

CTCF: an architectural protein bridging genome topology and function

Chin-Tong Ong and Victor G. Corces

Abstract | The eukaryotic genome is organized in the three-dimensional nuclear space in a specific manner that is both a cause and a consequence of its function. This organization is partly established by a special class of architectural proteins, of which CCCTC-binding factor (CTCF) is the best characterized. Although CTCF has been assigned various roles that are often contradictory, new results now help to draw a unifying model to explain the many functions of this protein. CTCF creates boundaries between topologically associating domains in chromosomes and, within these domains, facilitates interactions between transcription regulatory sequences. Thus, CTCF links the architecture of the genome to its function.

Chromosome conformation capture

(3C). A ligation-based technique that is used to map interactions between two specific genomic regions.

Eukaryotic genomes are dynamically packaged into multiple levels of organization, from nucleosomes to chromatin fibres to large-scale chromosomal domains that occupy defined territories of the nucleus. The three-dimensional interplay of protein–DNA complexes facilitates timely realization of intricate nuclear functions such as transcription, replication, DNA repair and mitosis¹. A combination of microscopy and chromosome conformation capture (3C)-related approaches² has revealed that CCCTC-binding factor (CTCF) is, in large part, responsible for bridging the gap between nuclear organization and gene expression. CTCF is the main insulator protein described in vertebrates. Initially characterized as a transcription factor that is capable of activating or repressing gene expression in heterologous reporter assays^{3,4}, CTCF was later found to display properties that are characteristic of insulators (that is, the ability to interfere with enhancer–promoter communication or to buffer transgenes from chromosomal position effects caused by heterochromatin spreading). These properties, which were observed using transgenic assays, were interpreted to suggest a role for insulators in restricting enhancer–promoter interactions and in establishing functional domains of gene expression.

In this Review, we discuss recent evidence from the use of 3C-related techniques, which indicates that the diverse properties of CTCF and other insulator proteins are based on their broader role in mediating both interchromosomal and intrachromosomal interactions between distant sites in the genome. As a result of these interactions, CTCF elicits specific functional outcomes that are context dependent, determined by the nature

of the two sequences brought together and by the proteins with which they interact. Consequently, CTCF contributes to the establishment of a three-dimensional structure of the chromatin fibre in the nucleus that is both an effector and a consequence of genome function. As the role of CTCF extends well beyond that originally attributed to insulator proteins and its functional effects are based on its ability to mediate interactions between distant sequences, we propose the term ‘architectural’ rather than ‘insulator’ to describe this type of protein.

Regulation of CTCF binding to DNA

CTCF is conserved in most bilaterian phyla but is absent in yeast, *Caenorhabditis elegans* and plants⁵. It contains a highly conserved DNA-binding domain with 11 zinc-fingers⁶; it is present at 55,000–65,000 sites in mammalian genomes⁷ and is normally located in linker regions surrounded by well-positioned nucleosomes⁸. Of these sites, ~5,000 are ultraconserved between mammalian species and tissues, and correspond to high-affinity sites⁹, whereas 30–60% of CTCF-binding sites show cell-type-specific distribution^{8,10–12}. The location of CTCF-binding sites with respect to genomic features provides insights into the possible roles of this protein. Approximately 50% of CTCF-binding sites are found in intergenic regions, ~15% are located near promoters and ~40% are intragenic (that is, within exons and introns)^{7,12} (FIG. 1). Surprisingly, and in view of the original role attributed to CTCF as an enhancer blocker, enhancer elements are enriched for this protein^{13,14}, which indicates that a subset of CTCF-binding sites may be important in

Department of Biology,
Emory University,
1510 Clifton Rd NE, Atlanta,
Georgia 30322, USA.
Correspondence to V.G.C.
e-mail: vcorces@emory.edu
doi:10.1038/nrg3663
Published online
11 March 2014

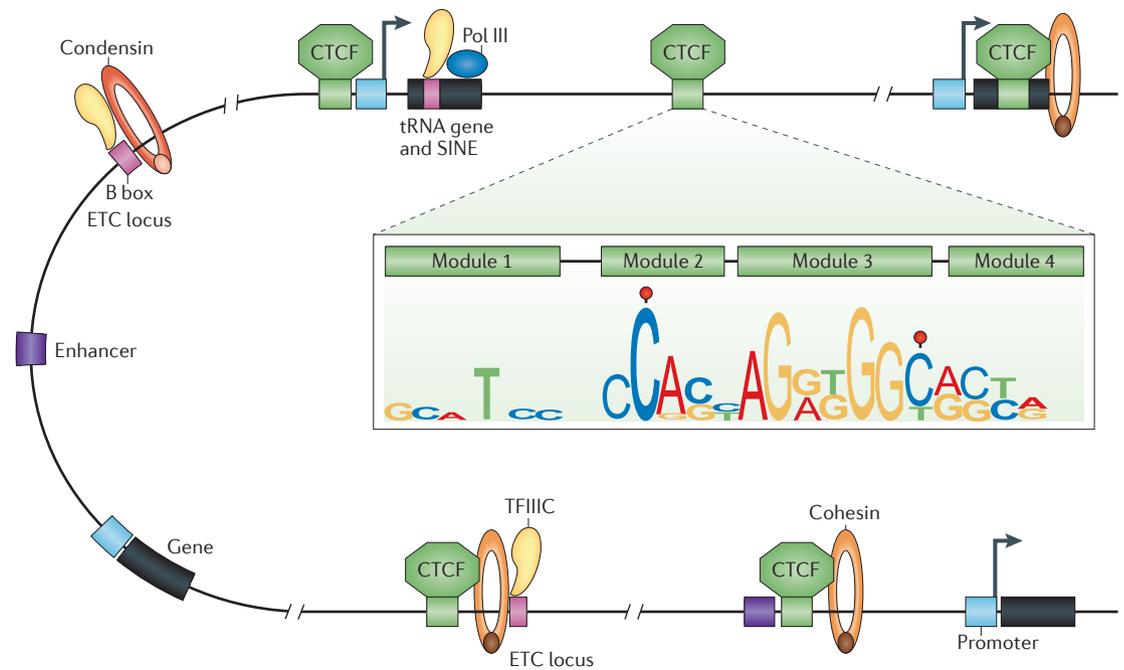


Figure 1 | Features of CTCF-binding sites in the genome. Binding sites of CCCTC-binding factor (CTCF) are associated with different genetic elements. The majority of these sites are intergenic and colocalize with cohesin. In addition, a proportion of CTCF-binding sites are located near RNA polymerase III (Pol III) type II genes (for example, tRNA genes and short interspersed nuclear elements (SINEs)) and ‘extra-TFIIC’ (ETC) loci, which suggests that TFIIC and CTCF cooperate in some nuclear processes. The 12-bp consensus sequence of CTCF-binding sites is embedded within binding modules 2 and 3 as determined by ChIP-exo experiments. DNA methylation (represented by red circles) of cytosine residues occurs at positions 2 and 12 of the consensus sequence in a subset of CTCF-binding sites.

regulating transcription to establish cell-lineage-specific programmes. Experiments using the ChIP-exo technique uncovered a 52-bp CTCF-binding motif that contains four CTCF-binding modules^{15,16} (FIG. 1).

The presence of CpGs in the DNA consensus sequence of the CTCF-binding site supports the notion that methylation of cytosine residues at carbon 5 of the base to form 5-methylcytosine (5mC) in CpG-containing sites may, at least partly, underlie CTCF target selectivity in different cell types¹⁷. Recent studies indicate that DNA methylation has a widespread role in regulating CTCF occupancy at many genes, including *CDKN2A* (which encodes INK4A and ARF)¹⁸, B-cell CLL/lymphoma 6 (*BCL6*)¹⁹ and brain-derived neurotrophic factor (*BDNF*)²⁰. One study has mapped the occupancy of CTCF in 19 human cell types; by comparing this information with DNA methylation data from parallel reduced representation bisulphite sequencing, it was found that 41% of cell-type-specific CTCF-binding sites are linked to differential DNA methylation²¹ (FIG. 2). Conversely, at 67% of sites that showed variability in DNA methylation, the presence of 5mC was associated with a concomitant downregulation of cell-type-specific CTCF occupancy. CTCF can also affect the methylation status of DNA by forming a complex with poly(ADP-ribose) polymerase 1 (PARP1) and DNA (cytosine-5)-methyltransferase 1 (DNMT1). CTCF activates PARP1, which can then inactivate DNMT1 by poly(ADP-ribosylation), and thus

maintains methyl-free CpGs in the DNA^{22,23}. An additional level of complexity in the interaction between CTCF and its target sequence can arise from the oxidation of 5mC to 5-hydroxymethylcytosine (5hmC)^{24,25}, 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)²⁶ by ten-eleven translocation (TET) enzymes. Genome-wide profiling analyses of 5hmC have shown that this modification and, to a lesser extent, 5fC are enriched genomic locations that contain CTCF-binding sites^{27,28}. Furthermore, identification of proteins that bind to different oxidized derivatives of 5mC discovered CTCF as a 5caC-specific binder²⁹. These results underscore the complexity and possible importance of the relationship between DNA methylation status and plasticity of CTCF occupancy. However, the presence of cell-type-specific CTCF-binding sites that are not differentially methylated suggests the existence of other mechanisms by which the DNA occupancy of this protein is regulated (FIG. 2).

One such mechanism is post-translational covalent modification of CTCF, such as sumoylation³⁰ and poly(ADP-ribosylation)³¹. In breast cancer cells, defective poly(ADP-ribosylation) of CTCF leads to its dissociation from the *CDKN2A* locus, which results in aberrant silencing of this tumour suppressor gene³². In *Drosophila melanogaster*, poly(ADP-ribosylation) of Centrosomal protein 190 kDa (Cp190) and CTCF facilitates their interaction, tethering to the nuclear matrix and intrachromosomal contacts³³.

ChIP-exo

An extension of chromatin immunoprecipitation followed by sequencing (ChIP-seq) that includes exonuclease trimming after ChIP to increase the resolution of the mapped transcription factor bound sites.

