

TADs and Their Borders: Free Movement or Building a Wall?

Li-Hsin Chang, Sourav Ghosh, Daan Noordermeer

▶ **To cite this version:**

Li-Hsin Chang, Sourav Ghosh, Daan Noordermeer. TADs and Their Borders: Free Movement or Building a Wall?. Journal of Molecular Biology, 2020, 432 (3), pp.643-652. 10.1016/j.jmb.2019.11.025 . hal-02995869

HAL Id: hal-02995869
https://hal.science/hal-02995869

Submitted on 19 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TADs and their borders: free movement or building a wall?

Li-Hsin Chang[1], Sourav Ghosh[1], Daan Noordermeer[1†]

Affiliation:

[1] Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, University Paris-sud, University Paris-Saclay, Gif-sur-Yvette, France

[†] Correspondence: daan.noordermeer@i2bc.paris-saclay.fr

Key words: TADs; Topologically Associating Domains; CTCF; loop extrusion; Hi-C

Abstract

The tridimensional (3D) organization of mammalian genomes combines structures from different length scales. Within this organization, Topologically Associating Domains (TADs) are visible in Hi-C heat maps at the sub-megabase scale. The integrity of TADs is important for correct gene expression, but in a context-dependent and variable manner. The correct structure and function of TADs require the binding of the CTCF protein at both borders, which appears to block an active and dynamic mechanism of 'Cohesin-mediated loop extrusion'. As a result, mammalian TADs appear as so-called 'loop domains' in Hi-C data, which are the focus of this review. Moreover, we present a reanalysis of TADs from three "golden-standard" mammalian Hi-C data sets. Despite the prominent presence of TADs in Hi-C heat maps from all studies, we find consistently that regions within these domains are only moderately insulated from their surroundings. Moreover, single-cell Hi-C and super-resolution microscopy have revealed that the structure of TADs and the position of their borders can vary from cell-to-cell. The function of TADs as units of gene regulation may thus require additional aspects, potentially incorporating the mechanism of loop extrusion as well. Recent developments in single-cell and multi-contact genomics and super-resolution microscopy assays will be instrumental to link TAD formation and structure to their function in transcriptional regulation.

TADs as functional units of gene regulation

A major biophysical challenge for cells is to compact their large genome—up to 2 meters for a mammalian cell—into a nucleus that is only a few microns in diameter, all while ensuring both proper genome function and the programmable flexibility required for cellular differentiation [1]. The first order of DNA compaction, the wrapping of 147 bp of DNA around a histone octamer core to form the nucleosome (‘10 nm fiber’), is well agreed upon [2-4]. Subsequent higher-order chromatin organization is formed by more diverse combinations of DNA structures, ranging from chromatin loops and contact domains at the smallest scales to chromosome territories that occupy large volumes within interphase nuclei [5, 6].

The development of Chromosome Conformation Capture (3C) and genomics-based derivatives have greatly advanced our capacity to explore structural aspects of 3D genome architecture [7, 8]. Particularly Hi-C has proven to be instrumental in our understanding of 3D genome organization, by providing information on all possible genome-wide interactions within a single experiment [9]. When analyzed at sufficient resolution (typically below 100 kb), Hi-C experiments reveal the existence of distinct structural domains [10]. These domains, referred to as Topologically Associating Domains (TADs), are characterized by increased intra-domain interactions and insulation from their surroundings, and thus appear as ‘pyramids’ in Hi-C heat maps [10, 11] (Fig. 1A). Whereas TADs in mammalian cells are generally identified with median sizes just below one megabase, in other organisms they may manifest themselves at considerably different sizes [10-16]. Moreover, studies with increasingly higher resolution have suggested the presence of additional finer-scale levels of sub-domains (‘sub-TADs’), nested within mammalian TADs [17-20].

Initial correlative studies revealed enrichment within TADs of co-regulated gene clusters, enhancer-promoter pairs and replication domains [10, 11, 21, 22]. Moreover, a large fraction of the borders that surround mammalian TADs bind the CTCF protein (CCCTC-binding factor), the only somatic insulator protein in somatic cells [10, 18] (Fig. 1A). Until now, CTCF binding at large subsets of TAD borders has only been reported in vertebrates (e.g. [10, 15]). Other TAD borders can be linked to (at least) two other biological phenomena. First, the presence of highly expressed genes can introduce a strong insulation between surrounding regions in the Hi-C heat map of a diverse range of organisms [10, 23-25]. Second, a sharp separation may be observed between regions bound by histone modifications associated with opposing transcriptional and regulatory output [12, 18, 26]. This latter

separation, in mammals usually at the smaller sub-domain levels, appears not to directly involve the borders though, but rather to be caused by the self-aggregation of regions that carry the same epigenetic marks due to phase separation, thereby forming so-called 'contact domains' or 'compartmental domains' ([5, 18, 27-29], see also [30] in this JMB issue for further discussion). Unless mentioned otherwise, in the remainder of this review we focus on TADs where CTCF is bound at the borders.

