Review

CellPress

Special Issue: Chromatin Dynamics

Architectural proteins: regulators of 3D genome organization in cell fate

Elena Gómez-Díaz and Victor G. Corces

Department of Biology, Emory University, Atlanta, GA 30322, USA

The relation between alterations in chromatin structure and changes in gene expression during cell differentiation has served as a paradigm to understand the link between genome organization and function. Yet, the factors involved and the mechanisms by which the 3D organization of the nucleus is established remain poorly understood. The use of Chromosome Conformation-Capture (3C)-based approaches has resulted in a new appreciation of the role of architectural proteins in the establishment of 3D genome organization. Architectural proteins orchestrate higher-order chromatin organization through the establishment of interactions between regulatory elements across multiple spatial scales. The regulation of these proteins, their interaction with DNA, and their co-occurrence in the genome, may be responsible for the plasticity of 3D chromatin architecture that dictates cell and time-specific blueprints of gene expression.

Nuclear organization

Chromosomes are tightly packed in the nucleus within chromosome territories [1–4]. The 3D arrangement of the chromatin fiber in these territories during interphase is not random and, in principle, could be either a consequence of genome function or a pre-established effector of nuclear activity [5]. Nuclear processes, such as transcription and replication, require the assembly of large multiprotein complexes at promoters, enhancers, and replication origins [5–7]. These proteins often contain multiple interacting domains and, therefore, may drive the formation of intra- and inter-chromosomal contacts that contribute to the establishment of a specific 3D arrangement of the chromatin fiber. Given that this arrangement may be a consequence of genome function, it should be, at least in part, cell type specific, correlating with the transcriptional state of the cell. In addition to this transcription-driven organization, the cell appears to also use specific protein complexes whose main role is to establish contacts between distant sites in the genome to facilitate its 3D organization and allow the execution of specific functional outcomes. These proteins, generally referred to as insulator proteins, were originally characterized for their ability to interfere with enhancer–promoter interactions and to shield the expression of transgenes from the effects of adjacent sequences [8]. More recent results suggest that insulator sequences and their associated proteins not only inhibit, but also facilitate enhancer–promoter interactions, as well as regulating other aspects of transcription, in addition to more general roles in chromosome organization [9]. Given the varied, and sometimes contradictory, functions mediated by these proteins, we refer to them here as architectural instead of insulator proteins.

Starting with the premise that architectural proteins can mediate interactions between different sequences to regulate genome function, here we discuss mechanisms by which the interaction of these proteins with DNA or other proteins can be regulated to create specific patterns of nuclear 3D organization to elicit distinct functional outcomes that may contribute to the establishment of specific cell lineages during development.

Architectural proteins: structure and organization

Architectural proteins have been described in organisms ranging from yeast to humans [10]. In *Saccharomyces cerevisiae* and *Saccharomyces pombe*, the main architectural protein characterized to date is the RNA polymerase III-associated factor TFIIIC, which is present at genes transcribed by this polymerase, such as tRNA genes, as well as at many nontranscribed regions of the genome known as extra TFIIIC (ETC) loci [11,12]. TFIIIC colocalizes with cohesin and condensin, which have been shown to be required for its function in protecting against the spreading of histone covalent modifications associated with transcription silencing [13]. The best-characterized architectural protein in vertebrates is CCCTC-binding factor (zinc finger protein) (CTCF), which also requires association with cohesin for its enhancer-blocking function [10]. Recent experiments showing that tRNA genes can block enhancer function and that TFIIIC colocalizes with CTCF at many ETC loci through the mouse and human genomes suggest a conservation in the function of TFIIIC as an architectural protein from yeast to humans [14]. Other proteins shown to colocalize or directly interact with CTCF in vertebrates include Yin Yang 1 (YY1), Kaiso, chromodomain helicase DNA-binding protein 8 (CHD8), poly(ADP-ribose) polymerase 1 (PARP1), MYC-associated zinc-finger protein (MAZ), JUND, zinc-finger protein 143 (ZNF143), PR domain zinc-finger protein 5 (PRDM5), and nucleophosmin [15–17]. *Drosophila* has also been a rich source of information aimed at understanding the structure and organization of this class of proteins. Several DNA-binding architectural proteins, including CTCF, Suppressor of Hairy-wing [Su(Hw)], Boundary Element Associated

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

Keywords: CTCF; transcription; chromatin; development; differentiation.

0962-8924/

 http://dx.doi.org/10.1016/j.tcb.2014.08.003

Factor 32 (BEAF-32), DNA Replication Related Element Binding Factor (DREF) and TFIIIC, interact directly with the DNA [18,19]. These DNA-binding proteins recruit other accessory architectural proteins that do not bind to DNA directly, including cohesin (Rad21), condensins (Cap-H2 and Barren), Modifier of mdg4 [Mod(mdg4)], Centrosomal Protein 190 (CP190), L(3)mbt, Female sterile homeotic [Fs(1)h-L], Chromator, Zeste-white 5 (Zw5), and GAGA factor (GAF) [18,19].

These proteins are present in the genome in different combinations at what are termed 'architectural protein binding sites' (APBSs) [20]. Some sites in the genome contain one DNA-binding architectural protein and several accessory proteins and are called low-occupancy APBSs. Others contain several DNA-binding proteins bound within a short genomic region that recruit all or most accessory proteins and are called high-occupancy APBSs. These two types of sites have different roles in genome organization and function [20]. Architectural proteins in *Drosophila* and mammals have been shown to interact with RNAs [21–25], perhaps as a means of stabilizing these large multi-protein complexes, but the mechanistic role of these transcripts in their function has not been studied in detail.

The role of architectural proteins in 3D genome organization

The recent use of 3C-derived approaches, such as 5C and Hi-C, to measure interaction frequencies has allowed the establishment of comprehensive interaction maps over large regions or whole genomes [3,19,26–29]. Results from these experiments suggest that the *Drosophila* and mammalian genomes are compartmentalized into discrete regions termed 'topologically associating domains' (TADs) [27,30]. TADs are regions of the genome that show a high frequency of intradomain interactions, whereas the frequency of interactions with other TADs is low. Therefore, the interaction-based division of the genome into TADs is caused by the presence of sequences and associated proteins within TADs that frequently interact with their neighbors while, concurrently, other sequences and associated proteins form TAD borders that preclude interactions between adjacent TADs. In both *Drosophila* and mammals, TAD borders contain highly transcribed genes, including housekeeping genes, and architectural proteins [19,30,31]. Overall, only 15% of CTCF sites are present at TAD borders in mouse and human cells. Instead, most CTCF-binding sites (85%) localize within TADs, where they mediate interactions aimed at regulating various steps of the transcription process [27,32]. These findings together suggest the existence of various functional subclasses, border-associated versus nonborder, APBSs. It appears that the functional difference between the two types of APBSs rests on the number of architectural proteins present. In mammals, there is a strong association between TAD borders and the presence of CTCF, TFIIIC, cohesin, and PRDM5. Similarly, in *Drosophila*, CTCF clusters at TAD borders with condensins, cohesin, TFIIIC, BEAF-32, Su(Hw), CP190, Mod(mdg4), DREF, Chromator, and L(3)mbt [20] (Figures 1 and 2).