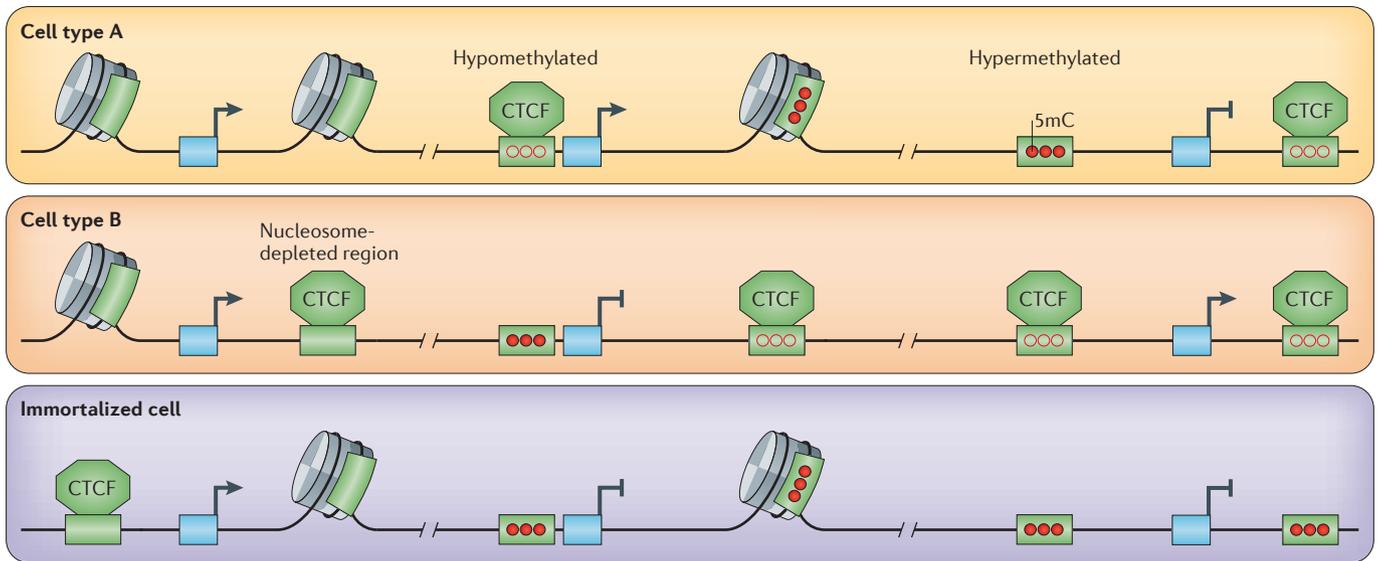


Figure 2 | Regulation of CTCF binding to DNA. Constitutive binding sites of CCCTC-binding factor (CTCF), which are bound by CTCF in cells from different tissues, are present in non-methylated and nucleosome-free regions. Cell-type-specific CTCF binding is partly regulated by differential DNA methylation and nucleosome occupancy across different cell types. This suggests that cells can use ATP-dependent chromatin remodelling complexes to regulate nucleosome occupancy at specific CTCF-binding sites and control the interaction of this protein with DNA. In addition, the methylation status of cell-type-specific CTCF-binding sites may be determined by a combination of activities of *de novo* methyltransferases and ten-eleven translocation (TET) enzymes that regulate the presence and levels of 5-methylcytosine (5mC) at specific sites. Immortalized cancer cell lines contain high levels of 5mC at CTCF-binding sites, which correlates with the low CTCF occupancy in these cells. Filled red circles represent methylated DNA, and open circles denote unmethylated DNA.

Interactions between CTCF and other proteins may represent an additional strategy by which the function of this protein can be regulated at various genomic locations during cell differentiation. Although different proteins — including transcription factor YY1, transcriptional regulator Kaiso, chromodomain helicase DNA-binding protein 8 (CHD8), PARP1, MYC-associated zinc-finger protein (MAZ), JUND, zinc-finger protein 143 (ZNF143), PR domain zinc-finger protein 5 (PRDM5) and nucleophosmin — have been implicated in CTCF function at specific loci³⁴, only cohesin has been shown to be required to stabilize most CTCF-mediated chromosomal contacts and to be essential for CTCF function at most sites in the genome^{35–41}. Interaction between CTCF and the cohesin complex takes place through the carboxy-terminal region of CTCF and the SA2 subunit of cohesin³⁵. Similar to CTCF, cohesin is present in intergenic regulatory regions, promoters, introns and 5' untranslated regions (5'UTRs) of genes during interphase of the cell cycle. Depending on the cell line, 50–80% of CTCF-binding sites in the genome are also occupied by cohesin, and downregulation of cohesin using RNA interference results in disruption of CTCF-mediated intrachromosomal interactions^{42–44}. A second protein that may cooperate with CTCF at a subset of sites in the genome is TFIIC, which is required for the transcription of tRNAs, 5S ribosomal RNA (rRNA), B2 short interspersed nuclear elements (SINEs) and other non-coding RNAs (ncRNAs) by RNA polymerase III (Pol III)⁴⁵. TFIIC also binds to many genomic

sites devoid of Pol III that are called 'extra-TFIIC' (ETC) loci (FIG. 1). In yeast, both the tRNA genes and ETC loci have been shown to cluster and tether DNA sequences to the nuclear periphery^{45,46}. Furthermore, B2 SINEs and human tRNA genes, both of which contain TFIIC-binding sites, can act as enhancer-blocking insulators in transgenic assays^{47,48}, and genome-wide analyses revealed that CTCF and its binding partner cohesin are found in the vicinity of many tRNA genes and ETC loci in mouse⁴⁹ and human cells^{50,51} (FIG. 1).

In addition, several observations suggest that RNAs can also cooperate with CTCF to stabilize its interactions with other proteins. *D. melanogaster* architectural proteins, such as Cp190, require the RNA helicase Rm62 for proper function, and interactions between Rm62 and Cp190 depend on the presence of RNA⁵². Similar observations have been made in mammalian cells, in which CTCF has been shown to interact with the DEAD-box RNA helicase p68 and its associated ncRNA, both of which are required for proper CTCF function⁵³. These observations, together with new findings which indicate that CTCF can itself bind to the *Jpx* RNA⁵⁴, support the idea that ncRNAs may have an important role in stabilizing interactions that are mediated by CTCF and its protein partners.

Mechanisms of CTCF function

A large body of evidence strongly supports the idea that the mechanism underlying the diverse functions of CTCF in genome biology is based on its ability to

mediate long-range interactions between two or more genomic sequences. This evidence first came from 3C analyses of the mouse *H19*–insulin-like growth factor 2 (*Igf2*) and haemoglobin subunit beta (*Hbb*) loci. At the imprinted maternal *H19*–*Igf2* locus, work using circular chromosome conformation capture (4C) indicated that the *H19* imprinting control region (ICR) forms extensive interchromosomal and intrachromosomal interactions across the genome, many of which require the presence of CTCF-binding sites within the ICR⁵⁵. Binding of CTCF at multiple DNase I hypersensitive sites is also required to maintain a specific chromatin architecture at the murine *Hbb* locus⁵⁶. Similarly, the *D. melanogaster* architectural proteins Suppressor of Hairy wing (Su(Hw)), Boundary element-associated factor of 32 kDa (BEAF-32) and Zeste-white 5 (Zw5; also known as Dwg) were shown to mediate long-range interactions both at the *Heat-shock-protein-70* (*Hsp70*) locus and between copies of the *gypsy* retrotransposon^{57,58}.

Taken together, these observations suggest that the effect of CTCF and other architectural proteins on gene expression is a result of their ability to bring sequences that are far apart in the linear genome into close proximity. Assuming that this mechanism underlies all functions of CTCF, results of both locus-specific and genome-wide studies can be interpreted to suggest that CTCF-mediated contacts regulate aspects of genome function in a context-dependent manner; that is, the functional outcomes of these interactions depend on the nature of the sequences adjacent to CTCF-binding sites and perhaps on the presence of other specific chromatin proteins. Below, we critically analyse existing evidence for these various roles in an attempt to generate a model that reflects the function of this protein in its normal genomic context.

The classical roles of CTCF

CTCF as a chromatin barrier. The function of CTCF and other architectural proteins was initially analysed using transgenic assays, in which results were often interpreted to suggest that these proteins act as barriers to the progressive spread of heterochromatin⁵⁹. Indeed, sequences with bona fide barrier activity in yeast have been shown to recruit histone acetyltransferases that antagonize the spreading of silencing histone modifications⁶⁰. However, although this interpretation of experimental results is often used to provide an explanation for those obtained in higher eukaryotes, recent studies now offer a different view of how architectural proteins may regulate gene expression. For example, one study has demonstrated the presence of a CTCF-dependent enhancer-blocking function but a CTCF-independent barrier function in the chicken HS4 insulator. In this case, the barrier function depends on upstream stimulatory factor 1 (USF1), which recruits histone acetyltransferases⁶¹.

Results from genome-wide studies of the localization of CTCF in relation to various histone modifications also do not offer strong support for a role of CTCF as a barrier. Human CD4⁺ T cells and HeLa cells contain ~30,000 domains of histone H3 trimethylated at lysine 27 (H3K27me3, which is a histone modification

characteristic of silenced chromatin), of which ~1,600 and ~800 contain CTCF at one of the domain borders, respectively⁸. This represents only 2–4% of the domain borders, which is a relatively low number if CTCF was primarily involved in the establishment or maintenance of these silenced domains. Similarly, it has been suggested that CTCF can contribute to the formation of chromosomal domains that are associated with the nuclear lamina. Human lung fibroblasts contain 2,688 borders that flank lamin-associated domains (LADs). These domains are enriched in oriented promoters, CpG islands and CTCF-binding sites. Approximately 9% of LAD borders contain CTCF (245 of 2,688; 120 additional borders contain a combination of promoters, CpGs and/or CTCF) within 10 kb of the boundary⁶². Although the correlation is tantalizing, it is unlikely that CTCF is the main contributor to the formation of these borders or that this is one of the primary functions of this protein.