Why CTCF is present at a large fraction of mammalian TAD borders, and if it is linked to transcriptional regulation, has been addressed in various pathologies and developmental contexts (see Table 1 for summary). The CTCF insulator protein was previously described as an 'enhancer blocker' that precludes transcriptional activation when bound in-between an enhancer-promoter pair [31]. Lupiáñez and colleagues showed that a group of congenital hand malformations could be traced to deletions, inversions and duplications around the *EPHA4* gene that had in common that they created chromosomal configurations where CTCF-bound sites between the enhancers of the *EPHA4* gene and neighboring inactive genes were removed [32]. In turn, ectopic enhancer-promoter interactions could be formed that activated these neighboring genes, which in turn caused the various hand malformations [32]. Similarly, in Cooks syndrome, a duplication that spanned parts of two TADs and CTCF-bound border in between, resulted in the formation of a new TAD ('neo-TAD') where the genes encountered a reorganized regulatory landscape that imposed aberrant activation [33]. In further analogy, mutations or epi-mutations of CTCF sites at TAD borders resulted in ectopic enhancer-promoter interactions and inappropriate gene activation in cancer cells as well [34, 35]. Combined, these studies suggest that the enhancer blocking function of CTCF is achieved by positioning genes and non-relevant enhancers in separate TADs.

More recent studies have provided nuance to the impact of perturbed CTCF binding at TAD borders. At the *Shh* locus, the limb-specific ZRS enhancer is located near the TAD border, where it is surrounded by multiple CTCF sites. Deletion of CTCF sites, either one or multiple, reduced *Shh* promoter-ZRS enhancer contacts but only had moderate influence on *Shh* gene activity. Removal of multiple CTCF sites increased interactions with neighboring sequences, which coincided with a mildly increased activity of a gene that is contained within this region [36, 37]. Similarly, at the *Sox9* locus, a fusion with the neighboring TAD required the deletion of 8 CTCF sites at and near the border. Yet, upon this fusion, the effect on gene activity was relatively mild, with *Sox9* activity moderately reduced and a twofold upregulation of the lowly expressed neighboring *Kcnj2* gene [38]. Combined, these studies show that TAD borders can be composed of more complex ensembles of CTCF-bound sites.

Moreover, the effect of TAD-border perturbations on gene expression may be relatively moderate. Intriguingly, acute depletion of CTCF in embryonic stem cells resulted in similar moderate changes to transcriptional output. After two days of CTCF removal, insulation between TADs was strongly reduced, yet the transcriptional deregulation of only nine hundred genes could be assigned to TAD deregulation [39]. In a specific cell type, the contribution of individual TADs on gene expression may thus be relatively small. This effect may be related to the cell-type specificity of enhancers, which reduces the chance of inappropriate enhancer-promoter contacts within a given cell type [40, 41].

In conclusion, genes and their associated enhancers are clustered in TADs, which act as functional but variable and context-dependent units of gene regulation. Demarcation and insulation of TADs are primarily dependent on the correct location of their borders, which are strongly enriched for the binding of the CTCF insulator protein (see also Table 1).

Cohesin-dependent loop extrusion and TAD formation

TADs can be identified from a Hi-C heat map, where they appear as pyramid-shaped domains with increased intra-domain interactions (Fig. 1A). Initial TAD-calling algorithms used a 'Directionality Index' to identify bin-by-bin transitions in the asymmetry of contacts, thus indicating the switch from one insulated TAD to the next [10]. More recently, algorithms to determine 'Insulation Scores' have been developed, which determine localized minima in interaction frequency by looking at a more regional scale, thus being more robust to individual outlying bins ([42, 43]; see [30] for further considerations).

How punctuated CTCF binding at TAD borders can result in increased intra-domain contacts, remains not fully established. An important mechanistic suggestion comes from the Hi-C data itself, where many TADs with CTCF at their borders display strongly enriched signal on the top of the pyramid [18]. This punctuated signal indicates that the borders of these domains loop together, which is particularly frequently observed when CTCF is bound at both sides. These TADs, with both increased intra-domain interactions and looping of the borders, are therefore also referred to as 'loop domains' [18, 44] (Fig. 1B). This feature distinguishes these domains from conventional chromatin loops, which are often associated with enhancer-promoter contacts, and 'contact / compartmental domains', which are associated with the aggregation of domains carrying identical histone modifications [5, 12, 18, 26, 45] (Fig. 1B).

CTCF binds non-symmetric DNA motifs in the genome, where it often co-localizes with the ring shaped Cohesin complex [46-48]. Intriguingly, up to 90% of CTCF motifs that

are at the base of TAD-spanning loops are in a convergent orientation [18, 49, 50]. The importance of the correct orientation of CTCF motifs has been experimentally confirmed by artificial inversions, which result in reduced TAD insulation [49, 51].