Work in mouse and human cells suggests that 60–70% of TADs are conserved between embryonic stem cells (ESCs) and differentiated cells, and even between mouse and human cells [27]. One interpretation of these observations is that TADs are static domains of genome organization that allow interactions among genes and regulatory sequences located in the same TAD, but preclude interactions between sequences located in different TADs [1,2,33]. However, it is important to consider that the concept of the TAD border is relative. Borders are determined computationally using algorithms that, either implicitly or explicitly, set thresholds for the relative frequency of interactions within and between TADs flanking the border. Based on the difference in frequency between inter- and intra-TAD interactions, it is possible to establish the concept of border strength [20,34,35]. Strong TAD borders are those for which Hi-C interaction matrices do not show interactions between sequences in the two adjacent TADs, whereas weak borders separate TADs with a high frequency of inter-TAD interactions. If one considers TAD borders as relative structures whose strength can be modulated, for example during cell differentiation, then it is possible to speculate that the apparent conservation of TADs between different cell types does not preclude the existence of interactions between genes and regulatory sequences present in different TADs. This has been observed for some enhancers involved in controlling the expression of genes during the differentiation of the mesoderm in *Drosophila*. In this case, some enhancers present in a TAD are able to contact promoters present in a different TAD [32].

What is responsible for the differences in the strength of borders separating different TADs? Recent results suggest that in both *Drosophila* and mammals, the strength of TAD borders directly correlates with the number of architectural proteins present at the border [20,34]. In *Drosophila*, high-occupancy APBSs containing 8–12 architectural proteins form strong TAD borders and show enhancer-blocking activity in functional reporter assays, whereas APBSs with five to eight proteins have weak border strength and weak enhancer-blocking activity. Interestingly, APBSs with two to five architectural proteins are enriched inside TADs and do not interfere with enhancer–promoter interactions in functional reporter assays [20]. These results agree with the view that architectural proteins located inside TADs may facilitate interactions between gene promoters and their regulatory sequences, whereas those present at TAD borders may preclude interactions between genes and regulatory sequences located in different TADs. The role of architectural proteins in controlling TAD organization and TAD border strength has been clearly demonstrated in several recent studies that analyzed the consequence of depletion of CTCF and cohesin on 3D genome organization [35–37]. Embryonic kidney cells depleted of cohesin showed a general loss of intrachromosomal interactions without affecting the TAD organization, whereas depletion of CTCF caused a similar decrease in the frequency of intra-TAD interactions concomitant with an increase in the frequency of interactions between adjacent TADs. Cohesin-deficient mouse astrocytes also showed a reduced number of CTCF- and cohesin-mediated long-range interactions together with a relaxation of TAD organization. This TAD relaxation could be a consequence of a decrease in TAD border strength due to the lack of

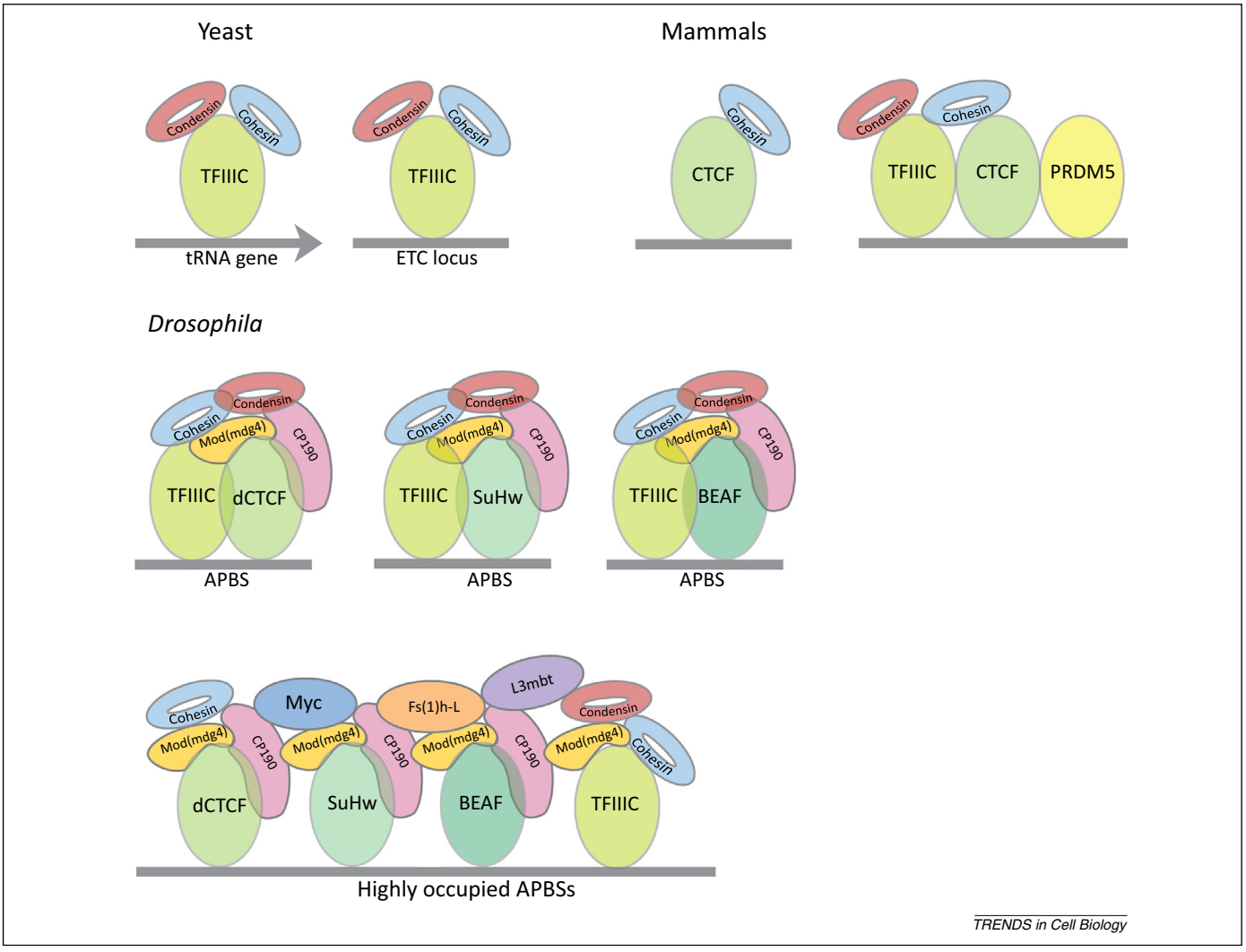

Figure 1. Structure and organization of architectural protein binding sites (APBS) in yeast, *Drosophila*, and mammals. Each DNA-binding architectural protein interacts with a particular sequence motif in the genome. For example, the TFIIIC protein interacts with the B-box sequence in tRNA genes or ETC sites. DNA-binding architectural proteins require interaction with accessory proteins to accomplish their function. For example, CTCF often interacts with Cohesin, whereas in *Drosophila* dCTCF, BEAF and Su(Hw) interact with Mod(mdg4) and CP190. High-occupancy binding sites are dense clusters of architectural proteins present at specific genomic regions and have been found in both *Drosophila* and mammals. See text for abbreviations of protein names.

cohesin binding or to an increase in the frequency of inter-TAD interactions, as observed in CTCF-depleted cells. Therefore, it is possible that cells may be able to regulate border strength by controlling the number of architectural proteins present at specific borders, thus allowing or constraining inter-TAD interactions to elicit novel patterns of gene expression during cell differentiation (Figure 2).