Despite the lack of strong evidence from genome-wide localization analyses, the role of CTCF has been interpreted as a domain barrier in several studies. Recent mapping of CTCF-mediated intrachromosomal and interchromosomal interactions in mouse embryonic stem cells (ESCs) using chromatin interaction analysis with paired-end tag sequencing (ChIA-PET), which combines chromatin immunoprecipitation (ChIP) with 3C analyses, seems to support the notion that CTCF may be, at least partly, responsible for the establishment of functional expression domains⁶³. A total of 1,480 CTCF-containing *cis*-interacting loci were identified by this strategy. Cluster analyses of intrachromosomal interactions with seven histone modification signatures and Pol II occupancy profiles uncovered four distinct categories of CTCF-mediated loops. One class (155 of 1,295 loops that are <1 Mb; 12%) contains active H3K4me1, H3K4me2 and H3K36me3 histone modifications inside but repressive H3K9me3, H3K20me3 and H3K27me3 modifications outside the loops. A second class (142 of 1,295; 11%) of loops has the reverse pattern of histone modifications. Although the evidence is only correlative and the number of CTCF-mediated interactions is much smaller than the total number of CTCF-binding sites in the genome, the existence of these two types of loops in which CTCF flanks histone modifications that have opposite effects on transcription is a suggestive, but not conclusive, proof of a role for this protein in separating functional domains of gene expression.

Some examples of locus-specific analyses also seem to support this view of CTCF-mediated barrier activity. For example, the Wilms tumour 1 homologue (WT1) transcription factor can either activate or repress expression of the mouse wingless-related MMTV integration site 4 (*Wnt4*) gene in a cell-type-specific manner by controlling the state of chromatin in a domain that has CTCF-defined boundaries. Mutation of CTCF leads to spreading of histone modifications outside the delimited genomic domain, which causes aberrant expression of neighbouring genes and suggests a role for CTCF in the establishment or maintenance of the *Wnt4* domain by creating a functional barrier⁶⁴.

Circular chromosome conformation capture (4C). The combination of inverse PCR and high-throughput sequencing with the chromosome conformation capture (3C) technique that allows the profiling of chromatin interactions between a known specific locus and multiple unknown sites.

DNase I hypersensitive sites
Chromosomal regions that are readily degraded by deoxyribonuclease I (DNase I) owing to decreased nucleosome occupancy. These sites are associated with open chromatin conformation and the binding of transcription factors.

Nuclear lamina
A scaffold of proteins comprised mainly of lamin A/C and B that is predominantly found in the nuclear periphery; it is associated with the inner surface of the nuclear membrane.

Chromatin interaction analysis with paired-end tag sequencing (ChIA-PET). A technique used to determine the chromosomal interactions that are mediated by a specific chromatin-binding protein by combining chromatin immunoprecipitation with a chromosome conformation capture (3C)-type analysis.

Although the ChIA-PET and *Wnt4* results are best explained by assuming that CTCF is capable of barrier activity, a similar explanation of results from studies at other loci may seem to be less straightforward. For example, groups of androgen-responsive genes that are demarcated by CTCF-binding sites tend to have similar epigenetic and expression profiles, which suggests that CTCF establishes domains in which these genes are co-regulated⁶⁵. Downregulation of CTCF results in a decrease in the expression of genes within the domain, whereas genes outside the domain are unaffected; this can be explained if CTCF is involved in targeting regulatory sequences to androgen-responsive promoters and, in its absence, transcription of these genes decreases. Similarly, the mouse homeobox A cluster (*Hoxa*) forms two distinct chromatin loops around CTCF-binding site 5 (CBS5) as ESCs differentiate into neural progenitor cells. The loop that contains the *Hoxa1–7* gene cluster upstream of CBS5 is marked by active H3K4me3 modifications, whereas the loop containing the downstream *Hoxa9–13* gene cluster is enriched for repressive H3K27me3 marks⁶⁶. Knockdown of CTCF results in the loss of three-dimensional conformation and the concomitant spread of H3K27me3 modifications across the locus. These results can be explained on the basis of a barrier function for CTCF, but it is equally possible that CTCF-binding sites in the *Hoxa* locus can participate in bringing together regulatory sequences for gene activation or repression. This explanation is supported by results obtained in *D. melanogaster*, in which the role of CTCF in the maintenance of H3K27me3-enriched domains that are delimited by CTCF and other architectural proteins was analysed in detail. When CTCF was knocked down, H3K27me3-enriched domains showed a significant reduction in the level of this histone modification within the domain. However, little or no spreading of H3K27me3 was observed outside the demarcated domains^{67,68}. This suggests that *D. melanogaster* CTCF helps to maintain the level of silencing within domains, but not its spreading, presumably by clustering H3K27me3-enriched loci and Polycomb group proteins into Polycomb bodies^{67,69}. On the basis of these results, we suggest that there is little causal evidence to support a generalized functional role for CTCF in separating domains with different epigenetic marks. Instead, alternative mechanistically different processes, such as those involving looping between regulatory sequences, may provide a better explanation for some of the observations that were previously interpreted in this context.

CTCF as an enhancer blocker. Although CTCF has been extensively characterized for its ability to block enhancer activity in transgenic assays, there has been little evidence to support such a role for this protein in its normal genomic context. However, some recent studies suggest that CTCF can indeed act as an enhancer blocker at specific loci. For example, induction of the *Ecdysone-induced protein 75B* (*Eip75B*) gene by treatment of *D. melanogaster* Kc cells with the steroid hormone ecdysone results in the downregulation

of one of the *Eip75B* transcripts that is expressed from an alternative upstream promoter. This is caused by activation of a poised CTCF-binding site by recruitment of Cp190, which increases its interaction with a distant CTCF-binding site and topologically separates the downregulated promoter of the *Eip75B* gene from its enhancer⁷⁰. Some genome-wide studies also suggest an enhancer-blocking function for CTCF. A search for conserved regulatory motifs in the human genome led to the finding of 15,000 CTCF-binding sites that separate adjacent genes which show markedly reduced correlation in gene expression when compared with genes that are in a similar arrangement but that are not separated by CTCF-binding sites⁷¹. A similar observation has been made for the *D. melanogaster* BEAF-32 protein⁷², which suggests that CTCF and other architectural proteins can allow neighbouring gene pairs to be differentially regulated. However, the classical enhancer-blocking function of CTCF seems to contradict more recent results that support a function for CTCF as a facilitator of enhancer function. Below, we describe in detail some of this new information to underscore the widespread, but not widely acknowledged, role for CTCF as a positive regulator of various transcriptional processes. Ultimately, models that explain how CTCF controls gene expression need to account for these two apparently contradictory functions — enhancer blocker and enhancer facilitator — of this protein.

An updated view of CTCF function

CTCF helps to tether distant enhancers to their promoters. Recent observations seem to contradict the idea of enhancer blocking as a predominant role for CTCF. For example, one study examined interactions between promoters and their regulatory sequences using the chromosome conformation capture carbon copy (5C) technique and found that 79% of long-range interactions between distal elements and promoters are not blocked by the presence of one or more intervening CTCF-bound sites⁷³. Instead, a proportion of these interacting distal elements are significantly enriched for CTCF and/or histone modifications that are characteristic of active enhancers (that is, H3K4me1, H3K4me2 and H3K27 acetylation (H3K27ac)), which strongly supports the concept that one of the main roles of CTCF in genome function may be to facilitate the interaction between regulatory sequences and promoters. Activation of transcription requires the assembly of specific activators, the Mediator complex and the basal transcription machinery in a process that involves long-range chromosomal interactions between distal enhancers and proximal promoter elements. An enrichment of CTCF-binding sites at promoters and intergenic regions has been observed in ChIP followed by sequencing (ChIP-seq) studies, which also suggests that one of the main functions of CTCF is to target regulatory elements to their cognate promoters. This conclusion is supported by the finding of a significant overlap between cell-type-specific CTCF-binding sites and enhancer elements⁷⁴, as well as by studies at several individual loci. For example, CTCF-mediated topological organization of the major histocompatibility

Chromosome conformation capture carbon copy (5C). A technique used to profile all chromatin interactions in specific regions of the genome by the hybridization of a mixture of DNA primers to chromosome conformation capture (3C) templates followed by high-throughput sequencing.

Mediator complex

The ~30-subunit co-activator complex that is required for successful transcription of RNA polymerase II (Pol II) promoters of metazoans genes. Its interaction with Pol II and site-specific factors facilitates enhancer–promoter communication.

complex class II (MHC-II; also known as HLA-D) locus precedes transcriptional activation⁷⁵. Activation of MHC-II gene expression by interferon- γ (IFN γ) treatment requires the looping of the XL9 enhancer element and its cognate promoters that is mediated by CTCF, MHC class II transactivator (CIITA) and specific transcription factors⁷⁶.

CTCF has also been shown to be important in regulating the expression of complex gene clusters in which regulatory sequences are far from some of their target genes. For example, in human islets, CTCF maintains long-range interactions between the insulin (*INS*) and synaptotagmin 8 (*SYT8*) genes that are necessary for *SYT8* transcription⁷⁷. In the mammalian brain, neuronal diversity is attained through a combination of stochastic promoter choice and alternative pre-mRNA processing of the protocadherin (*PCDH*) genes. Each *PCDH* mRNA contains a variable 5' exon followed by a common region. The *PCDH* gene cluster is comprised of more than 50 different 5' exons, each preceded by its own promoter (FIG. 3). CTCF and cohesin bind to most of these promoters⁷⁸ and the distant enhancer element HS5-1 (REF. 79). Alternative isoform expression requires CTCF-mediated DNA looping between the HS5-1 enhancer and active *PCDHA* promoters^{80,81} (FIG. 3). Conditional knockout of CTCF in mouse postmitotic projection neurons leads to reduced expression of *PCDH* genes, neuronal defects

and abnormal behaviour, which suggests that CTCF is required to tether the HS5-1 enhancers to the various promoters⁸².