How CTCF motif orientation guides TAD formation may be explained through the co-localization with the Cohesin complex. Based on this link, a 'Cohesin-dependent loop extrusion' model has been developed, which reproduces Hi-C data *in silico* with high degrees of similarity to experimentally obtained data [52-55]. In this model, the Cohesin complex randomly associates with the chromatin fiber, where it starts the progressive formation of a DNA loop through bi-directional extrusion (Fig. 1C). Loop extrusion is blocked, on one side only, when a CTCF molecule is encountered that is bound in the correct orientation. When Cohesin has encountered CTCF on both borders of the TAD, the loop is stabilized, as reflected by a visible loop in the Hi-C map (Fig. 1C). The visible enrichment of intra-domain interactions indicates that this process is simultaneously accompanied by more transient DNA contacts within the extruded loop, suggesting it adopts a more crumbled configuration. Together, these two patterns of 3D interactions are reflected as a loop domain in the Hi-C heat map (Fig. 1B).

Importantly, to reconstruct TADs *in silico*, the loop extrusion in the model needs to be a dynamic process. This implies that the Cohesin complex does not only continuously associate with the DNA at random positions in the genome, but also dissociates from the DNA after a certain amount of time [54]. In practice, this dissociation will mostly occur after Cohesin has co-occupied a CTCF-bound site for a certain period of time. Various experimental studies have dissected how the combined action of the Cohesin complex and CTCF influences TAD structure, thereby confirming key aspects of the Cohesin-dependent loop extrusion model. Rapid elimination of the Cohesin component Rad21/Ssc1, or of its loading factor Nipbl, results in elimination of all TADs that appear as loop domains. In contrast, contact domains appear mostly untouched or even increased in structure [44, 56-58]. These studies therefore uncouple the process of loop domain formation from that of contact domains. Loss of TADs upon mutations in the ATPase domains of Smc3, another Cohesin complex member, further confirms that loop extrusion is energy consuming, and thus an active mechanism [59]. In the same study, a subset of TADs was identified where one or both borders showed contiguous enriched contacts within the entire domain, visible as "architectural stripes". These stripes are mostly found at TAD borders that combine strong insulation and multiple instances of CTCF binding with a nearby Cohesin loading area [59]. At such borders, the Cohesin complex will be blocked instantly on one side. According to

loop extrusion model, such unidirectional loop formation is indeed predicted to manifest itself as horizontal and vertical stripes in the Hi-C matrix. Similar to Cohesin removal, the rapid removal of CTCF from TAD borders reduced the insulated nature of loop domains, whereas contact domains remained intact as well [39]. Interestingly though, the elimination of Cohesin and CTCF differentially affected genome compaction: whereas overall compaction was not affected by CTCF elimination, the loss of Cohesin association caused a global decompaction of the genome [39, 57]. This supports the model that Cohesin can create loops in the absence of CTCF, but that it needs to be blocked at specific sites in the genome to create noticeable loop domains.

An intriguing addition to the model of loop formation has been identified upon the deletion of Wapl, the protein that is required for the release of Cohesin from the DNA [58, 60]. As expected from reinforced residence of the Cohesin complex at TAD borders, intra-TAD interactions become stronger. Unexpectedly though, TADs increased in size as well, which was visible as multiple TADs fusing into a single loop domain [60]. Prolonged blocking of Cohesin at a TAD borders may therefore result in a 'read through' of loop extrusion towards the next site occupied by CTCF in the correct orientation. This phenomenon may be due to a 'traffic jam' effect, whereby multiple TAD borders cluster together [61]. In turn, this may be caused by the dynamics of CTCF binding to the DNA, whose dissociation is considerably more dynamic than that of the Cohesin complex (~1-2 minutes vs. ~20 minutes residence time) [62]. Cohesin-dependent loop extrusion may therefore not only depend on dynamic association and dissociation of the Cohesin complex, but may also involve a dynamic CTCF-binding component.

Combined, the results from *in silico* modeling and experimental studies provide accumulating evidence that TADs, as visible as loop domains in Hi-C heat maps, are shaped by a dynamic and active process that consists of a constant and energy-consuming bi-directional extrusion of DNA loops by the Cohesin complex. Within this process, CTCF functions to temporarily fix extruded chromatin loops at defined sites in the genome, which in turn creates specific insulated TADs.

TADs as structural units of genome organization: consistency or heterogeneity?

How TADs are structured and how they are involved in context-depending transcriptional regulation is increasingly understood. In contrast, important questions remain on how stable TADs are over time, if they are consistent between different cells, and what is the nature of

their internal structure. Most insights on TAD organization have been obtained from Hi-C studies on large cell populations, whose averaged interaction data preclude the interpretation of truly dynamic aspects.