Regulation of architectural protein localization

CTCF is located at 55 000–65 000 sites in the genome of mammalian cells [38]. Of these, approximately 45% reside within intergenic regions, approximately 15% are located near promoters, and approximately 40% are present in introns and exons [38,39]. In *Drosophila*, CTCF and other architectural proteins are present in the genome at approximately 20-times fewer sites, in agreement with the difference in genome size, and their distribution with respect to genome features, such as promoters, introns, and exons, is similar to that of CTCF in mammals [30]. This conserved distribution at intergenic regions, 5′ transcribed untranslated regions (UTRs) and introns suggests that, in addition to their role at TAD borders, architectural proteins also have roles in the regulation of enhancer–promoter interactions, transcription pausing or elongation, and splicing.

Although 60–70% of the TADs are conserved among stem and differentiated cells corresponding to various lineages [27], the rest are not, suggesting that cells have the ability to alter the localization of architectural proteins during cell differentiation to regulate TAD border strength as well as various aspects of the transcription process. Thus, an important question in the field is how the distribution of architectural proteins is regulated to effect different functional outcomes during cell fate specification. Recent results suggest that the location of various architectural proteins is modulated by controlling their interaction with DNA or with other proteins via post-translational modification [40]. Covalent modifications of mammalian CTCF by poly(ADP-ribosyl)ation affects its ability to bind DNA [41], whereas the same modification of *Drosophila* CTCF affects its ability to interact with CP190 [40]. The interaction of CTCF with DNA can be also modulated by changes in the methylation status at its binding site [42]. Recent studies combining ChIP-seq and bisulfite sequencing in multiple human cell types revealed that 41% of

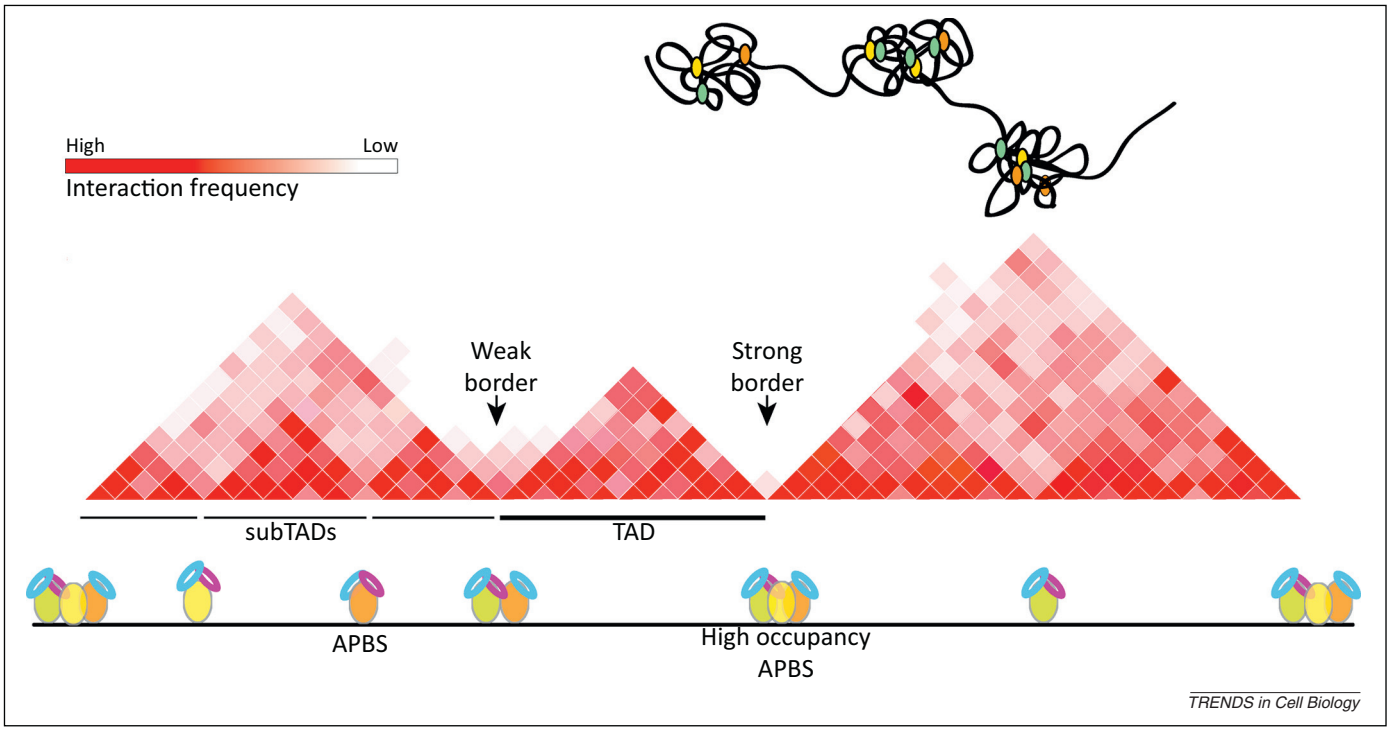

Figure 2. Model for how architectural proteins influence genome organization at various length scales [topologically associating domains (TADs) and sub-TADs] via long-range interactions. TADs are defined as regions of the genome undergoing high frequency of local interactions. They are separated by borders that preclude interactions between adjacent TADs. Highly occupied architectural protein binding sites (APBSs), containing multiple architectural proteins, are enriched at TAD borders, whereas low-occupancy APBSs are enriched inside TADs. Dynamic changes in the number and colocalization of architectural proteins may modulate TAD border strength across different cell types, allowing or restricting inter-TAD interactions to establish new patterns of gene expression during cell type specification.

cell specific CTCF binding is linked to differential DNA methylation [42]. Other studies have also reported a negative correlation between CTCF DNA binding and the DNA methylation status of CpGs within the CTCF binding sites [43], although the picture seems to be more complex. CTCF sites bound by this protein show the same methylation level as all other sites in the genome, and the binding affinity of CTCF correlates with the level of unmethylation, suggesting that CTCF binds with low affinity to sites in the genome that are partially methylated [44]. Equally intriguing is the fact that CTCF can actively inhibit DNA methylation at CTCF-binding sites by interacting with PARylated PARP1, which in turns inhibits DNA methyltransferase 1 (DNMT1) activity [45]. Thus, it is not clear from these data whether DNA methylation has a causal role in CTCF binding or is a consequence of this process.