A third recent example that underscores the role of CTCF in promoting enhancer–promoter interactions comes from studies in mouse ESCs, in which the TATA-binding protein-associated factor 3 (TAF3) — a component of the core promoter-recognition complex TFIID — is required for endodermal differentiation. In addition to promoters, TAF3 localizes to distal sites that contain CTCF and cohesin, and the two sequences form a loop in a TAF3-dependent manner⁸³ (FIG. 4). Given the role of TAF3 in regulating lineage commitment in ESCs, the distal elements that contain binding sites for both CTCF and TAF3 might have acquired H3K4me1 and H3K4me2 pre-patterning in ESCs to become endodermal enhancers, thus supporting the idea that CTCF can tether distal regulatory sequences to their target promoters.

Observations from genome-wide analyses of intra-chromosomal interactions also support a role of CTCF in facilitating contacts between transcription regulatory sequences. An analysis of CTCF-mediated interactions using ChIA-PET in mouse ESCs suggests that this protein is involved in clustering promoters of different genes, perhaps to establish ‘transcription factories’. Interestingly, 28% of genes with promoters that are brought into close proximity (<10 kb) to p300 sites by CTCF-mediated

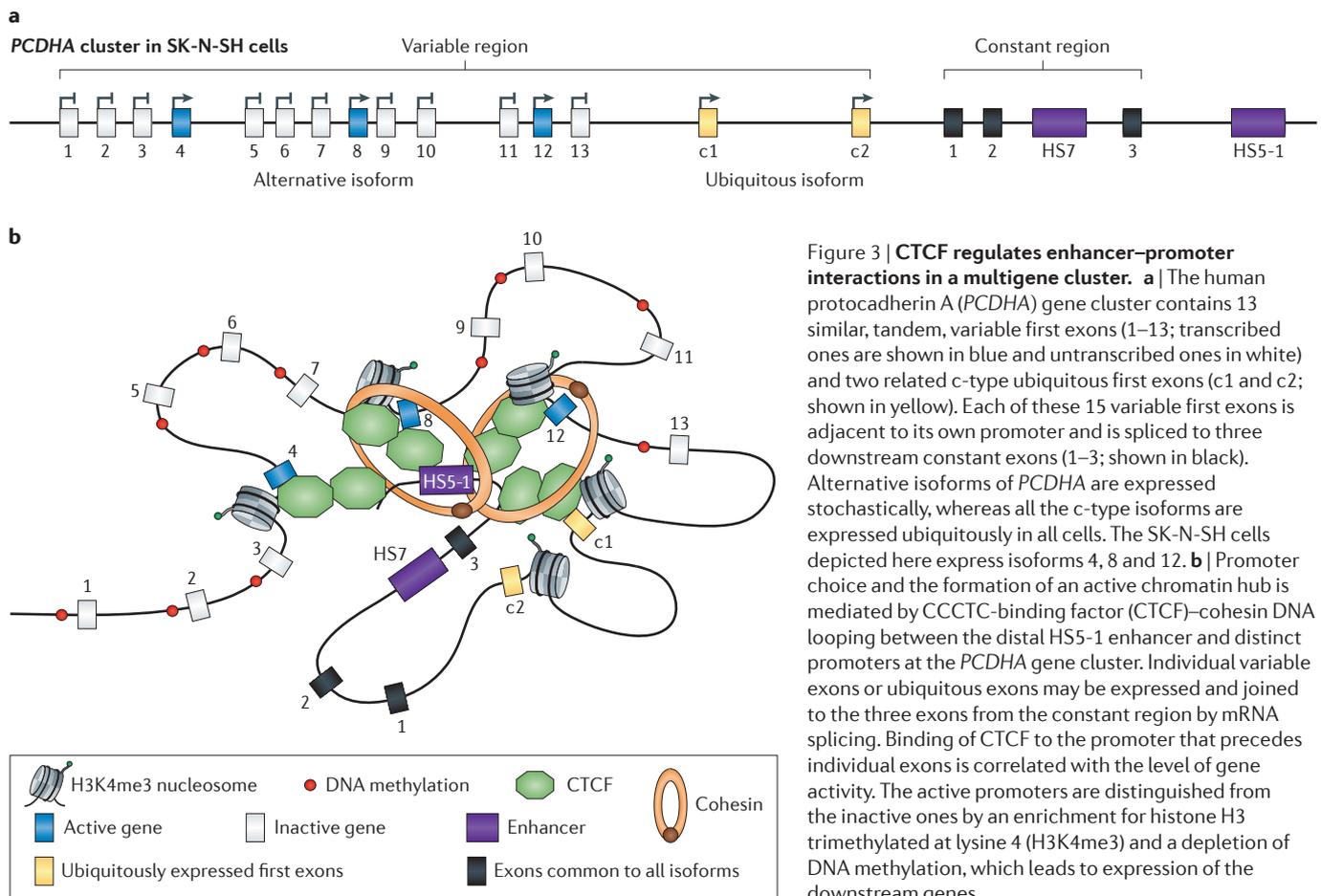


Figure 3 | CTCF regulates enhancer–promoter interactions in a multigene cluster. **a** | The human protocadherin A (*PCDHA*) gene cluster contains 13 similar, tandem, variable first exons (1–13; transcribed ones are shown in blue and untranscribed ones in white) and two related c-type ubiquitous first exons (c1 and c2; shown in yellow). Each of these 15 variable first exons is adjacent to its own promoter and is spliced to three downstream constant exons (1–3; shown in black). Alternative isoforms of *PCDHA* are expressed stochastically, whereas all the c-type isoforms are expressed ubiquitously in all cells. The SK-N-SH cells depicted here express isoforms 4, 8 and 12. **b** | Promoter choice and the formation of an active chromatin hub is mediated by CCCTC-binding factor (CTCF)–cohesin DNA looping between the distal HS5-1 enhancer and distinct promoters at the *PCDHA* gene cluster. Individual variable exons or ubiquitous exons may be expressed and joined to the three exons from the constant region by mRNA splicing. Binding of CTCF to the promoter that precedes individual exons is correlated with the level of gene activity. The active promoters are distinguished from the inactive ones by an enrichment for histone H3 trimethylated at lysine 4 (H3K4me3) and a depletion of DNA methylation, which leads to expression of the downstream genes.

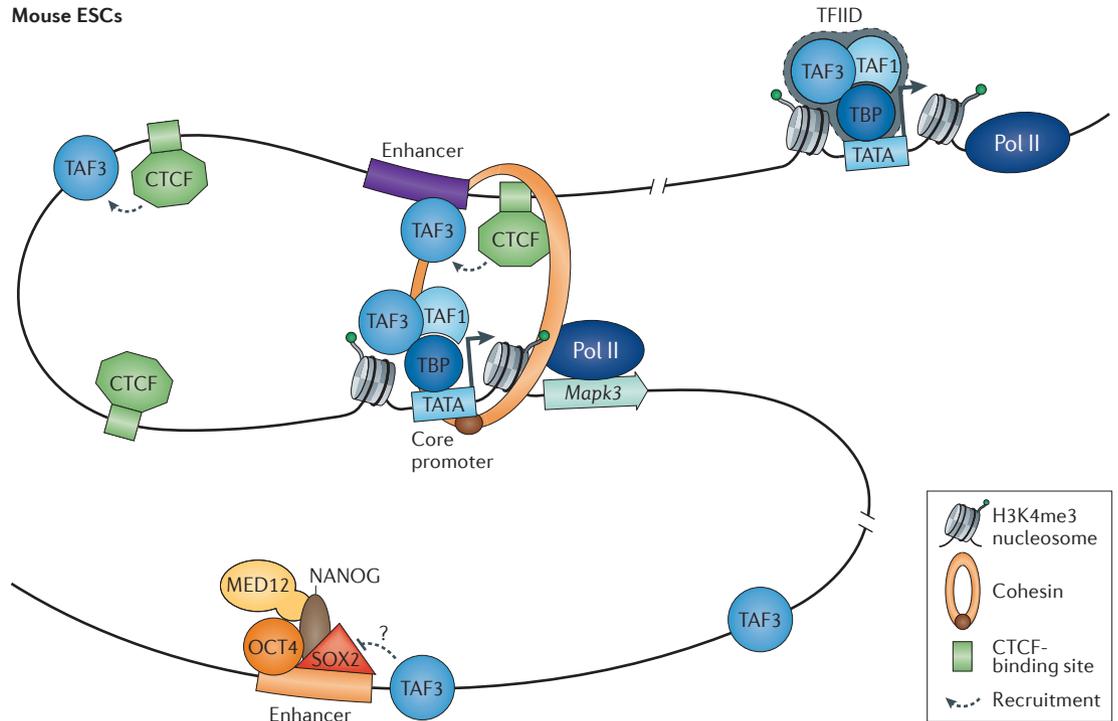


Figure 4 | CTCF facilitates endodermal enhancer–promoter interactions in ESCs. Recruitment of TATA-binding protein-associated factor 3 (TAF3) at endodermal enhancers by CCCTC-binding factor (CTCF) and chromatin looping activates the mitogen-activated protein kinase 3 (*Mapk3*) gene in mouse embryonic stem cells (ESCs). Apart from being a component of TFIIID at core promoters, TAF3 may also associate with other transcription factors across the genome in ESCs. For example, TAF3 represses the activity of pluripotency-associated transcription factors (octamer-binding protein 4 (OCT4; also known as POU5F1), SOX2 and homeobox protein NANOG). H3K4me3, histone H3 trimethylated at lysine 4; MED12, Mediator complex subunit 12; Pol II, RNA polymerase II; TBP, TATA-binding protein.

Adaptive response

The acquired immune response to the specific antigen presented on a pathogen that typically triggers immunological memory.