Nonetheless, a more detailed appraisal of population-based Hi-C heat maps does provide valuable insights on how the chromatin is organized within and between TADs. Whereas pairing of CTCF-bound borders in a loop domain is represented by the strong signal at the top of the triangle in the Hi-C heat map, the internal organization of a TAD is described by the signal within the triangle itself. Moreover, the degree of insulation between neighboring TADs is measured by the depletion of signal between two TADs (Fig. 1A; absence of signal in the 'inversed' pyramid between TADs). Similarly, the strength of borders between TADs influences insulation as well. Strong borders, visible as sharp valleys in the Hi-C heat map, are indicative of a separation between highly discrete TADs. In contrast, the presence of a more gradual valley between two TADs indicates a border that is more transitional and dynamic in nature (Fig. 2A). Inspection of high-resolution Hi-C data reveals that many TADs are separated by zones of transition that can span a considerable distance (Fig. 2B, see below for additional characterization). TAD-calling algorithms will call discrete borders within such zones of transition, yet the actual insulation at these zones is more gradual (Fig. 2B; dashed versus solid line in Hi-C heat map). In turn, according to the population-based Hi-C data, sequences near the borders do not strictly localize within one TAD (Fig. 2B; virtual 4C tracks). Although the presence of zones of transition may be due to artifacts in the Hi-C assay, similar interaction patterns are observed in the crosslink- and ligation-free DamC assay as well [63]. A more biological explanation for the appearance of these zones of transition may be that TAD borders can be dynamic, linked both to the discontinuous nature of CTCF binding and the observed clustering of CTCF binding sites at the majority of TAD borders [62, 64, 65]. In line with the model of Cohesin-dependent loop extrusion, the actual position of a TAD border may therefore not be fixed but vary within these defined zones of transition over time and between cells.

More recently developed single-cell Hi-C methods have started to shed light into the structural organization within and between TADs [56, 66-68]. Compared to population-based Hi-C, chromatin interactions within individual cells tend to be more stochastic and to often span multiple TADs. Yet, combination of single-cell Hi-C data from multiple individual cells tends to faithfully reproduce TAD structure as observed in population-based Hi-C data [67, 68]. TADs in population-based Hi-C data may thus represent the statistically most prominent

average 3D configuration, which is built up from a variety of structures that is present within the cell population.

To get a more detailed and quantitative insight into the structural organization of TADs, we reanalyzed currently highest resolution mouse and human Hi-C data sets, both of which are commonly used as reference contact maps [13, 18]. Whereas TADs appear as (relatively) discrete pyramids in the Hi-C heat map of both data sets, a comparison of interaction frequencies between equally distant loci that are located either within or between TADs reveals only a maximum twofold degree of insulation (Fig. 2C). A similar moderate difference was previously reported for lower resolution mouse Hi-C data as well [10, 69]. A comparison of TAD structure and borders between the both mouse embryonic stem cell data sets confirms their similarities (Fig. S1A-D). The calling of these structural aspects of TAD organization is therefore highly robust, despite considerable experimental and analytical differences. Interestingly, the maximum difference between intra- and inter-TAD interaction frequencies is only observed at a considerable distance from the borders between the TADs (Fig. 2C; dashed lines at 290 and 785 kb). This distance for maximum insulation exceeds the span of many TADs and appears not reduced at smaller TADs (Fig. 2C and Fig S1E). The zones of transition that separate TADs therefore encompass both the borders and (at least partially) the domains themselves. In turn, these results suggest that the insulating capacity of TADs is relatively weak, yet this capacity scales with the size of the TAD.

Importantly though, a low level of insulation is not incompatible with the model of loop extrusion, as the extrusion process does not prevent the resulting loops in different domains from crossing borders and intermingling with each other. In contrast, it may suggest that the creation of regulatory units within TADs is not determined by the insulated nature of these domains alone. Rather, it may be that the ongoing loop extrusion that structures TADs is involved in guiding enhancer-promoter contacts as well. Indirect support for this hypothesis may be found at architectural stripes, where active (super-)enhancer elements that are in close proximity to stripe anchors have considerably elevated contacts with other regulatory elements [59]. Prevalent promoter-enhancer contacts between chromosomes in dipteran insects (including transvection) and an artificial mammalian system using an ectopic super-enhancer and its endogenous target gene have ruled out though that the formation of regulatory contacts depend on loop extrusion alone [70, 71].

Recent developments in super-resolution microscopy, particularly the combination of STORM (STochastic Optical Reconstruction Microscopy) with Oligopaint labeling [72], constitute an alternative approach to characterize TAD structure and insulation in individual

cells. Multiple studies in Drosophila cells were able to confirm the presence of TADs, albeit with a certain degree of heterogeneity and intermingling [73-75]. The small size of these domains and their overlap with specific histone modifications suggest they are more associated with contact domains, rather than the process of Cohesin-dependent loop extrusion. In mammals, larger spatially segregated nano-compartments that correspond to TADs could also be visualized in individual cells [76]. Similar to Hi-C data, an enrichment of intra-TAD interactions could be detected, yet with relatively little specificity. Moreover, frequent contacts between TADs were detected as well, but with a reduced consistency between cells. Similar patterns of moderately enriched, 2- to 3-fold, intra-TAD interactions were reported in a study that mapped large number of pair-wise interactions at a genome-wide scale [77].

Interestingly, the depletion of Cohesin in these single-cell microscopy studies did not result in a loss of TAD-like structures in individual cells, which contradicts published population-based Hi-C data [44, 57, 76]. In contrast, the preferential positioning of boundaries at CTCF bound sites was lost, which is more in line with single-cell Hi-C studies [67, 68]. The overall 3D organization of TADs may therefore be shaped by additional nested structures, including smaller loop domains that are linked to epigenetic state.