Taken together, these studies underscore the complexity and possible importance of DNA methylation and protein covalent modifications in modulating the occupancy and interactions of architectural proteins [46]. It is possible that, by controlling the interactions of architectural proteins with DNA and other proteins, the cell can regulate their location in the genome and, therefore, control different steps of the transcription process to establish or maintain patterns of gene expression during cell differentiation.

Architectural proteins mediate functional chromatin interactions during cell fate specification

The mechanisms by which cell-to-cell differences in chromatin architecture arise and how these various topologies can result in diverse functional outcomes remains a major gap in our understanding of cell fate-specification processes. Below, we review recent examples that illustrate how architectural proteins are responsible for the establishment of cell specific 3D chromatin structures that may contribute to the spatiotemporal regulation of transcription during pluripotency and along various differentiation pathways.

Architectural proteins, types, and co-occurrence drive the transcriptional plasticity of ESCs

Data gathered from independent studies using 5C, Hi-C, and ChIA-PET, comparing 3D chromatin organization in human and mouse ESCs, indicate that the pluripotent genome displays unique topological and functional features. These include global low-levels of transcription, a lack of long-range contacts at a global scale, and disorganization of the heterochromatin in the nuclei [47–49]. As ESCs differentiate, there is a dynamic reorganization of the network of interactions genome wide. This involves compartmentalization of the genome into high-frequency interaction domains coupled with very tight spatiotemporal regulation of transcription [50–54] (Figure 2). Interestingly, ESC-specific TADs are mainly shaped around the pluripotency factors Octamer 4 (*Oct4*), Sex-determining region Y (SRY) Box 2 (*Sox2*), and *Nanog*, and interactions occur between genome regions that are rich in superenhancers and genes that control the pluripotent state [48,55].

An important question arising from these studies is how these long-range interaction maps are reconfigured in the transition from ESCs to differentiated cells, and whether architectural proteins have a key role in pluripotency. Previous studies using ChIA-PET and ChIP-sequencing have suggested a role for CTCF, Mediator, and cohesin as

chromatin organizers in ESCs, showing that they engage in functional interactions with pluripotent genes and transcription factors [56,57]. This is supported by results obtained during the differentiation of ESCs into the endodermal lineage, where CTCF has been shown to directly recruit TAF3, a TBP-associated core promoter factor, to distal regulatory sequences. TAF3 present at CTCF/cohesin sites cooperates with these two proteins in mediating interactions between these enhancers and promoters during the differentiation of ESCs into endoderm [58,59]. Changes in CTCF occupancy during differentiation of ESCs are associated with alterations in nucleosome positioning and DNA demethylation [60,61].

A comprehensive analysis of APBS occupancy patterns in the context of ESC differentiation was obtained by comparing 5C interaction maps in ESCs and neural progenitor cells (NPCs). The study revealed two classes of interactions: ESC-specific enhancer–promoter short-range contacts involving cohesin and Mediator but not CTCF, and larger loops coinciding with CTCF and cohesin binding. Loops at the sub-megabase scale show clear reorganization during differentiation, whereas CTCF-mediated megabase loops remain invariant and were proposed to have a role in chromosome folding [62]. These observations can be interpreted in the context of a model in which the regulation in the occupancy of various subclasses of architectural proteins results in changes in chromatin organization that allow the cell to switch between various transcription programs. Consistent with this hypothesis, two independent studies in ESCs using conditional knockdowns in cohesin and Mediator, found either an artificial induction of differentiation of ESCs or impaired reprogramming into induced pluripotent stem cells (iPSCs) [50,63]. Whether architectural protein binding alone defines pluripotency, or whether pluripotency is instead driven by state-specific transcription factors and enhancers, remains unanswered.

Together, these findings support a key role of architectural proteins in the dynamic folding of the genome during cell fate specification. Yet, several important issues remain. For example, how general is the relationship between architectural proteins, pluripotency transcription factors, and/or enhancers? Are architectural proteins causal to changes in the pluripotent state, or a consequence of the binding of pluripotency transcription factors? Do TADs and TAD borders have a regulatory role in the transition between pluripotent and differentiated chromatin states?

CTCF and cohesin regulate lymphocyte differentiation

Lymphocyte differentiation provides a compelling example of the role of architectural proteins and chromatin 3D architecture in generating cell diversity. B and T lymphocytes have a unique antigen receptor that is highly variable and cell specific, and the basis of adaptive immunity. The variable portion of the B cell immunoglobulin (Ig) and T cell receptor (Tcr) loci is encoded by multiple copies of variable (V), diversity (D), and joining (J) gene segments that span across large genomic regions. Antigenic diversity in B and T lymphocytes is generated by gene rearrangements of these V, D, and J gene segments catalyzed by the RAG1/2 recombinase. Growing evidence suggests that changes in 3D chromatin architecture are key to the generation of B and T lymphocyte receptor diversity [64]. The antigen receptor loci are particularly enriched in binding sites for CTCF and cohesin [65–68], leading to the proposal that these two proteins function together in modulating lymphocyte differentiation by at least two mechanisms. First, by forming rosette-like structures that facilitate lineage-specific enhancer–promoter communication and differentially activate transcription. In the Tcrα/δ locus of $CD4^+$ $CD8^+$ double-positive thymocytes, binding of CTCF and cohesin at sites flanking the TEA promoter and the Eα enhancer is required for the long-range promoter–enhancer interactions that control Tcrκ transcription (Figure 3A) [65,66]. This is supported by functional studies in mice, where Rad21-deficient thymocytes show reduced interactions between the Tcrα enhancer Eα and the TEA promoter, and reduced TEA transcription, while provision of pre-rearranged TCR transgenes largely rescues thymocyte differentiation [65]. A second mechanism by which CTCF contributes to B and T cell development is by alternatively facilitating and repressing V(D)J rearrangements via modulation of chromatin accessibility at the antigen receptor locus [66,69,70]. This has been shown using 3C-based analyses in pre-pro-B cells that reveal long-range interactions between CTCF-binding sites near the Silencer intervening sequence (SIS), Vκ gene segments, and the boundaries of the *Igκ* locus. These interactions physically restrict the communication between the Jκ-Cκ-enhancer and the proximal Vκ promoter, thereby promoting rearrangement with distal Vκ segments, whereas the conditional knockout of CTCF results in more interactions between the intronic Igκ enhancer and the proximal Vκ segments and a bias toward proximal Vκ recombination [71]. Likewise, in the *IgH* locus, ChIP sequencing and 3C data show colocalization of CTCF and Rad21 at approximately 60 sites throughout the V_H region and two CTCF-binding sites within the Intergenic control region 1 (IGCR1). These sites form the bases of the multiloop rosette structures that mediate ordered and lineage-specific V_H-to-DJ_H recombination by biasing distal over proximal V_H rearrangements [68,72]. That is, IGCR1, which is positioned between the V_H and D_H clusters, suppresses the rearrangement of proximal V_H segments by forming a CTCF-mediated loop that isolates the proximal V_H promoter from the influence of the downstream Eμ enhancer (Figure 3B). Similarly, in $CD4^+$ $CD8^+$ double-positive thymocytes, the Tcrα enhancer Eα activates 3′ Vα promoters and the TEA promoter at the 5′ end of the Jα array to initiate Vα-to-Jα rearrangement. It has been shown that cohesin depletion in $CD4^+$ $CD8^+$ double-positive mouse thymocytes impairs the functional separation between Tcrκ and the neighboring housekeeping gene *Dad1* [65]. More recent 3C data further highlights the role of CTCF as an important regulator of Tcrα locus recombination. Vα-to-Jα recombination occurs within a chromatin hub that is dependent on long-range interactions between CTCF-binding sites and the Tcrα enhancer. The loss of CTCF in DP thymocytes dysregulates chromatin looping and locus contraction impairing Vα-to-Jα rearrangement [66].

