The three-dimensional genome: principles and roles of long-distance interactions
M Jordan Rowley and Victor G Corces

The linear sequence of eukaryotic genomes is arranged in a specific manner within the three-dimensional nuclear space. Interactions between distant sites partition the genome into domains of highly associating chromatin. Interaction domains are found in many organisms, but their properties and the principles governing their establishment vary between different species. Topologically associating domains (TADs) extending over large genomic regions are found in mammals and Drosophila melanogaster, whereas other types of contact domains exist in lower eukaryotes. Here we review recent studies that explore the mechanisms by which long distance chromatin interactions determine the 3D organization of the genome and the relationship between this organization and the establishment of specific patterns of gene expression.

Address
Department of Biology, Emory University, 1510 Clifton Rd NE, Atlanta, GA 30322, USA

Corresponding author: Corces, Victor G (vcorces@emory.edu)

Introduction: a three-dimensional genome
The eukaryotic nucleus is a complex three-dimensional environment in which genome function depends not only on the linear arrangement of regulatory sequence elements but also on their spatial organization for effective control of gene expression [1,2]. Modulation of transcription occurs in part through spatial proximity of regulatory elements and gene promoters [1,2]. These interactions are essential for organismal development and response to environmental stimuli [2,3,4,5] in eukaryotes, including yeast, worms, plants, flies, and mammals [6–11]. Analysis of the role of chromatin 3D organization in gene expression is progressing rapidly, largely due to the development of chromosome conformation capture methods such as Hi–C [12]. Studies of long-range chromatin interactions have highlighted principles of three-dimensional genome organization, and whole genome chromatin contact maps have provided significant insights into how the 3D organization of the genome relates to gene expression [1,2]. Due to these advances we are beginning to understand overall chromosomal organization in the nucleus, how this organization is established, and how it can modulate gene expression. Here we discuss recent work that has helped answer important questions about the establishment and role of chromatin organization in genome regulation.

Units of organization
Whole-genome chromatin conformation capture (involving ligation and sequencing of spatially proximal DNA fragments — Hi–C as described in [13]) has been performed in several organisms and the results indicate that some features of chromatin 3D organization are consistent between some species (Figure 1). In many species and tissue types there are easily observable features of chromatin contact maps consisting of large genomic regions organized as contact domains [1,2]. Sequences within these Topologically Associated Domains (TADs) interact more frequently with sites inside than outside the domain. TADs with a median size of 880 kb have been found in mammals (Figure 1a) whereas Drosophila TADs have a median size of 107 kb [1] (Figure 1b). TADs in Caenorhabditis elegans are fairly weak, with the more easily defined domains located in the X chromosome of hermaphrodites [12] (Figure 1c). In Schizosaccharomyces pombe TAD-like contact domains (termed globules) are present at sizes ranging from 50 to 100 kb [14] (Figure 1e). However, large TAD-like structures are not as easily identifiable in some model organisms such as Saccharomyces cerevisiae and Arabidopsis thaliana [6–9] (Figure 1d,f). This poses the question of whether TADs are truly conserved features of chromatin organization in eukaryotes.

TADs vary in size throughout an individual genome and are overall shorter in the smaller genome of Drosophila (Figure 1a,b). Since TAD borders form at sites of active transcription and in regions with high gene density, it is likely that these features occur more often in smaller genomes [15]. To search for TADs in S. cerevisiae, the Hi–C protocol was modified to increase resolution by using micrococcal nuclease (MNase) digestion in lieu of restriction fragmentation (Micro-C) [15]. Using this method high-frequency contact domains were observed in S. cerevisiae [15] (Figure 1f). These domains are indeed smaller than those of mammals and Drosophila, between 2 and 10 kb in size, and contain only a few genes each
It is unknown whether these small domains are similar to TADs, especially compared to the relatively large domains found in *S. pombe* [14*] (Figure 1e). However, cohesin which is thought to be an important contributor to TAD border formation in many organisms, is enriched at domain borders in *S. pombe* as is the cohesin loader Scc2 in *S. cerevisiae* [14*,15*] (Figure 1e,f). It is therefore likely that the difference in domain size is related to the distribution of active genes and architectural proteins across the respective genomes [14*,15*].