Conclusions and Perspectives

In this review, we have discussed the current evidence for the existence of TADs, how they are structured and how they are involved in transcriptional regulation. Whereas their presence can be unambiguously detected in population-based Hi-C heat maps, both quantitative interpretation of Hi-C data and the comparison of single-cell Hi-C and microscopy data suggest a certain degree of heterogeneity in TAD structure. As such, these data imply that TADs are dynamic structures, both regarding their internal structure and the position of their borders, with a relatively low degree of insulation from their surroundings. How this lack of structural consistency can be reconciled with their function as regulatory units in the genome remains to be determined. Here, we hypothesize that the setting of transcriptional programs may incorporate the process of TAD formation through loop extrusion as well.

To unambiguously link TAD formation and structure to transcriptional regulation, we will need to comprehensively address these aspects at highest precision in individual cells. Particularly instrumental will be the recent developments in super-resolution microscopy, including the combined visualization of DNA conformation and RNA [78, 79], and the development of genomics-based approaches that can identify and quantify ensembles of chromatin contacts. Recently, several technologies have emerged that either allow genome-

wide non-targeted characterization of chromatin hubs (C-walks [80], GAM (Genome Architecture Mapping) [81], SPRITE [82], and ChIA-Drop [83]) or the quantitative measurement of three-way interactions at defined sites (Multi-Contact 4C [61] and Tri-C [84]). The use of these technologies, and potential further developments, together with advanced modeling approaches and cellular models with modified aspects of TAD structure, will provide the tools to comprehensively unravel the dynamic nature of TADs.

Methods

Reanalysis of published Hi-C data

Unprocessed sequencing data from published Hi-C studies was obtained from the GEO data repository: human GM12878 B-lymphoblastoid cells: GSE63525 [18] and mouse embryonic stem cells: GSE96107 [13]. Data was processed at 5 kb resolution using the HiC-Pro tool, using standard settings and ICE-normalization of the resulting matrices [85]. TADs and their borders were called using the TADtool from the 5 kb Hi-C matrices based on insulation score [43]. Lower resolution Hi-C data [10], including normalized matrices and positions of TADs, were downloaded from http://chromosome.sdsc.edu/mouse/hi-c/download.html. For the comparison of TAD borders in mouse embryonic stem cell Hi-C data (Fig. S1C), data from [10] was converted to mm10 using the LiftOver tool (http://genome.ucsc.edu).

Autosomal read pairs spanning 5 kb – 2 Mb were filtered for presence within TADs, followed by sorting for being within the same TAD or in different TADs. Median interaction signal for all read pairs was calculated based on category and distance.

Acknowledgments

We thank the members of the Noordermeer lab for useful discussion. L.C. and S.G. are supported by funding from the Agence Nationale de Recherche to D.N. (grants ANR-14-ACHN-0009-01, ANR-16-TERC-0027-01 and ANR-17-CE12-0001-02).

Declarations of interest

None

Author contributions

S.G. reanalyzed published Hi-C data. L.-H.C. and D.N. wrote the manuscript with input from S.G.

References

[1] Mozziconacci J, Merle M, Lesne A. The 3D genome encodes gene regulation during development. J Mol Biol. This issue (2019).

[2] Khorasanizadeh S. The nucleosome: from genomic organization to genomic regulation. Cell. 116 (2004) 259-272.

[3] Schalch T, Duda S, Sargent DF, Richmond TJ. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature. 436 (2005) 138-141.

[4] Fussner E, Strauss M, Djuric U, Li R, Ahmed K, Hart M, et al. Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. EMBO Rep. 13 (2012) 992-996.

[5] Rowley MJ, Corces VG. Organizational principles of 3D genome architecture. Nat Rev Genet. 19 (2018) 789-800.

[6] Stadhouders R, Filion GJ, Graf T. Transcription factors and 3D genome conformation in cell-fate decisions. Nature. 569 (2019) 345-354.

[7] Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 295 (2002) 1306-1311.

[8] Davies JO, Oudelaar AM, Higgs DR, Hughes JR. How best to identify chromosomal interactions: a comparison of approaches. Nat Methods. 14 (2017) 125-134.

[9] Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 326 (2009) 289-293.

[10] Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 485 (2012) 376-380.

[11] Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 485 (2012) 381-385.

[12] Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, et al. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell. 148 (2012) 458-472.

[13] Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, et al. Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell. 171 (2017) 557-572 e524.