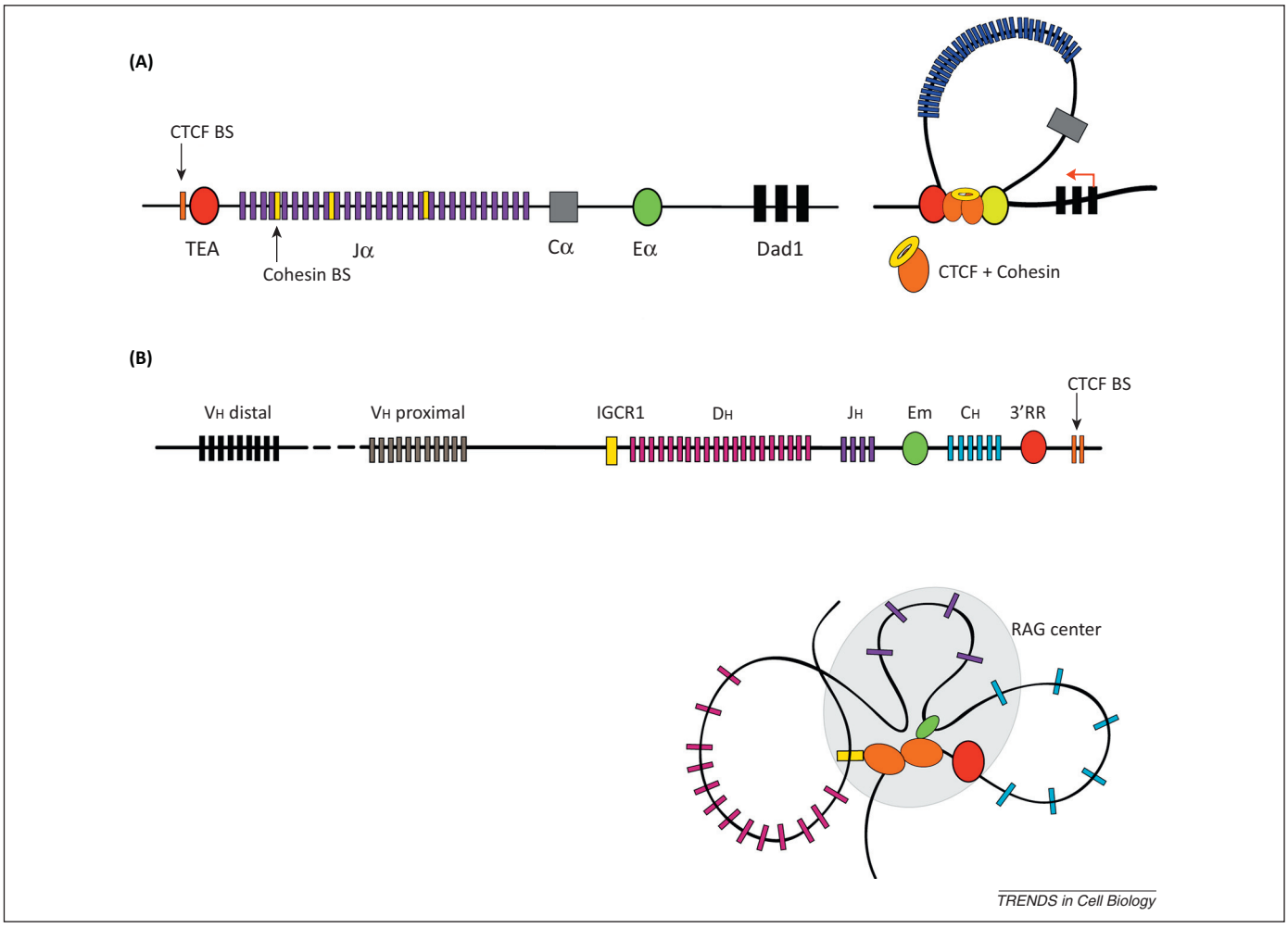

Figure 3. CCCTC-binding factor (zinc finger protein) (CTCF) and cohesin regulate antigen receptor diversity in T and B lymphocytes. Antigen receptor diversity of B and T cells is generated by the rearrangement of different variable (V), diversity (D), and joining (J) gene segments in individual lymphocytes. CTCF influences the outcome of V(D)J recombination by regulating enhancer–promoter interactions and locus compaction. The general organization of the T cell receptor (TCRα) and immunoglobulin (Ig)H loci are shown. **(A)** In the TCRα locus of thymocytes, cobinding of the CTCF/cohesin complex at the TEA promoter and the Eα enhancer results in a DNA loop that is required to activate transcription of the nearby housekeeping gene *Dad1*. **(B)** In the IgH locus of pre-pro-B cells, CTCF-mediated looping between the Eμ enhancer and 3′ regulatory region (3′RR) with distinct D_H-J_H-C_H gene segments is required for ordered (D_H–J_H) recombination. CTCF binding at intergenic control region 1 (IGCR1) blocks the influence of the Eμ enhancer on proximal variable (V_H) regions.