It is clear from these studies that even simple eukaryotes contain structured contact domains and it is likely that chromatin domain organization is a conserved principle of eukaryotic life. However, contact maps of *Arabidopsis* do not display large TAD structures despite a genome size similar to that of *Drosophila* [7–9] (Figure 1d). The *Arabidopsis* genome does contain more genes, and thus gene density may contribute to smaller domains not as easily visible at the resolutions used. Indeed, a recent study using higher resolution contact maps has described small domain-like structures that are not as large or distinct as TADs in other organisms [9]. Another factor that may be confounding clear TAD detection is that samples used for *Arabidopsis* Hi–C have been from whole tissue and thus represent a population of different cell types [7–9]. Although almost nothing is known about the genome-wide distribution of architectural proteins in *Arabidopsis*, several features of chromatin organization stand out from Hi–C experiments carried out in this organism. Most prominent are inter-centromeric interactions and heterochromatic knot structures defined by high local and inter-chromosomal interactions [8] (Figure 1d). These organizational structures indicate that, despite a
lack of knowledge on architectural proteins, *Arabidopsis* has distinct features of chromatin organization.

Although it is tempting to think of TADs as well defined discrete genomic units, in reality TAD structure is more complex. Although sharp boundaries between TADs can sometimes be seen, many boundaries are fuzzy and/or have multiple possible locations (Figure 1). In addition, smaller structures can be seen within TADs corresponding to domains with even higher contact frequencies [10,16]. It is not clear how these smaller domains are different from TADs or if they simply represent similar organizational features at a finer scale. The complexity of Hi–C contact maps makes domain calling difficult and algorithms for this purpose vary and can produce either low resolution or differing boundary calls [11,17,18]. The complexity of TAD organization could suggest dynamic chromatin interactions and/or borders in which structures seen by Hi–C represent average or usual contact points in a population. At a finer scale than domains, point to point chromatin contacts such as enhancer–promoter interactions, represent important features of chromatin organization [1,2]. These contacts can be somewhat difficult to see by Hi–C due to the high frequency of interactions within TADs and many algorithms seek to optimally choose significant contacts [19,20]. More directly capturing interactions with promoter probes (Capture Hi–C/HiCap) has allowed these interactions to be explored in more detail [21*,22]. These studies (along with Pol II precipitated interactions) found that promoter–promoter interactions are fairly frequent and form multigene complexes [22,23]. Although no total interaction bias was seen for inactive versus active promoters, the level of expression between contact points was correlated, which implies the existence of a matrix of expression regulation [21*,22].

Promoter contacts as well as the complex point to point interaction matrix contribute to overall domain organization. Exploration of these and other fine scale organization principles is an important area of future research and may help to explain the formation and function of contact domains as discussed below.

**Establishment of long-range interactions**

How two specific distal sites can find each other in the three-dimensional space is an open question. It is known that architectural proteins such as CTCF and cohesin play a major role in long-range contact formation [1,24*]. CTCF has not been identified in plants, yeast, or *C. elegans* [25,26] which may account for the smaller or weaker organizational units observed in these organisms (Figure 1c–f). However, the high abundance of detectable long-distance interactions in these organisms [6–9] and the actual presence of contact domains in yeast [14*,15] suggest that other factors can function similar to CTCF in genome organization. In support of this idea, while no known CTCF protein has been found in *C. elegans*, large TAD-like domains exist [12*] (Figure 1c).

Although these domains are relatively weak on autosomes, they have better defined borders on the X chromosome of hermaphrodites [12*]. Interestingly, these borders coincide with the Dosage Compensation Complex (DCC), which contains condensin [12*] (Figure 1c). Depletion of the DCC reduces the strength of TAD borders [12*], thus the DCC (condensin) acts as a border defining complex without the necessity of CTCF. It is likely that other factors can function similarly in other organisms either for the establishment of domain borders or long-distance chromatin interactions.

Even in organisms that have CTCF, other architectural proteins contribute to genome organization [1,27]. In *Drosophila*, CTCF and other architectural proteins are present at thousands of sites throughout the genome called Architectural Protein Binding Sites (APBSs) [1]. Clustering of APBSs occurs at TAD borders and is thought to effectively insulate chromatin contacts from crossing these sites and/or promote long distance contacts solely in one direction [1] (Figure 1b). In mammals, CTCF motifs often cluster in the genome, but they usually lie too close together to resolve distinct peaks using ChIP-seq. Analysis of the distribution of BORIS and CTCF in the same cells enabled examination of occupancy at clustered sites [28]. The BORIS protein binds the same motif as CTCF, but it is expressed only in the male germ line and in tumor cells. In BORIS positive cells, CTCF and BORIS often bind next to each other at 2× CTCF sites containing two CTCF motifs separated by 30–50 bp. In cells that do not express BORIS, these sites are thought to be occupied by two CTCF proteins, as detected by DNase-seq [28]. These 2× CTCF sites are enriched at active promoters and enhancers, suggesting that clustered CTCF sites mediate interactions involved in transcription activation [28]. CTCF sites have also been shown to cluster with TFIH and Prdm5 at TAD borders [29]. It is therefore tempting to speculate that clustering of architectural proteins may play a role in the formation of TAD borders in both mammals and *Drosophila* [29].