[14] Liu C, Cheng YJ, Wang JW, Weigel D. Prominent topologically associated domains differentiate global chromatin packing in rice from Arabidopsis. Nat Plants. 3 (2017) 742-748.
[15] Kaaij LJT, van der Weide RH, Ketting RF, de Wit E. Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development. Cell Rep. 24 (2018) 1-10 e14.
[16] Fishman V, Battulin N, Nuriddinov M, Maslova A, Zlotina A, Strunov A, et al. 3D organization of chicken genome demonstrates evolutionary conservation of topologically associated domains and highlights unique architecture of erythrocytes' chromatin. Nucleic Acids Res. 47 (2019) 648-665.
[17] Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell. 153 (2013) 1281-1295.
[18] Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 159 (2014) 1665-1680.
[19] Hsieh T-HS, Slobodyanyuk E, Hansen AS, Cattoglio C, Rando OJ, Tjian R, et al. Resolving the 3D landscape of transcription-linked mammalian chromatin folding. bioRxiv. (2019) 638775.
[20] Krietenstein N, Abraham S, Venev SV, Abdennur N, Gibcus JH, Hsieh T-HS, et al. Ultrastructural details of mammalian chromosome architecture. bioRxiv. (2019) 639922.
[21] Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature. 489 (2012) 109-113.
[22] Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, et al. A map of the cis-regulatory sequences in the mouse genome. Nature. 488 (2012) 116-120.
[23] Hou C, Li L, Qin ZS, Corces VG. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. Mol Cell. 48 (2012) 471-484.
[24] Le TB, Imakaev MV, Mirny LA, Laub MT. High-resolution mapping of the spatial organization of a bacterial chromosome. Science. 342 (2013) 731-734.
[25] Eser U, Chandler-Brown D, Ay F, Straight AF, Duan Z, Noble WS, et al. Form and function of topologically associating genomic domains in budding yeast. Proc Natl Acad Sci U S A. 114 (2017) E3061-E3070.
[26] Noordermeer D, Leleu M, Splinter E, Rougemont J, De Laat W, Duboule D. The dynamic architecture of Hox gene clusters. Science. 334 (2011) 222-225.

[27] Larson AG, Elnatan D, Keenen MM, Trnka MJ, Johnston JB, Burlingame AL, et al. Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. Nature. 547 (2017) 236-240.

[28] Strom AR, Emelyanov AV, Mir M, Fyodorov DV, Darzacq X, Karpen GH. Phase separation drives heterochromatin domain formation. Nature. 547 (2017) 241-245.

[29] Tatavosian R, Kent S, Brown K, Yao T, Duc HN, Huynh TN, et al. Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation. J Biol Chem. 294 (2019) 1451-1463.

[30] de Wit E. TADs as the caller calls them. J Mol Biol. This issue (2019).

[31] Merkenschlager M, Nora EP. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. Annu Rev Genomics Hum Genet. 17 (2016) 17-43.

[32] Lupianez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 161 (2015) 1012-1025.

[33] Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schopflin R, et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature. 538 (2016) 265-269.

[34] Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature. 529 (2016) 110-114.

[35] Hnisz D, Weintraub AS, Day DS, Valton AL, Bak RO, Li CH, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science. 351 (2016) 1454-1458.

[36] Paliou C, Guckelberger P, Schopflin R, Heinrich V, Esposito A, Chiariello AM, et al. Preformed chromatin topology assists transcriptional robustness of Shh during limb development. Proc Natl Acad Sci U S A. (2019).

[37] Williamson I, Kane L, Devenney PS, Anderson E, Kilanowski F, Hill RE, et al. Developmentally regulated Shh expression is robust to TAD perturbations. bioRxiv. (2019) 609941.

[38] Despang A, Schopflin R, Franke M, Ali S, Jerkovic I, Paliou C, et al. Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. Nat Genet. 51 (2019) 1263-1271.

[39] Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, et al. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell. 169 (2017) 930-944 e922.

[40] Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 459 (2009) 108-112.

[41] Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature. 457 (2009) 854-858.

[42] Crane E, Bian Q, McCord RP, Lajoie BR, Wheeler BS, Ralston EJ, et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature. 523 (2015) 240-244.

[43] Kruse K, Hug CB, Hernandez-Rodriguez B, Vaquerizas JM. TADtool: visual parameter identification for TAD-calling algorithms. Bioinformatics. 32 (2016) 3190-3192.

[44] Rao SSP, Huang SC, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon KR, et al. Cohesin Loss Eliminates All Loop Domains. Cell. 171 (2017) 305-320 e324.

[45] Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell. 10 (2002) 1453-1465.

[46] Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell. 128 (2007) 1231-1245.

[47] Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, et al. Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell. 132 (2008) 422-433.

[48] Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature. 451 (2008) 796-801.

[49] de Wit E, Vos ES, Holwerda SJ, Valdes-Quezada C, Verstegen MJ, Teunissen H, et al. CTCF Binding Polarity Determines Chromatin Looping. Mol Cell. 60 (2015) 676-684.

[50] Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, et al. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell. 163 (2015) 1611-1627.

[51] Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, et al. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell. 162 (2015) 900-910.

[52] Alipour E, Marko JF. Self-organization of domain structures by DNA-loop-extruding enzymes. Nucleic Acids Res. 40 (2012) 11202-11212.

[53] Sanborn AL, Rao SS, Huang SC, Durand NC, Huntley MH, Jewett AI, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A. 112 (2015) E6456-6465.

[54] Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. 15 (2016) 2038-2049.