CTCF/cohesin mediate monoallelic gene expression in neuronal differentiation

The differentiation of the hundreds of specialized neuronal types present in the brain requires the establishment of specific patterns of gene expression. Among the many genes that are transcribed in a neuron-specific manner, the mechanisms underlying the expression of protocadherins have been studied in great detail. Protocadherins (Pcd) are part of the larger family of calcium-dependent cell adhesion molecules in the central nervous system. In mammals, there are more than 50 protocadherin isoforms grouped into three gene clusters named α, β, and γ. Interestingly, the genomic organization of the Pcd gene clusters resembles that of the immunoglobulin and T cell receptor genes, albeit the mechanism of regulation differs slightly in that it does not involve somatic rearrangements. In neurons, single cell diversity results from the monoallelic gene expression of a protocadherin gene cluster, so that only one isoform is transcribed at a time. This is achieved by stochastic promoter choice from the 15-variable first exons, followed by alternative pre-mRNA *cis*-splicing of the chosen alternative exons to three downstream constant exons [73]. Two independent studies in human and mouse cell lines provide evidence that CTCF and cohesin-mediated interactions are ultimately responsible for the monoallelic expression at the protocadherin α cluster [74,75]. In the first of these studies, Maniatis and colleagues used a human diploid neuroblastoma cell line SK-N-SH expressing a select number of alternative *Pcdh* isoforms. In the case of Pcdhα, the cluster comprises a set of 14–15 variable exons, each containing its own promoter and two *cis*-regulatory elements with enhancer activity (HS7 and HS5-1). CTCF/cohesin co-bound sites interact with the TSS and first exon of α4, α8, and α12 isoforms, and activate specific transcription of these isoforms. In the second study, the same authors showed that DNA looping at Pcdhα requires specific cobinding of the CTCF/cohesin complex to two symmetrically aligned binding sites in both the transcriptionally active promoters and the HS5-1 enhancer. In addition, this study identified a unique regulatory role for cohesin, which binds to another enhancer (HS7) independently of CTCF. Functional analyses demonstrated

that CTCF or cohesin deletion and/or deletion of the CTCF-bound HS5-1 enhancer dysregulates chromatin architecture at this locus and results in nonspecific expression of Pcdhα isoforms [76,77]. The findings suggest a primary role for CTCF/cohesin in establishing interactions between the two downstream enhancers and individual exon promoters that drive Pcdhα specific enhancer-promoter communication (Figure 4).

A question that arises from these studies is whether CTCF/cohesin may function by additional mechanisms, for example by regulating chromatin accessibility and compaction at this locus. In addition, the mechanisms underlying the exclusion of homologous alleles remain unclear and will be an important issue for future work.

Architectural proteins, 3D organization, and Hox gene regulation during limb development

In vertebrates, Hox genes, present in four clusters named A–D, are activated sequentially relative to their positions within their genomic loci, leading to an anterior–posterior patterning of gene expression along the body axis. Recent studies using various 3C techniques suggest that dynamic changes in chromatin architecture are key to transcriptional regulation of Hox gene clusters and underlie the collinearity in transcription during limb and trunk development [78–80]. In mouse limbs, the HoxD locus is located at the cusp of adjacent TADs [81]. The early *HoxD1-9* genes are expressed in the proximal limb and regulated by enhancers located at the 3′ telomeric gene desert, whereas the late *HoxD12-10* genes are expressed in the distal limb and regulated by enhancers located in a 5′ centromeric gene desert [81]. The transition from early to late limb development involves topological and functional switches between the regulatory archipelagos located at either side of the Hox gene cluster. Following this switch, new sets of interactions are progressively established and collinearity progresses with two subsequent waves of transcription [79]. Thus far, however, the regulatory sequences and the mechanisms underlying the conformational and functional switches between domains remain obscure. It has been hypothesized that CTCF-binding sites located at the TAD borders act as enhancer-blocking barriers that insulate early and late HoxD genes. Consistent with an involvement of CTCF in HoxD regulation, ChIP-chip analyses revealed CTCF-binding sites flanking seven of the nine HoxD genes, as well as CTCF sites in the centromeric and telomeric gene deserts. The conditional inactivation of CTCF in mice results in massive apoptosis leading to a nearly complete loss of limb structure [82]. The situation is more complex in the case of the HoxA cluster, where studies in human cell lines and mouse embryos have reported different HoxA architectures [80,83]. However, a common theme in these studies is the selective gene activation through chromatin looping, which seems to depend on CTCF. Supporting evidence for this conclusion comes from a recent study using 5C in a human leukemia cell line, showing that HoxA gene activation coincides with a progressive loss of contacts throughout the region and the reconfiguration of CTCF-mediated