CTCF is found not only at TAD borders, but also inside TADs, suggesting that the mere presence of this protein is not sufficient for TAD border formation [16,30]. Chromatin 3D organization in mammals seems to rely in part on the orientation of CTCF motifs with respect to each other in order to establish contacts between specific regions in the genome, such that interactions tend not to extend beyond CTCF motifs found in reverse–forward orientation [16,24*,31*,32] (Figure 1a). Recently, it was shown that deletion of reverse–forward oriented CTCF motifs at looping boundaries results in merging of domain/loop structures and causes interactions to extend beyond the deleted site [31*,33**]. This was also true if the CTCF motif orientation was disrupted solely by inversion [31*,33**].
only one CTCF motif in the reverse–forward pair does not merge the domains (i.e. the border is not lost) but does allow more interactions to occur between domains [33**]. This is probably because the remaining CTCF site is still able to insulate interactions from one direction, but not the other. Interestingly, it was found that even palindromic motifs are recognized by CTCF in only one orientation and that inversion of even these sites results in reverse CTCF binding orientation and loss of cohesin binding [31**]. Other studies hint at a more complex situation in the establishment of interactions between CTCF sites. Analysis of CTCF-mediated interactions in mouse embryonic stem cells (ESCs) and neural progenitor cells (NPCs) using 4C indicates that contacts between CTCF sites, even over distances of 5.8 Mb, show a preference in directionality, with 65% of loops forming between CTCF sites in convergent, 1% divergent, and 34% in the same orientation [34]. Although deletion of specific CTCF sites results in disruption of loops between convergent sites, inversion of a site does not result in the formation of new loops with sites now arranged in a convergent orientation. Further work will be necessary to resolve these discrepancies, which may be the result of the particular genomic contexts being analyzed in different studies.

Another recent study demonstrated the significance of CTCF orientation by inserting RAG recombinase sites into different regions of the genome [35]. At insertion sites, recombination events correlated with Hi-C interaction intensity and were restricted to domains marked by forward–reverse oriented CTCF motifs [35]. This was dramatically exemplified at the IgH locus where forward–reverse CTCF sites at IGCRI act as a domain boundary [35]. The function of this boundary was shown by deletion of these sites, which resulted in aberrations in recombination [35]. The importance of CTCF motif orientation has been documented by several groups [16**,24**,31**,32] giving rise to models where linear sequence determines three-dimensional genome structure [33**,36,37]. In these models, cohesin tracks along the chromosome such that DNA is pulled through a cohesin ring into a loop until appropriately oriented CTCF motifs are reached [33**,36,37].

**Functional aspects of chromatin organization**

One role of 3D chromatin organization may be to enable enhancer–promoter contacts over long linear distances [2,22]. Proper regulation of transcription requires interactions between enhancers and promoters via long distance contacts [1,2]. This strategy can complicate the interpretation of GWAS analyses because SNPs affecting specific traits are not necessarily located proximally to the affected genes. Indeed it was recently shown that QTLs are more accurately identified if long distance contacts are accounted for [38**]. Individuals with differential histone modifications in regulatory regions had corresponding SNPs at distally interacting transcription factor motifs. These differences in histone modifications among individuals correlate with differences in gene expression between non-adjacent genes that interact in the 3D space of the nucleus [38**]. Similarly, another study identified variable chromatin modules (VCMs) composed of enhancer marks and genes with coordinated expression [39**]. Coordinated variability in chromatin occurs mostly within topological domains and vcmQTLs are enriched for transcription factor binding sites and changes in gene expression [39**]. These studies also indicate that genes may share regulatory regions and that changes at single sites may affect chromatin state and gene expression at multiple locations [38**,39**]. Thus TAD organization and finer scale contacts are important features to consider when evaluating factors affecting gene expression. In fact the overall function of TADs in controlling gene expression could be the creation of local genomic environments in which enhancer–promoter interactions occur [10,16**,21**,22], and it is possible that TADs are simply low resolution views of clusters of enhancer–promoter matrices.