[55] Nuebler J, Fudenberg G, Imakaev M, Abdennur N, Mirny LA. Chromatin organization by an interplay of loop extrusion and compartmental segregation. Proc Natl Acad Sci U S A. 115 (2018) E6697-E6706.

[56] Gassler J, Brandao HB, Imakaev M, Flyamer IM, Ladstatter S, Bickmore WA, et al. A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. EMBO J. 36 (2017) 3600-3618.

[57] Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, et al. Two independent modes of chromatin organization revealed by cohesin removal. Nature. 551 (2017) 51-56.

[58] Wutz G, Varnai C, Nagasaka K, Cisneros DA, Stocsits RR, Tang W, et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. EMBO J. 36 (2017) 3573-3599.

[59] Vian L, Pekowska A, Rao SSP, Kieffer-Kwon KR, Jung S, Baranello L, et al. The Energetics and Physiological Impact of Cohesin Extrusion. Cell. 173 (2018) 1165-1178 e1120.

[60] Haarhuis JHI, van der Weide RH, Blomen VA, Yanez-Cuna JO, Amendola M, van Ruiten MS, et al. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. Cell. 169 (2017) 693-707 e614.

[61] Allahyar A, Vermeulen C, Bouwman BAM, Krijger PHL, Verstegen M, Geeven G, et al. Enhancer hubs and loop collisions identified from single-allele topologies. Nat Genet. 50 (2018) 1151-1160.

[62] Hansen AS, Pustova I, Cattoglio C, Tjian R, Darzacq X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. Elife. 6 (2017).

[63] Redolfi J, Zhan Y, Valdes-Quezada C, Kryzhanovska M, Guerreiro I, Iesmantavicius V, et al. DamC reveals principles of chromatin folding in vivo without crosslinking and ligation. Nat Struct Mol Biol. 26 (2019) 471-480.

[64] Madani Tonekaboni SA, Mazrooei P, Kofia V, Haibe-Kains B, Lupien M. CREAM: Clustering of genomic REgions Analysis Method. bioRxiv. (2018) 222562.

[65] Kentepozidou E, Aitken SJ, Feig C, Stefflova K, Ibarra-Soria X, Odom DT, et al. Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. bioRxiv. (2019) 668855.
[66] Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature. 502 (2013) 59-64.
[67] Flyamer IM, Gassler J, Imakaev M, Brandao HB, Ulianov SV, Abdennur N, et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. Nature. 544 (2017) 110-114.
[68] Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature. 544 (2017) 59-64.
[69] Dekker J, Mirny L. The 3D Genome as Moderator of Chromosomal Communication. Cell. 164 (2016) 1110-1121.
[70] Pirrotta V. Transvection and chromosomal trans-interaction effects. Biochim Biophys Acta. 1424 (1999) M1-8.
[71] Noordermeer D, de Wit E, Klous P, van de Werken H, Simonis M, Lopez-Jones M, et al. Variegated gene expression caused by cell-specific long-range DNA interactions. Nat Cell Biol. 13 (2011) 944-951.
[72] Beliveau BJ, Boettiger AN, Avendano MS, Jungmann R, McCole RB, Joyce EF, et al. Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. Nat Commun. 6 (2015) 7147.
[73] Boettiger AN, Bintu B, Moffitt JR, Wang S, Beliveau BJ, Fudenberg G, et al. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature. 529 (2016) 418-422.
[74] Cattoni DI, Cardozo Gizzi AM, Georgieva M, Di Stefano M, Valeri A, Chamousset D, et al. Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. Nat Commun. 8 (2017) 1753.
[75] Szabo Q, Jost D, Chang JM, Cattoni DI, Papadopoulos GL, Bonev B, et al. TADs are 3D structural units of higher-order chromosome organization in Drosophila. Sci Adv. 4 (2018) eaar8082.
[76] Bintu B, Mateo LJ, Su JH, Sinnott-Armstrong NA, Parker M, Kinrot S, et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science. 362 (2018).

[77] Finn EH, Pegoraro G, Brandao HB, Valton AL, Oomen ME, Dekker J, et al. Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. Cell. 176 (2019) 1502-1515 e1510.

[78] Cardozo Gizzi AM, Cattoni DI, Fiche J-B, Espinola SM, Gurgo J, Messina O, et al. Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. Molecular Cell. 74 (2019) 212-222.e215.

[79] Mateo LJ, Murphy SE, Hafner A, Cinquini IS, Walker CA, Boettiger AN. Visualizing DNA folding and RNA in embryos at single-cell resolution. Nature. 568 (2019) 49-54.

[80] Olivares-Chauvet P, Mukamel Z, Lifshitz A, Schwartzman O, Elkayam NO, Lubling Y, et al. Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. Nature. 540 (2016) 296-300.

[81] Beagrie RA, Scialdone A, Schueler M, Kraemer DC, Chotalia M, Xie SQ, et al. Complex multi-enhancer contacts captured by genome architecture mapping. Nature. 543 (2017) 519-524.

[82] Quinodoz SA, Ollikainen N, Tabak B, Palla A, Schmidt JM, Detmar E, et al. Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. Cell. 174 (2018) 744-757 e724.