Pro-B cells

B cells at their earliest developmental stage in the bone marrow that are defined as the CD19⁺ cytoplasmic IgM⁻ or B220⁺CD43⁺ population and that have incomplete rearrangement of the immunoglobulin heavy chain.

Double-positive thymocytes

Immature T cells characterized by the expression of CD4 and CD8 cell surface markers that will differentiate into single-positive thymocytes after their T cell receptors interact with self-peptide major histocompatibility complex ligands in the thymus.

Pre-pro-B cells

The lymphoid progenitors found in the bone marrow that contain the CLP-2s surface marker and lack heavy-chain diversity (D)–joining (J) rearrangements.

contacts are upregulated in mouse ESCs, and knock-down of CTCF results in downregulation of some of these genes, which supports the notion that CTCF may be involved in mediating enhancer–promoter interactions during transcription initiation⁶³.

CTCF regulates recombination at the antigen receptor loci. The role of CTCF in mediating enhancer–promoter communication may also contribute to the regulation of other nuclear processes such as V(D)J recombination. The B cell immunoglobulin (Ig) and T cell receptor (Tcr) loci comprise multiple copies of variable (V), diversity (D), joining (J) and constant (C) gene segments that span across large genomic regions (FIG. 5). During the adaptive response, unique epigenetic features and three-dimensional chromatin architecture at these loci provide the framework for recombinase-activating gene (RAG)-mediated DNA recombination of the gene segments to generate antigen receptor diversity⁸⁴. Although CTCF-mediated long-range chromatin interactions are not essential for the progression of V(D)J recombination, they may influence lineage- and/or developmental stage-specific segment choice during recombination.

Looping between distant CTCF-binding sites may bring distant gene segments together. In mouse pro-B cells, chromatin looping of CTCF-binding sites at the immunoglobulin heavy chain complex (*Igh*) locus

occurs independently of the E μ enhancer and contributes to the compaction of the locus^{85,86} (FIG. 5). In double-positive thymocytes, CTCF-mediated looping between the E α enhancer and specific promoters within the *Tcra–Tcrd* locus facilitates V α –J α over V δ –D δ –J δ rearrangement⁸⁷. By establishing interactions between specific sequences, CTCF may also impede other sequences from contacting each other. In fact, this may be the basis for the enhancer-blocking function of CTCF. In the *Igh* locus, two CTCF-binding sites within intergenic control region 1 (IGCR1) mediate ordered and lineage-specific V $_H$ –D $_H$ recombination and bias distal over proximal V $_H$ rearrangements⁸⁸. Positioned between the V $_H$ and D $_H$ clusters, IGCR1 suppresses the transcriptional activity and the rearrangement of proximal V $_H$ segments by forming a CTCF-mediated loop that presumably isolates the proximal V $_H$ promoter from the influence of the downstream E μ enhancer (FIG. 5). Similarly, in pre-pro-B cells, CTCF promotes distal over proximal V $_k$ rearrangement by blocking the communication between specific enhancer and promoter elements in the *Igk* locus⁸⁹.

CTCF regulates transcriptional pausing and alternative mRNA splicing. The existence of a proportion of CTCF-binding sites in the 5'UTR and introns of genes suggests a role for CTCF in regulating transcriptional events downstream of the initiation step. Indeed, recent studies

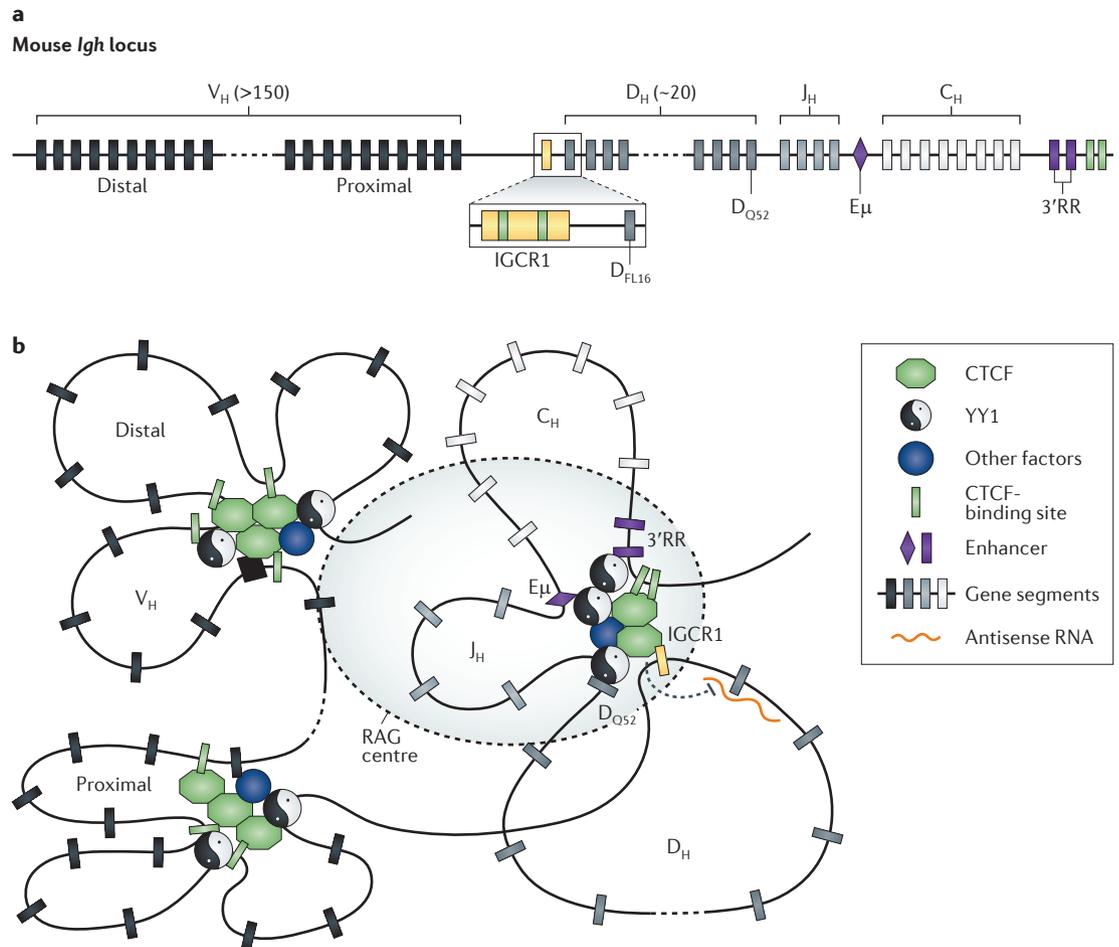


Figure 5 | CTCF regulates V(D)J recombination. V(D)J recombination at antigen receptor loci is regulated by chromatin accessibility, which correlates with active histone modifications and transcription. CCCTC-binding factor (CTCF) may influence the outcome of V(D)J recombination by regulating enhancer–promoter interactions and locus compaction. **a** | Organization of the mouse immunoglobulin heavy chain complex (*Igh*) locus is shown. **b** | CTCF-mediated looping of diversity (D_H)–joining (J_H)–constant (C_H) segments imposes ordered (that is, D_H – J_H) recombination by controlling the communication of enhancers (E_{μ} and 3' regulatory region (3'RR)) with distinct gene segments. Binding of CTCF at intergenic control region 1 (IGCR1) blocks the influence of the E_{μ} enhancer on proximal variable (V_H) segments and prevents the spread of active histone modifications from D_H into the proximal V_H region. In addition, it inhibits the level of antisense transcription within the D_H region and modulates locus compaction in collaboration with other factors (for example, transcription factor YY1, DNA-binding protein Ikaros, paired box protein PAX-5 and transcription factor 3 (TCF3; also known as E2A)). As a consequence, CTCF within IGCR1 may bias the rearrangement of distal over proximal V_H segments with DJ_H joins. RAG, recombinase-activating gene.

indicate that CTCF can control both pausing of Pol II and alternative mRNA splicing. For example, CTCF binds to both the first intron and upstream regulatory elements in the mouse myeloblastosis oncogene (*Myb*) locus. During erythroid differentiation, looping between the first intron, promoter and upstream enhancer elements that is mediated by CTCF and key erythroid transcription and elongation factors is required for Pol II-mediated transcriptional elongation and high expression of the *Myb* gene⁹⁰. This three-dimensional architecture is lost upon differentiation, when CTCF interferes with Pol II elongation at the first intron, which leads to low expression of *Myb*. In this case, the dual functions of CTCF in transcription initiation and pausing seem to rely on its ability to stabilize long-range

interactions with regulatory sequences and to impede the elongation of Pol II. The effect of CTCF on Pol II elongation may be widespread, given that the genome-wide presence of CTCF at promoter-proximal regions in 5'UTRs strongly correlates with high pausing indexes⁹¹. In other cases, hindering elongation of Pol II by CTCF may result in the inclusion or exclusion of specific exons in the mature mRNA. One example of this phenomenon occurs at the *CD45* gene in humans, which expresses alternatively spliced transcripts during lymphocyte differentiation. Binding of CTCF to exon 5 of the gene promotes its inclusion in the *CD45* mRNA, whereas disruption of CTCF binding results in exclusion of this exon. Interestingly, it seems that DNA methylation of CTCF recognition sequences in

exon 5 determines whether this protein binds to exon sequences, as knockdown of DNMT1 during late stages of lymphocyte differentiation leads to CTCF binding and inclusion of exon 5 in *CD45* transcripts⁹² (FIG. 6).

