48. Zuin, J., Dixon, J.R., van der Reijden, M.I.J.A., Ye, Z., Kolovos, P., Brouwer, R.W.W., van de Corput, M.P.C., van de Werken, H.J.G., Knoch, T.A., van IJcken, W.F.J. *et al.* (2014) Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl Acad. Sci. USA*, **111**, 996–1001.

49. Seitan, V.C., Faure, A.J., Zhan, Y., McCord, R.P., Lajoie, B.R., Ing-Simmons, E., Lenhard, B., Giorgetti, L., Heard, E., Fisher, A.G. *et al.* (2013) Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Res.*, **23**, 2066–2077.

50. Sofueva, S., Yaffe, E., Chan, W.-C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S.M., Schroth, G.P., Tanay, A. and Hadjur, S. (2013) Cohesin-mediated interactions organize chromosomal domain architecture. *embo J.*, **32**, 3119–3129.

51. Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C., Mirny, L. *et al.* (2016) Two independent modes of chromosome organization are revealed by cohesin removal. 10.1101/094185.

52. Barbieri, M., Xie, S.Q., Torlai Triglia, E., Chiariello, A.M., Bianco, S., de Santiago, I., Branco, M.R., Rueda, D., Nicodemi, M. and Pombo, A. (2017) Active and poised promoter states drive folding of the extended HoxB locus in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.*, **109**, 1427.

53. Brackley, C.A., Brown, J.M., Waithe, D., Babbs, C., Davies, J., Hughes, J.R., Buckle, V.J. and Marenduzzo, D. (2016) Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models. *Genome Biol.*, **17**, 59.

54. Davies, J.O.J., Telenius, J.M., McGowan, S.J., Roberts, N.A., Taylor, S., Higgs, D.R. and Hughes, J.R. (2015) Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat. Meth.*, **86**, 1202–1210.

55. Sanyal, A., Lajoie, B.R., Jain, G. and Dekker, J. (2012) The long-range interaction landscape of gene promoters. *Nature*, **489**, 109–113.

56. Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V. *et al.* (2012) A map of the cis-regulatory sequences in the mouse genome. *Nature*, **488**, 116–120.

57. Ruf, S., Symmons, O., Uslu, V.V., Dolle, D., Hot, C., Ettwiller, L. and Spitz, F. (2011) Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat. Genet.*, **43**, 379–386.

58. Symmons, O., Uslu, V.V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L. and Spitz, F. (2014) Functional and topological characteristics of mammalian regulatory domains. *Genome Res.*, **24**, 390–400.

59. Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F. and de Laat, W. (2002) Looping and interaction between hypersensitive sites in the active β-globin locus. *Molecular Cell*, **10**, 1453–1465.

60. Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R. and Higgs, D.R. (2014) Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.*, **46**, 205–212.

61. Hug, C.B., Grimaldi, A.G., Kruse, K. and Vaquerizas, J.M. (2017) Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell*, **169**, 216–228.e19.

62. Higgs, D.R., Vernimmen, D. and Wood, B. (2008) Long-range regulation of alpha-globin gene expression. *Adv. Genet.*, **61**, 143–173.

63. Noordermeer, D. and de Laat, W. (2008) Joining the loops: beta-globin gene regulation. *IUBMB Life*, **60**, 824–833.

64. Selvaraj, S., Dixon, J.R., Bansal, V. and Ren, B. (2013) Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nat. Biotechnol.*, **31**, 1111–1118.

65. Hsu, S.C., Gilgenast, T.G., Bartman, C.R., Edwards, C.R., Stonestrom, A.J., Huang, P., Emerson, D.J., Evans, P., Werner, M.T., Keller, C.A. *et al.* (2017) The BET protein BRD2 cooperates with CTCF to enforce transcriptional and architectural boundaries. *Molecular Cell*, **66**, 102–116.e7.

66. Wang, Y., Zhang, B., Zhang, L., An, L., Xu, J., Li, D., Choudhary, M.N., Li, Y., Hu, M., Hardison, R. *et al.* (2017) The 3D genome browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *bioRxiv*, 10.1101/112268.