Consistent with the idea that TADs represent matrices of contacts among regulatory regions, interactions within TADs are different between cell types and are influenced by cellular differentiation as well as environmental conditions [4*,40,41]. Individual domain contact strength is different between lineage specific cell types while wide-spread changes occur in histone marks and CTCF binding [40]. The changes to chromatin interactions are best correlated with the enhancer mark H3K4me1 such that increased interactions correspond to increased active enhancer density [40]. This suggests that interactions can change to connect regulatory regions being actively used [41]. In support of this idea, differences in TAD organization can easily be seen when comparing different cell types. For example, specific domains that are well defined in embryonic cell lines are missing in lung fibroblasts (Figure 2a). This indicates that TAD structure is variable among cell types. Furthermore, TAD organization in *Drosophila* is altered during the heat-shock response [4*] (Figure 2b). Architectural proteins are re-distributed from TAD borders to inside TADs, resulting in a decrease in TAD border strength and an increase in inter-TAD interactions [4*] (Figure 2b). These changes in 3D organization may allow long-range interactions between Polycomb (Pc) bound enhancers and promoters, forming new Pc bodies and causing general transcription silencing, a characteristic of the stress response to temperature elevation [4*].

Although the effect of transcription inhibition on TAD organization appears to be small [4*,10]. RNA may play an important role in the 3D organization of the genome. Enhancers often contain IncRNAs, which may be important for recruitment of transcription factors or architectural proteins, or for other means of establishing long-range
contacts [42–45]. Studies of individual loci have found that elimination of specific lncRNAs do in fact alter chromatin interactions [42,44,45]. However, lncRNAs are also involved in several aspects of chromatin structure, including histone modifications and nucleosome positioning [43,46], and thus the effect on long range interactions may be indirect. This is probably not the case at the HOX4 locus, as depletion of blincRNA did not affect CTCF binding, but drastically altered local chromatin interactions [44]. The relationship between chromatin organization and lncRNAs is likely not one-sided, and just as long-distance interactions influence gene expression, the same may be true for lncRNA transcription.

Overall the chromatin environment in the eukaryotic nucleus involves long-distance contacts that form higher-order domains. The effects on chromatin interaction structures likely depend on communication among several features such as covalent histone modifications, nucleosome position, transcription factor binding, lncRNAs, and gene expression. Teasing apart independent roles in long-range contact formation may be difficult as each component must take part in the overall chromatin community.

Summary
Although our knowledge of the three-dimensional genome has advanced significantly, we still do not understand much about the specificity underlying the establishment of long-range interactions. From a wide perspective it seems that chromatin association domains exist in most species tested [4,12,14,15,47] (Figure 1), but the functional significance of these domains is only partially understood. Several layers of genome organization exist [1,2,16,47] and the contributions of each to the control of gene expression is an important direction for future studies. Understanding this relationship will require knowledge of how these domains form and, more specifically, how two sites in the genome are chosen for contact. Research on proteins or lncRNA with undiscovered architectural function may provide answers to these questions [29,42,44]. Truly, the study of genome organization is an exciting field at the cusp of really understanding the basic mechanisms that control gene expression.

Acknowledgements
Work in the authors’ lab was supported by U.S. Public Health Service Award R01 GM035463. MJR was supported by award F32 GM113570 from the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

• of special interest
•• of outstanding interest

4. This study examined chromatin interactions in patients with limb malformations and found that each was associated with disruption of a TAD boundary. Deletion of TAD borders resulted in increased interactions from genes to neighboring TADs containing enhancer elements. This suggested that enhancers act past the previous TAD border which results in misregulation of gene expression.
Using this approach, the authors suggest that the Hi-C contact map represents a high-resolution view of the genome organization. This method was able to capture contact maps that are consistent with the genomic organization of the species. This approach also highlights the importance of the Hi-C data in understanding the architectural properties of the genome.


Different models of chromatin organization are compared and it was found that chromatin extrusion best fits actual Hi-C data. Normal Hi-C data was compared to simulated data, as was data where CTCF sites at borders were systematically inverted or deleted. This effectively explores the role of CTCF motif orientation in domain organization.


This study indicates that SNPs and chromatin variation can influence gene/trait expression over long distances. These variations could affect distal enhancers, transcription factors, or chromatin modifications that interact with several different genomic regions. It suggests that accounting for Hi-C contacts improves the power of genome-wide association studies.


By comparing 47 individual genome-wide maps of chromatin factors this study found regions of the genome where variations in chromatin correlate (termed variable chromatin modules). These modules generally occur within contact domains and are associated with transcription factor binding sites. This indicates that changes within domains can alter multiple sites connected by long-range interactions.