Genome topology may rationalize CTCF roles

Results from experiments that are aimed at mapping all interactions in the genome using Hi-C suggest that genomes of higher eukaryotes are organized into topologically associating domains (TADs), which are defined by a high frequency of interactions within domains and a low frequency of interactions between adjacent domains (FIG. 7). In *D. melanogaster*, TAD boundaries are gene-dense regions that are enriched for highly transcribed genes and clusters of architectural protein-binding sites, including those of CTCF, BEAF-32, Su(Hw), Modifier of *mdg4* (Mod(*mdg4*)), Chromator (Chro) and Cp190 (REFS 93,94). Similarly, TAD borders in mammals are enriched for binding sites of CTCF and double-strand break repair protein rad21 homologue (RAD21), housekeeping and tRNA genes, and SINES⁹⁵

(FIG. 7). The enrichment of CTCF and RAD21 at TAD borders may have a causal role in determining their establishment. This conclusion is supported by results from experiments in which a 58-kb region located at the border between the TADs of *Tsix* (X (inactive)-specific transcript, opposite strand) and *Xist* (inactive X specific transcripts) in the mouse X chromosome was deleted. Elimination of these sequences — which include a CTCF-binding site and the *Xist*, *Tsix* and regulatory region 18 (*Rr18*; also known as *Xite*) genes — leads to increased interactions in the previous inter-TAD border region and to the formation of a new TAD border at an adjacent location⁹⁶.

Recent studies using Hi-C to investigate the role of CTCF and cohesin in the three-dimensional organization of the mammalian genome support a similar role for CTCF as a boundary protein between TADs, although the details vary between the different studies^{97–99}. Depletion of cohesin in HEK293 human embryonic kidney cells results in a general loss of intrachromosomal interactions without affecting the TAD organization, whereas depletion of CTCF causes a similar decrease in the frequency of intradomain interactions concomitant with an increase in the frequency of interactions between adjacent TADs⁹⁷. Cohesin-deficient postmitotic mouse astrocytes also show a reduced number of long-range interactions that are mediated by CTCF and cohesin but additionally display a relaxation of TAD organization⁹⁸. This TAD relaxation could be a consequence of a reduction in TAD border strength due to the lack of cohesin binding or to an increase in the frequency of inter-TAD interactions as observed in CTCF-depleted HEK293 cells. A similar decrease in the frequency of cohesin-mediated interactions was observed in cohesin-depleted developing mouse thymocytes that were arrested in G1 phase, and an increase in the frequency of alternative interactions resulted in changes to gene expression⁹⁹.

In mammals, only 15% of genomic CTCF-binding sites are present at TAD borders, whereas the other 85% are present inside TADs⁹⁵; this indicates that CTCF and cohesin alone are insufficient to separate different TADs — a conclusion supported by the fairly mild effects on TAD organization in cells that are depleted of CTCF or cohesin. In *D. melanogaster*, CTCF forms clusters with other architectural proteins at TAD borders, and vertebrate CTCF might adopt a similar strategy. Several lines of evidence suggest TFIIC as a candidate architectural protein that cooperates with CTCF at TAD borders in vertebrates. As discussed above, TFIIC colocalizes with CTCF near many tRNA genes and ETC loci in mammalian cells^{49–51}, and it also binds to SINES and tRNA genes, both of which functionally behave as enhancer-blocking insulators in humans^{47,48} and are enriched at TAD borders⁹⁵. As CTCF has been shown to recruit cohesin in mammalian cells, and TFIIC interacts with both cohesin and condensin in yeast, CTCF and TFIIC might act as docking sites for these proteins to stabilize interactions that are required for the formation of TAD borders. Such borders do not allow cross-interactions between sequences in the two adjacent TADs. Thus, it is possible that sequences at these borders

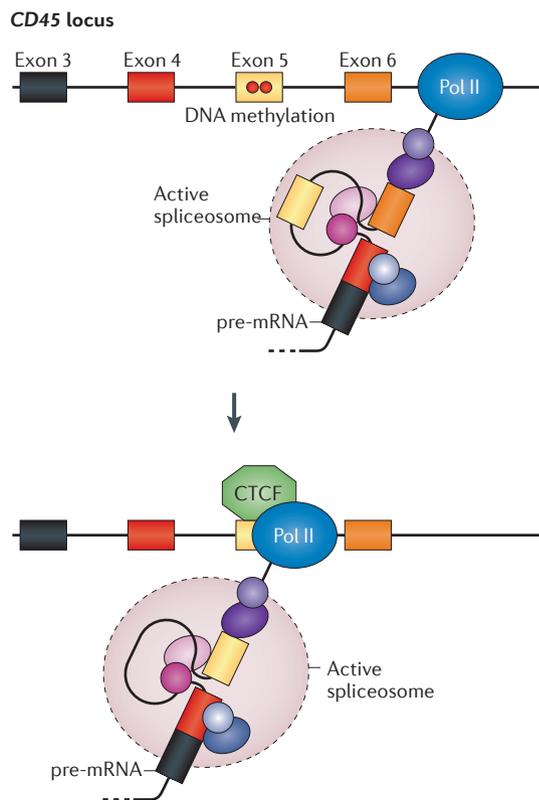


Figure 6 | CTCF promotes alternative mRNA splicing. Mutually exclusive DNA methylation and CCCTC-binding factor (CTCF) binding may regulate alternative splicing. At the *CD45* locus, DNA methylation (represented by red circles) at exon 5 inhibits CTCF binding, which leads to fairly unimpeded transcriptional elongation by RNA polymerase II (Pol II) and subsequent exclusion of exon 5 during splicing of the resultant mRNA (upper panel). By contrast, hypomethylation of exon 5 leads to CTCF binding and Pol II stalling, which promotes the inclusion of exon 5 (lower panel).

Hi-C
An extension of the chromosome conformation capture (3C) technique that incorporates a biotin-labelled nucleotide at the ligation junction to allow selective purification of chimeric DNA ligated products for high-throughput sequencing. This method generates matrices of interaction frequencies across the genome.

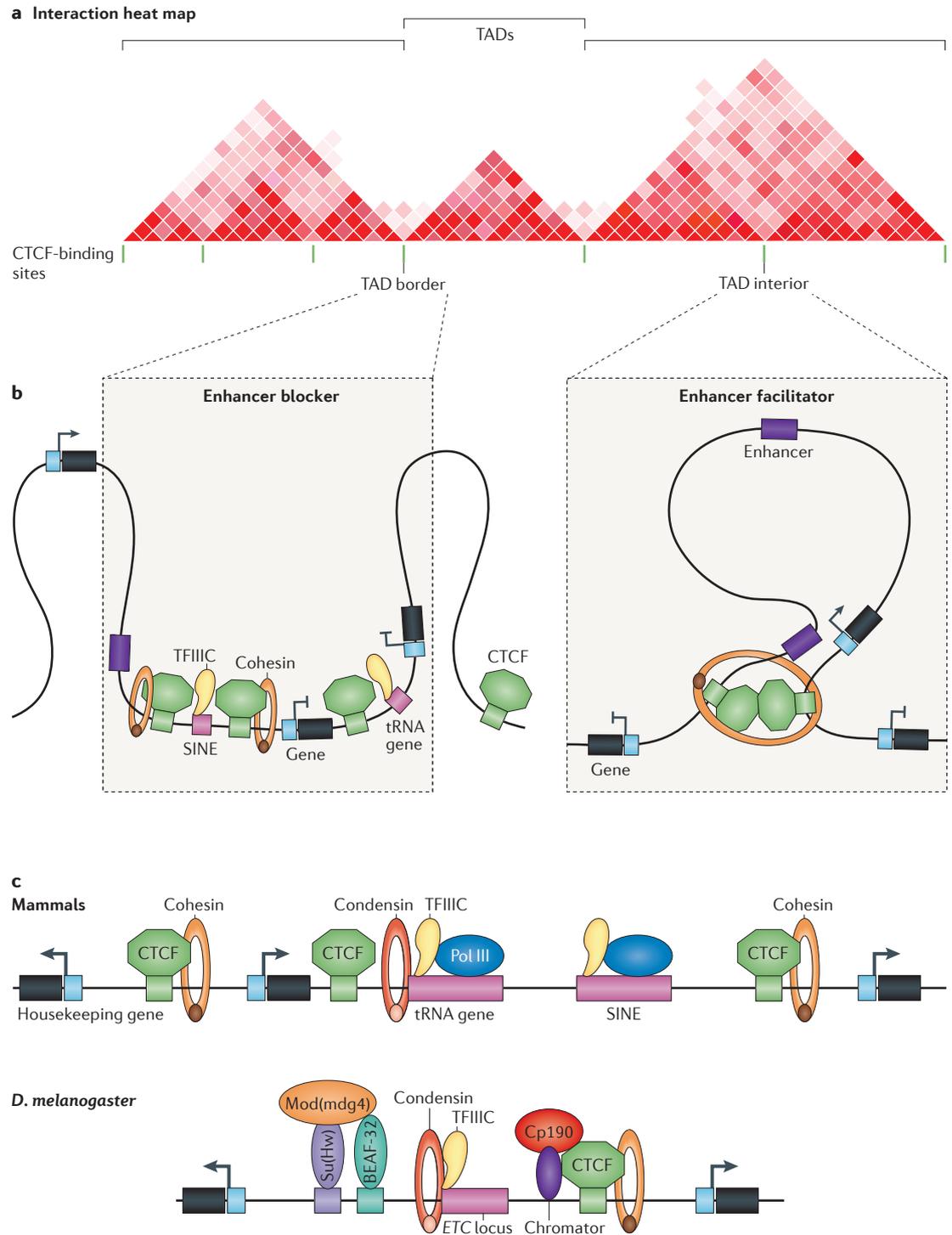


Figure 7 | CTCF regulates three-dimensional genome architecture. **a** | Schematic data generated by Hi-C in mammalian cells are shown in an interaction heat map of a ~2.5-Mb chromosomal segment. The topologically associating domains (TADs) and their borders are indicated. **b** | The presence of multiple binding sites for CTCF-binding factor (CTCF) and TFIIIC at TAD borders may contribute to the establishment of the border. This arrangement may provide an explanation for the observed function of CTCF as an enhancer blocker. Conversely, CTCF-binding sites within TADs may facilitate enhancer–promoter looping through the recruitment of cohesin. The blue box denotes the promoter of the gene. **c** | Chromatin features of TAD borders in mammals and *Drosophila melanogaster* are shown. The TAD borders in mammals are enriched for housekeeping and tRNA genes, short interspersed nuclear elements (SINEs) and CTCF-binding sites. In *D. melanogaster*, they are enriched for highly transcribed genes and clusters of binding sites for various architectural proteins, such as Suppressor of Hairy wing (Su(Hw)), Modifier of mdg4 (Mod(mdg4)) and Boundary element-associated factor of 32 kDa (BEAF-32). The roles of TFIIIC, cohesin and condensin proteins in mediating TAD border formation remain to be determined. Cp190, Centrosomal protein 190 kDa.

represent the enhancer-blocking insulators that have previously been characterized in transgenic assays (FIG. 7). Additional studies will be needed to clarify whether clustering of CTCF, TFIIC, cohesin and condensin occur at TAD boundaries of mammalian cells and whether their presence is required for border formation.