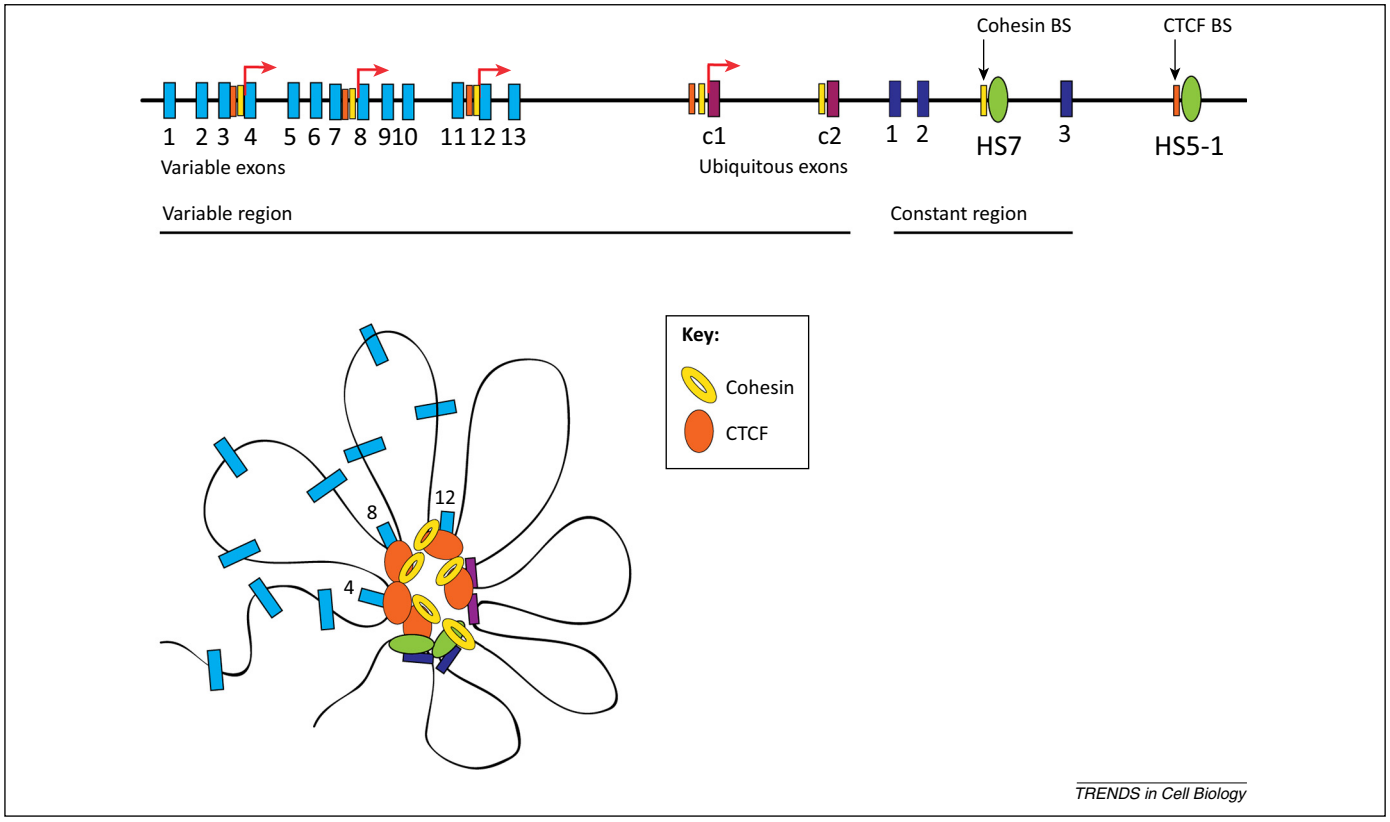

Figure 4. CCCTC-binding factor (zinc finger protein) (CTCF) and cohesin mediate monoallelic gene expression in neurons. The human protocadherin A (PCDHα) gene cluster contains 13 variable exons (1–13) and two c-type first exons (c1 and c2), which are expressed ubiquitously in neurons. Monoallelic gene expression of alternative isoforms occurs stochastically via a promoter choice mechanism that determines the splice site and, as such, which variable exon is included in a Pcdh mRNA. Promoter choice requires the formation of a chromatin hub that is mediated by the cobinding of the CTCF/cohesin complex to the distal HS5-1 enhancer and two symmetrically aligned binding sites (yellow, cohesin; orange, CTCF) in the active promoters (α4, α8, and α12). An additional binding site for cohesin exists in the HS-7 enhancer.

interactions between the two TAD boundaries [83,84]. However, CTCF-dependent chromatin looping at the HoxA/D gene clusters is still insufficient to explain the topological and functional changes that preclude the transition between early and late regulation during limb development. Furthermore, the fact that Hox gene clusters display different topologies and apparently different transcription regulatory mechanisms across cell types and developmental processes [78,80,81,84] questions the role of CTCF as the sole player in this process. In fact, Polycomb complexes have been shown to be directly involved in regulating changes in the topology of Hox loci in different developmental settings [85]. Therefore, it is likely that the presence of both Pc-G and architectural proteins at TAD borders and within TADs might be responsible for shaping the 3D organization of Hox gene clusters.

Concluding remarks

By mediating communication between distant DNA sequences, architectural proteins contribute to the organization of the genome into topological and functional domains. However, the particulars of the different classes of architectural proteins associated with these domains, and how they facilitate or preclude interactions, remain obscure. In the context of cell differentiation, an emerging theme from recent studies is that the dynamic regulation of the localization of architectural proteins, and their interactions with DNA and other proteins, modulate the network of contacts that result in cell specific chromatin configurations. This provides a novel mechanism for cell state-specific regulation of transcription in pluripotency and cell fate specification. During the transition from ESC to differentiated cells, genome-wide interaction maps are reshaped around cell type-specific enhancers and master transcription factors, at the same time that the binding landscapes of various architectural proteins are disrupted. However, whether architectural proteins are directly responsible for these changes is unclear. Filling this gap will require understanding the dynamics of architectural protein co-occupancy and their integration with TFs throughout the genome. Meanwhile, locus-specific studies, such as those in lymphocytes and neurons, have provided compelling and direct evidence of the importance of chromatin looping mediated by architectural proteins (CTCF/cohesin) in regulating differentiation. Much of our current knowledge is based on data obtained in different cell lines or tissue types that primarily lack functional validation. Thus, whether architectural protein binding is sufficient and necessary to engage in functional chromatin loops, remains unclear. Future research should investigate the mechanisms regulating architectural protein localization and cooperative binding, as well as the dynamics of 3D landscapes across various cell types and differentiation stages. Answers to these questions are key to our understanding of the regulation of differentiation and developmental processes.

References

1 Nora, E.P. *et al.* (2013) Segmental folding of chromosomes: a basis for structural and regulatory chromosomal neighborhoods? *Bioessays* 35, 818–828
2 Bickmore, W.A. (2013) The spatial organization of the human genome. *Annu. Rev. Genomics Hum. Genet.* 14, 67–84
3 Lieberman-Aiden, E. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293
4 Bauer, C.R. *et al.* (2012) Condensin II promotes the formation of chromosome territories by inducing axial compaction of polyploid interphase chromosomes. *PLoS Genet.* 8, e1002873
5 Van Bortle, K. and Corces, V.G. (2012) Nuclear organization and genome function. *Annu. Rev. Cell Dev. Biol.* 28, 163–187
6 Gilbert, D. *et al.* (2010) Space and time in the nucleus developmental control of replication timing and chromosome architecture. *Cold Spring Harb. Symp. Quant. Biol.* 75, 143–153
7 Li, B. *et al.* (2007) The role of chromatin during transcription. *Cell* 128, 707–719
8 Kellum, R. and Schedl, P. (1991) A position-effect assay for boundaries of higher order chromosomal domains. *Cell* 64, 941–950
9 Phillips-Cremins, J.E. and Corces, V.G. (2013) Chromatin insulators: linking genome organization to cellular function. *Mol. Cell* 50, 461–474
10 Ong, C.T. and Corces, V.G. (2014) CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* 15, 234–246
11 Hiraga, S. *et al.* (2012) TFIIIC localizes budding yeast ETC sites to the nuclear periphery. *Mol. Biol. Cell* 23, 2741–2754
12 Moqtaderi, Z. *et al.* (2010) Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. *Nat. Struct. Mol. Biol.* 17, 635–640
13 D'Ambrosio, C. *et al.* (2008) Identification of cis-acting sites for condensin loading onto budding yeast chromosomes. *Genes Dev.* 22, 2215–2227
14 Van Bortle, K. and Corces, V.G. (2012) tDNA insulators and the emerging role of TFIIIC in genome organization. *Transcription* 3, 277–284
15 Galli, G.G. *et al.* (2013) Genomic and proteomic analyses of Prdm5 reveal interactions with insulator binding proteins in embryonic stem cells. *Mol. Cell. Biol.* 33, 4504–4516
16 Zlatanova, J. and Caiafa, P. (2009) CTCF and its protein partners: divide and rule? *J. Cell Sci.* 122, 1275–1284
17 Xie, W. *et al.* (2013) Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* 153, 1134–1148
18 Schwartz, Y.B. *et al.* (2012) Nature and function of insulator protein binding sites in the *Drosophila* genome. *Genome Res.* 22, 2188–2198
19 Sexton, T. *et al.* (2012) Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* 148, 458–472
20 Van Bortle, K. *et al.* (2014) Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol.* 15, R82
21 Lei, E.P. and Corces, V.G. (2006) RNA interference machinery influences the nuclear organization of a chromatin insulator. *Nat. Genet.* 38, 936–941
22 Moshkovich, N. *et al.* (2011) RNAi-independent role for Argonaute2 in CTCF/CP190 chromatin insulator function. *Genes Dev.* 25, 1686–1701
23 Saldaña-Meyer, R. *et al.* (2014) CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev.* 28, 723–734
24 Sun, S. *et al.* (2013) Jpx RNA activates Xist by evicting CTCF. *Cell* 153, 1537–1551
25 Yao, H. *et al.* (2010) Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.* 24, 2543–2555
26 Belton, J.M. *et al.* (2012) Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* 58, 268–276
27 Dixon, J.R. *et al.* (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380
28 Simonis, M. *et al.* (2007) An evaluation of 3C-based methods to capture DNA interactions. *Nat. Methods* 4, 895–901
29 de Wit, E. and de Laat, W. (2012) A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* 26, 11–24
30 Hou, C. *et al.* (2012) Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol. Cell* 48, 471–484
31 Nora, E.P. *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385