[83] Zheng M, Tian SZ, Capurso D, Kim M, Maurya R, Lee B, et al. Multiplex chromatin interactions with single-molecule precision. Nature. 566 (2019) 558-562.

[84] Oudelaar AM, Davies JOJ, Hanssen LLP, Telenius JM, Schwessinger R, Liu Y, et al. Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. Nat Genet. 50 (2018) 1744-1751.

[85] Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. 16 (2015) 259.

[86] Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu M, et al. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. Science. 340 (2013) 1195.

[87] Sun JH, Zhou L, Emerson DJ, Phyo SA, Titus KR, Gong W, et al. Disease-Associated Short Tandem Repeats Co-localize with Chromatin Domain Boundaries. Cell. 175 (2018) 224-238 e215.

Figure and table legends

Figure 1. TADs and loop extrusion.

A. Schematic of two TADs in a Hi-C heat map, with the arrow indicating the border between the TADs. Simulated CTCF binding is depicted below, with convergently oriented CTCF peaks at the borders of both TADs (orientation of CTCF sites shown as blue and red triangles).

B. Appearance of different types of chromatin structures in a Hi-C heat map. Left: cartoon of 3D chromatin structure; right: corresponding simplified Hi-C heat maps.

C. Model for Cohesin-dependent loop extrusion. After random association of the Cohesin complex with the chromosome (**1**, **2**), a loop will start to bi-directionally extrude (**2**, **3**). Extrusion on either side will continue until a correctly oriented CTCF-bound site is encountered (left blue triangle; **4**). A complete block is only achieved when both sides have encountered convergently orientated CTCFs (blue and red triangles; **5**).

Figure 2. The insulated nature of TADs and their borders.

A. Schematic of strong and weak TAD borders. Weak TAD borders are separated by a 'zone of transition' that appears as a gradual valley between two TADs.

B. Reanalyzed Hi-C data showing a nearly 100 kb 'zone of transition' between two TADs in mouse embryonic stem cells [13]. Top: Hi-C heat map with the algorithmically called TAD border indicated with dotted lines, and the more gradual transition between the TADs with a solid line. Middle: virtual 4C signal for two bins at either side of the zone of transition [12]. Bottom: published CTCF ChIP-seq signal in mouse embryonic stem cells [22].

C. Genome-wide assessment of TAD insulation in human GM12878 B-lymphoblastoid cells [18] (red) and mouse embryonic stem cells [13] (blue) using reanalyzed Hi-C data at 5 kb resolution. Left: median genome-wide interaction counts for intra-TAD (solid lines) and inter-TAD (dashed lines) read-pairs at distances from 5 kb to 2 Mb. Right: enrichment of intra-TAD signal at different distance scales. Vertical dotted lines indicate the distance between read pairs with maximum signal enrichment (human B-lymphoblastoid cells: 290 kb – mouse embryonic stem cells: 785 kb).

Table 1. Key studies showing the influence of TAD borders on gene expression

Supplemental figure legend

Figure S1. Comparison of reanalyzed Hi-C data in mouse embryonic stem cells

A. Example of published and reanalyzed Hi-C data from two different studies in mouse embryonic stem cells. Top: 40 kb resolution Hi-C data from Dixon *et al.* [10]; bottom: 10 kb resolution Hi-C data from Bonev *et al.* [13]. Despite biological, experimental and analytical differences (left), an overall similar framework of TADs is called. Differences between both data sets are indicated above (see also panel **B.**). 10 Mb region was selected for the presence of multiple types of differences between data sets.

B. Differences in TAD calling between two different studies in mouse embryonic stem cells. Except for a minor fraction of TADs (category "Disagreement spanning multiple TADs", making up 2% of TADs in Dixon *et al.*), differences can be traced back to TADs or individual borders that are called in only one of two data sets (see panel **A.** for examples of categories).

C. Differences in TAD border calling between two different studies in mouse embryonic stem cells. Two-thirds of TAD borders that can be unambiguously linked between data sets map within 80 kb, despite the initial difference in resolution (40 kb *vs.* 5 kb resolution) and mapping to different versions of the mouse genome (mm10 *vs.* mm9) for both data sets.

D. Genome-wide assessment of TAD insulation between two different studies in mouse embryonic stem cells. In both data sets, a comparably moderate degree of insulation and a maximum insulation at a considerable distance from the borders is observed. Left: median genome-wide interaction counts for intra-TAD (solid lines) and inter-TAD (dashed lines) read-pairs. Right: enrichment of intra-TAD signal at different distance scales. Notice that data from Dixon *et al.* is visualized at a 40 kb resolution.

E. Genome-wide assessment of differences in insulation between TADs in mouse embryonic stem with spans below or above median size (median = 615 kb) [13]. Smaller TADs attain higher intra-TAD signal enrichment, yet this maximum occurs at a larger distance from the borders as compared to large TADs. Left: median genome-wide interaction counts. Right: enrichment of intra-TAD signal at different distance scales. TADs with sizes below median are limited to their maximum span, with signal becoming highly variable in the last bins due to scarcity of the signal (dotted line).