The majority of CTCF-binding sites (~85%) are found within TADs and are, by definition, unable to form a border. What is the role of CTCF at these sites? Studies in pre-pro-B cells using Hi-C suggest that CTCF located within TADs is primarily involved in mediating short-range intra-TAD interactions¹⁰⁰. As discussed above, the function of these CTCF-mediated interactions may be to direct enhancers within the TAD to the appropriate gene promoter. In large mammalian genomes, the resolution of Hi-C data is limited by the number of sequencing reads, which restricts the amount of structural information that can be obtained at the sub-TAD level. The use of 5C over large (1–2 Mb) genomic regions has made possible the mapping of finer topologies at the sub-megabase scale¹⁰¹. These topologies originate from interactions that are mediated by CTCF, cohesin and Mediator either alone or in various combinations. Many of these interactions change during cell differentiation and occur between genomic regions containing epigenetic signatures that are characteristic of enhancers and promoters. Furthermore, it seems that different combinations of these three architectural proteins mediate interactions at different length scales, whereas CTCF in combination with cohesin is enriched in constitutive interactions in mouse ESCs, and these interactions do not change when these cells differentiate into neural progenitor cells. These results are suggestive of a functional specialization of CTCF-mediated contacts as a consequence of interactions between this protein and its various partners. The formation of different complexes with other proteins inside TADs and at TAD borders may underlie the different functions of CTCF in genome organization and provide an explanation for its apparently contradictory properties as both an enhancer facilitator and an enhancer blocker (FIG. 7).

Conclusions and perspectives

The emerging theme from recent studies is that CTCF functions as an architectural protein that contributes to the establishment of genome topology. This is attained at two levels that are likely to be interrelated and that account for most previous observations. At a global level, interactions mediated by CTCF and other architectural proteins result in the formation of TADs. At a more local sub-megabase scale, CTCF may be involved in ‘fine-tuning’ intrachromosomal interactions within TADs to regulate various aspects of gene expression.

Cooperation of CTCF with other protein partners, which are possibly regulated by covalent modifications, may determine its functional specificity. First, association with other proteins — such as TFIIC, cohesins and condensins — at specific genomic locations may result in the formation of TAD borders by precluding interactions across these sites, thus rationalizing the observed enhancer-blocking properties of this protein. Second, association of CTCF with other proteins, such as cohesin and Mediator, may define the range and stability of chromosomal interactions within TADs, which provides an explanation for its other roles in transcription. The ability of CTCF to bind to RNA opens the possibility for ncRNAs in helping to stabilize these contacts and perhaps regulate their function. Finally, covalent modifications of CTCF and its partners are also likely to influence their regulatory potential. The principal outcome of CTCF-mediated contacts is to regulate transcription at various levels, including initiation, promoter selection, promoter-proximal pausing and splicing. The role of CTCF in determining three-dimensional genome organization has so far been considered mostly in the context of its effect on transcription during G1 phase, but both the protein and the architectural properties of the genome it controls are also likely to be important at other stages of the cell cycle, including DNA replication during S phase and chromosome condensation during mitosis¹⁰². In particular, important issues for future studies include how the TAD organization relates to the structure of metaphase chromosomes and how this affects gene expression at the beginning of G1 phase.

1. Van Bortle, K. & Corces, V. G. Nuclear organization and genome function. *Annu. Rev. Cell Dev. Biol.* **28**, 163–187 (2012).
2. de Laat, W. & Dekker, J. 3C-based technologies to study the shape of the genome. *Methods* **58**, 189–191 (2012).
3. Baniahmad, A., Steiner, C., Kohne, A. C. & Renkawitz, R. Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. *Cell* **61**, 505–514 (1990).
4. Lobanenko, V. V. *et al.* A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken *c-myc* gene. *Oncogene* **5**, 1743–1753 (1990).
5. Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E. & Wiehe, T. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc. Natl. Acad. Sci. USA* **109**, 17507–17512 (2012).
6. Ohlsson, R., Renkawitz, R. & Lobanenko, V. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet.* **17**, 520–527 (2001).
7. Chen, H., Tian, Y., Shu, W., Bo, X. & Wang, S. Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PLoS ONE* **7**, e41374 (2012).
8. Cuddapah, S. *et al.* Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* **19**, 24–32 (2009).
9. Schmidt, D. *et al.* Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148**, 335–348 (2012).
10. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
11. Kim, T. H. *et al.* Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**, 1231–1245 (2007).
12. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
13. DeMare, L. E. *et al.* The genomic landscape of cohesin-associated chromatin interactions. *Genome Res.* **23**, 1224–1234 (2013).
14. Song, L. *et al.* Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* **21**, 1757–1767 (2011).
15. Rhee, H. S. & Pugh, B. F. Comprehensive genome-wide protein–DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408–1419 (2011).
16. Nakahashi, H. *et al.* A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep.* **3**, 1678–1689 (2013).
In this study, genome-wide mapping of different CTCF mutants reveals that combinatorial clustering of its 11 zinc-fingers recognizes a wide range of DNA modules.
17. Engel, N., West, A. G., Felsenfeld, G. & Bartolomei, M. S. Antagonism between DNA hypermethylation and enhancer-blocking activity at the H19 DMD is uncovered by CpG mutations. *Nature Genet.* **36**, 883–888 (2004).
18. Rodriguez, C. *et al.* CTCF is a DNA methylation-sensitive positive regulator of the *INK/ARF* locus. *Biochem. Biophys. Res. Commun.* **392**, 129–134 (2010).
19. Lai, A. Y. *et al.* DNA methylation prevents CTCF-mediated silencing of the oncogene *BCL6* in B cell lymphomas. *J. Exp. Med.* **207**, 1939–1950 (2010).