32 Ghavi-Helm, Y. *et al.* (2014) Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* http://dx.doi.org/10.1038/nature13417
33 Gibcus, J.H. and Dekker, J. (2013) The hierarchy of the 3D genome. *Mol. Cell* 49, 773–782
34 Van Bortle, K. *et al.* (2012) *Drosophila* CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. *Genome Res.* 22, 2176–2187
35 Sofueva, S. *et al.* (2013) Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* http://dx.doi.org/10.1038/emboj.2013.237
36 Zuin, J. *et al.* (2013) Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. U.S.A.* http://dx.doi.org/10.1073/pnas.1317788111
37 Seitan, V.C. *et al.* (2013) Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Res.* http://dx.doi.org/10.1101/gr.161620.113
38 Kim, T.H. *et al.* (2007) Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 128, 1231–1245
39 Chen, H. *et al.* (2012) Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PLoS ONE* 7, e41374
40 Ong, C.T. *et al.* (2013) Poly(ADP-ribosyl)ation regulates insulator function and intrachromosomal interactions in *Drosophila*. *Cell* 155, 148–159
41 Yu, W. *et al.* (2004) Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. *Nat. Genet.* 36, 1105–1110
42 Wang, H. *et al.* (2012) Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* 22, 1680–1688
43 Mukhopadhyay, R. *et al.* (2004) The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. *Genome Res.* 14, 1594–1602
44 Feldmann, A. *et al.* (2013) Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet.* 9, e1003994
45 Zampieri, M. *et al.* (2012) ADP-ribose polymers localized on CTCF–Parp1–Dnmt1 complex prevent methylation of CTCF target sites. *Biochem. J.* 441, 645–652
46 Jones, P.A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* 13, 484–492
47 Efroni, S. *et al.* (2008) Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* 2, 437–447
48 de Wit, E. *et al.* (2013) The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature* 501, 227–231
49 Denholtz, M. and Plath, K. (2012) Pluripotency in 3D: genome organization in pluripotent cells. *Curr. Opin. Cell Biol.* 24, 793–801
50 Apostolou, E. *et al.* (2013) Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming. *Cell Stem Cell* 12, 699–712
51 Wang, P. *et al.* (2012) Higher-order genomic organization in pluripotent stem cells. *Protein Cell* 3, 483–486
52 Wei, Z. *et al.* (2013) Klf4 organizes long-range chromosomal interactions with the oct4 locus in reprogramming and pluripotency. *Cell Stem Cell* 13, 36–47
53 Zhang, Y. *et al.* (2013) Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature* 504, 306–310
54 Peric-Hupkes, D. *et al.* (2010) Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* 38, 603–613
55 Whyte, W.A. *et al.* (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307–319
56 Kagey, M.H. *et al.* (2010) Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430–435
57 Handoko, L. *et al.* (2011) CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.* 43, 630–638
58 Liu, Z. *et al.* (2011) Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. *Cell* 146, 720–731
59 Delgado-Olguín, P. *et al.* (2011) CTCF promotes muscle differentiation by modulating the activity of myogenic regulatory factors. *J. Biol. Chem.* 286, 12483–12494
60 Plasschaert, R.N. *et al.* (2013) CTCF binding site sequence differences are associated with unique regulatory and functional trends during embryonic stem cell differentiation. *Nucleic Acids Res.* http://dx.doi.org/10.1093/nar/gkt910
61 Teif, V.B. *et al.* (2014) Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development. *Genome Res.* http://dx.doi.org/10.1101/gr.164418.113
62 Phillips-Cremins, J.E. *et al.* (2013) Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* 153, 1281–1295
63 Zhang, H. *et al.* (2013) Intrachromosomal looping is required for activation of endogenous pluripotency genes during reprogramming. *Cell Stem Cell* 13, 30–35
64 Seitan, V.C. *et al.* (2012) Cohesin, CTCF and lymphocyte antigen receptor locus rearrangement. *Trends Immunol.* 33, 153–159
65 Seitan, V.C. *et al.* (2011) A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature* 476, 467–471
66 Shih, H.Y. *et al.* (2012) Tcra gene recombination is supported by a Tcra enhancer- and CTCF-dependent chromatin hub. *Proc. Natl. Acad. Sci. U.S.A.* 109, E3493–E3502
67 Degner, S.C. *et al.* (2009) Cutting edge: developmental stage-specific recruitment of cohesin to CTCF sites throughout immunoglobulin loci during B lymphocyte development. *J. Immunol.* 182, 44–48
68 Degner, S.C. *et al.* (2011) CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9566–9571
69 Guo, C. *et al.* (2011) Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. *Cell* 147, 332–343
70 Guo, C. *et al.* (2011) CTCF-binding elements mediate control of V(D)J recombination. *Nature* 477, 424–430
71 Ribeiro de Almeida, C. *et al.* (2011) The DNA-binding protein CTCF limits proximal Vkappa recombination and restricts kappa enhancer interactions to the immunoglobulin kappa light chain locus. *Immunity* 35, 501–513
72 Guo, C.G. *et al.* (2011) CTCF-binding elements mediate control of V(D)J recombination. *Nature* 477, 424
73 Chen, W.V. and Maniatis, T. (2013) Clustered protocadherins. *Development* 140, 3297–3302
74 Guo, Y. *et al.* (2012) CTCF/cohesin-mediated DNA looping is required for protocadherin alpha promoter choice. *Proc. Natl. Acad. Sci. U.S.A.* 109, 21081–21086
75 Monahan, K. *et al.* (2012) Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of protocadherin-alpha gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 109, 9125–9130
76 Hirayama, T. *et al.* (2012) CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. *Cell Rep.* 2, 345–357
77 Kehayova, P. *et al.* (2011) Regulatory elements required for the activation and repression of the protocadherin-α gene cluster. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17195–17200
78 Noordermeer, D. *et al.* (2011) The dynamic architecture of Hox gene clusters. *Science* 334, 222–225
79 Andrey, G. *et al.* (2013) A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* 340, 1234167
80 Berlivet, S. *et al.* (2013) Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs. *PLoS Genet.* 9, e1004018
81 Montavon, T. *et al.* (2011) A regulatory archipelago controls Hox genes transcription in digits. *Cell* 147, 1132–1145
82 Soshnikova, N. *et al.* (2010) Functional analysis of CTCF during mammalian limb development. *Dev. Cell* 19, 819–830
83 Kim, Y.J. *et al.* (2011) Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7391–7396
84 Rousseau, M. *et al.* (2013) Hox in motion: tracking HoxA cluster conformation during differentiation. *Nucleic Acids Res.* http://dx.doi.org/10.1093/nar/gkt998
85 Eskeland, R. *et al.* (2010) Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* 38, 452–464