20. Chang, J. *et al.* Nicotinamide adenine dinucleotide (NAD)-regulated DNA methylation alters CTCF-binding factor (CTCF)/cohesin binding and transcription at the *BDNF* locus. *Proc. Natl Acad. Sci. USA* **107**, 21836–21841 (2010).
21. Wang, H. *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* **22**, 1680–1688 (2012).
This paper shows that cell-type-specific CTCF-binding sites across 19 diverse human cell types are linked to the differential DNA methylation patterns within the cells.
22. Zampieri, M. *et al.* ADP-ribose polymers localized on Ctfp-Parp1–Dnmt1 complex prevent methylation of Ctfp target sites. *Biochem. J.* **441**, 645–652 (2012).
23. Guastafierro, T. *et al.* CCCTC-binding factor activates PARP-1 affecting DNA methylation machinery. *J. Biol. Chem.* **283**, 21873–21880 (2008).
24. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
25. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
26. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
27. Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368–1380 (2012).
28. Song, C. X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691 (2013).
29. Spruijt, C. G. *et al.* Dynamic readers for 5-(hydroxy) methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
30. MacPherson, M. J., Beatty, L. G., Zhou, W., Du, M. & Sadowski, P. D. The CTCF insulator protein is posttranslationally modified by SUMO. *Mol. Cell. Biol.* **29**, 714–725 (2009).
31. Yu, W. *et al.* Poly(ADP-ribosylation) regulates CTCF-dependent chromatin insulation. *Nature Genet.* **36**, 1105–1110 (2004).
32. Witcher, M. & Emerson, B. M. Epigenetic silencing of the *p16^{INK4a}* tumor suppressor is associated with loss of CTCF binding and a chromatin boundary. *Mol. Cell* **34**, 271–284 (2009).
33. Ong, C. T., Van Bortle, K., Ramos, E. & Corces, V. G. Poly(ADP-ribosylation) regulates insulator function and intrachromosomal interactions in *Drosophila*. *Cell* **155**, 148–159 (2013).
This work shows that poly(ADP-ribosylation) regulates CTCF-mediated intrachromosomal interactions and its association with other architectural protein in *D. melanogaster*.
34. Zlatanova, J. & Caiafa, P. CTCF and its protein partners: divide and rule? *J. Cell Sci.* **122**, 1275–1284 (2009).
35. Xiao, T., Wallace, J. & Felsenfeld, G. Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol. Cell. Biol.* **31**, 2174–2183 (2011).
36. Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).
37. Parelho, V. *et al.* Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132**, 422–433 (2008).
38. Rubio, E. D. *et al.* CTCF physically links cohesin to chromatin. *Proc. Natl Acad. Sci. USA* **105**, 8309–8314 (2008).
39. Stedman, W. *et al.* Cohesins localize with CTCF at the KSHV latency control region and at cellular *c-myc* and *H19Igf2* insulators. *EMBO J.* **27**, 654–666 (2008).
40. Galli, G. G. *et al.* Genomic and proteomic analyses of Prdm5 reveal interactions with insulator binding proteins in embryonic stem cells. *Mol. Cell. Biol.* **33**, 4504–4516 (2013).
41. Xie, D. *et al.* Dynamic trans-acting factor colocalization in human cells. *Cell* **155**, 713–724 (2013).
42. Hadjur, S. *et al.* Cohesins form chromosomal cis-interactions at the developmentally regulated *IFNG* locus. *Nature* **460**, 410–413 (2009).
43. Hou, C., Dale, R. & Dean, A. Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl Acad. Sci. USA* **107**, 3651–3656 (2010).
44. Nativo, R. *et al.* Cohesin is required for higher-order chromatin conformation at the imprinted *IGF2–H19* locus. *PLoS Genet.* **5**, e1000739 (2009).
45. Kirkland, J. G., Raab, J. R. & Kamakaka, R. T. TFIIIC bound DNA elements in nuclear organization and insulation. *Biochim. Biophys. Acta* **1829**, 418–424 (2013).
46. Noma, K., Cam, H. P., Maraia, R. J. & Grewal, S. I. A role for TFIIIC transcription factor complex in genome organization. *Cell* **125**, 859–872 (2006).
47. Lunyak, V. V. *et al.* Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* **317**, 248–251 (2007).
48. Raab, J. R. *et al.* Human tRNA genes function as chromatin insulators. *EMBO J.* **31**, 330–350 (2012).
49. Carriere, L. *et al.* Genomic binding of Pol III transcription machinery and relationship with TFIIIS transcription factor distribution in mouse embryonic stem cells. *Nucleic Acids Res.* **40**, 270–283 (2012).
50. Oler, A. J. *et al.* Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. *Nature Struct. Mol. Biol.* **17**, 620–628 (2010).
51. Moqtaderi, Z. *et al.* Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. *Nature Struct. Mol. Biol.* **17**, 635–640 (2010).
In this study, genome-wide mapping of distinct components of the Pol III complex in human cells reveals the association of TFIIIC with many ETC loci that are located near CTCF-binding sites.
52. Lei, E. P. & Corces, V. G. RNA interference machinery influences the nuclear organization of a chromatin insulator. *Nature Genet.* **38**, 936–941 (2006).
53. Yao, H. *et al.* Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.* **24**, 2543–2555 (2010).
54. Sun, S. *et al.* *Jpx* RNA activates *Xist* by evicting CTCF. *Cell* **153**, 1537–1551 (2013).
This work shows that the interaction of the *Jpx* ncRNA with CTCF is necessary for the eviction of CTCF from the promoter to facilitate *Xist* transcription.
55. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genet.* **38**, 1341–1347 (2006).
56. Splinter, E. *et al.* CTCF mediates long-range chromatin looping and local histone modification in the β -globin locus. *Genes Dev.* **20**, 2349–2354 (2006).
57. Blanton, J., Gaszner, M. & Schedl, P. Protein:protein interactions and the pairing of boundary elements *in vivo*. *Genes Dev.* **17**, 664–675 (2003).
58. Byrd, K. & Corces, V. G. Visualization of chromatin domains created by the *gypsy* insulator of *Drosophila*. *J. Cell Biol.* **162**, 565–574 (2003).
59. Kellum, R. & Schedl, P. A position-effect assay for boundaries of higher order chromosomal domains. *Cell* **64**, 941–950 (1991).
60. Oki, M. & Kamakaka, R. T. Barrier function at *HMR*. *Mol. Cell* **19**, 707–716 (2005).
61. Huang, S., Li, X., Yusufzai, T. M., Qiu, Y. & Felsenfeld, G. USF1 recruits histone modification complexes and is critical for maintenance of a chromatin barrier. *Mol. Cell. Biol.* **27**, 7991–8002 (2007).
62. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
63. Handoko, L. *et al.* CTCF-mediated functional chromatin interaction in pluripotent cells. *Nature Genet.* **43**, 630–638 (2011).
This paper shows that CTCF-mediated chromatin interactions may separate chromatin domains and facilitate enhancer–promoter contacts.
64. Essafi, A. *et al.* A Wt1-controlled chromatin switching mechanism underpins tissue-specific *Wnt4* activation and repression. *Dev. Cell* **21**, 559–574 (2011).
65. Taslim, C. *et al.* Integrated analysis identifies a class of androgen-responsive genes regulated by short combinatorial long-range mechanism facilitated by CTCF. *Nucleic Acids Res.* **40**, 4754–4764 (2012).
66. Kim, Y. J., Cecchini, K. R. & Kim, T. H. Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus. *Proc. Natl Acad. Sci. USA* **108**, 7391–7396 (2011).
67. Schwartz, Y. B. *et al.* Nature and function of insulator protein binding sites in the *Drosophila* genome. *Genome Res.* **22**, 2188–2198 (2012).
68. Van Bortle, K. *et al.* *Drosophila* CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. *Genome Res.* **22**, 2176–2187 (2012).
69. Li, H. B. *et al.* Insulators, not Polycomb response elements, are required for long-range interactions between Polycomb targets in *Drosophila melanogaster*. *Mol. Cell. Biol.* **31**, 616–625 (2011).
70. Wood, A. M. *et al.* Regulation of chromatin organization and inducible gene expression by a *Drosophila* insulator. *Mol. Cell* **44**, 29–38 (2011).
71. Xie, X. *et al.* Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *Proc. Natl Acad. Sci. USA* **104**, 7145–7150 (2007).
72. Yang, J., Ramos, E. & Corces, V. G. The BEAF-32 insulator coordinates genome organization and function during the evolution of *Drosophila* species. *Genome Res.* **22**, 2199–2207 (2012).
73. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).
By mapping long-range interactions between distal elements and transcription start sites, this study finds that CTCF-binding sites are enriched at many enhancer elements and that enhancer–promoter interactions may not be disrupted by intervening CTCF-binding sites.
74. Shen, Y. *et al.* A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012).
75. Majumder, P. & Boss, J. M. CTCF controls expression and chromatin architecture of the human major histocompatibility complex class II locus. *Mol. Cell. Biol.* **30**, 4211–4223 (2010).
76. Majumder, P., Gomez, J. A., Chadwick, B. P. & Boss, J. M. The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *J. Exp. Med.* **205**, 785–798 (2008).
77. Xu, Z., Wei, G., Chepelev, I., Zhao, K. & Felsenfeld, G. Mapping of *INS* promoter interactions reveals its role in long-range regulation of *SYT8* transcription. *Nature Struct. Mol. Biol.* **18**, 372–378 (2011).
78. Golan-Mashiach, M. *et al.* Identification of CTCF as a master regulator of the clustered protocadherin genes. *Nucleic Acids Res.* **40**, 3378–3391 (2012).
79. Kehayova, P., Monahan, K., Chen, W. & Maniatis, T. Regulatory elements required for the activation and repression of the protocadherin- α gene cluster. *Proc. Natl Acad. Sci. USA* **108**, 17195–17200 (2011).
80. Guo, Y. *et al.* CTCF/cohesin-mediated DNA looping is required for protocadherin α promoter choice. *Proc. Natl Acad. Sci. USA* **109**, 21081–21086 (2012).
81. Monahan, K. *et al.* Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of protocadherin- α gene expression. *Proc. Natl Acad. Sci. USA* **109**, 9125–9130 (2012).
82. Hirayama, T., Tarusawa, E., Yoshimura, Y., Galjart, N. & Yagi, T. CTCF is required for neural development and stochastic expression of clustered *Pcdh* genes in neurons. *Cell Rep.* **2**, 345–357 (2012).
83. Liu, Z., Scannell, D. R., Eisen, M. B. & Tjian, R. Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. *Cell* **146**, 720–731 (2011).
This study shows that CTCF directly recruits TAF3 to promoter distal sites that act as enhancers for endoderm lineage differentiation in ESCs.
84. Bossen, C., Mansson, R. & Murre, C. Chromatin topology and the regulation of antigen receptor assembly. *Annu. Rev. Immunol.* **30**, 337–356 (2012).
85. Guo, C. *et al.* Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. *Cell* **147**, 352–343 (2011).
86. Degner, S. C. *et al.* 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. USA* **108**, 9566–9571 (2011).
87. Shih, H. Y. *et al.* *Tcra* gene recombination is supported by a *Tcra* enhancer- and CTCF-dependent chromatin hub. *Proc. Natl Acad. Sci. USA* **109**, E3493–E3502 (2012).

88. Guo, C. *et al.* CTCF-binding elements mediate control of V(D)J recombination. *Nature* **477**, 424–430 (2011).
This study shows that binding of CTCF to a key V(D)J recombination regulatory region mediates chromatin looping to regulate ordered and lineage-specific rearrangement of the *Igh* locus.
89. Ribeiro de Almeida, C. *et al.* The DNA-binding protein CTCF limits proximal V κ recombination and restricts κ enhancer interactions to the immunoglobulin κ light chain locus. *Immunity* **35**, 501–513 (2011).
90. Stadhouders, R. *et al.* Dynamic long-range chromatin interactions control *Myb* proto-oncogene transcription during erythroid development. *EMBO J.* **31**, 986–999 (2012).
91. Paredes, S. H., Melgar, M. F. & Sethupathy, P. Promoter-proximal CCCTC-factor binding is associated with an increase in the transcriptional pausing index. *Bioinformatics* **29**, 1485–1487 (2013).
92. Shukla, S. *et al.* CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74–79 (2011).
This paper shows that DNA methylation and binding of CTCF at exons regulate alternative splicing by controlling the elongation rate of Pol II.
93. Hou, C., Li, L., Qin, Z. S. & Corces, V. G. Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol. Cell* **48**, 471–484 (2012).
94. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
95. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
This study shows that the mammalian genome is organized into TADs that are separated by boundaries enriched for CTCF and other architectural proteins.
96. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
97. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. USA* **111**, 996–1001 (2013).
98. Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* **32**, 3119–3129 (2013).
99. Seitan, V. C. *et al.* Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Res.* **23**, 2066–2077 (2013).
References 96–99 show that CTCF and cohesin are required to maintain the organization of TADs.
100. Lin, Y. C. *et al.* Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nature Immunol.* **13**, 1196–1204 (2012).
101. Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
102. Naumova, N. *et al.* Organization of the mitotic chromosome. *Science* **342**, 948–953 (2013).

Acknowledgements

Work in the authors' laboratory is supported by US Public Health Service Award R01 GM035463 from the US National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the US National Institutes of Health.

Competing interests statement

The authors declare no competing interests.