The Company of **Biologists**

REVIEW

Developing in 3D: the role of CTCF in cell differentiation

Rodrigo G. Arzate-Mejía[1], Félix Recillas-Targa[1] and Victor G. Corces[2],*

ABSTRACT

CTCF is a highly conserved zinc-finger DNA-binding protein that mediates interactions between distant sequences in the genome. As a consequence, CTCF regulates enhancer-promoter interactions and contributes to the three-dimensional organization of the genome. Recent studies indicate that CTCF is developmentally regulated, suggesting that it plays a role in cell type-specific genome organization. Here, we review these studies and discuss how CTCF functions during the development of various cell and tissue types, ranging from embryonic stem cells and gametes, to neural, muscle and cardiac cells. We propose that the lineage-specific control of CTCF levels, and its partnership with lineage-specific transcription factors, allows for the control of cell type-specific gene expression via chromatin looping.

KEY WORDS: Transcription, Epigenetics, Development, Genome organization, Cell differentiation

Introduction

CTCF is a zinc-finger protein that was initially described as a transcriptional repressor of the *Myc* gene (Klenova et al., 1993; Lobanenkov et al., 1990). It is conserved among bilaterians, and phylogenetic analyses suggest an early origin in the evolution of Metazoa (Heger et al., 2012). CTCF is composed of multiple domains (see Box 1) that allow it to bind to different DNA motifs and various regulatory proteins (Fig. 1). CTCF was initially shown to bind to insulator sequences within the α and β-globin loci (Bell et al., 1999; Chung et al., 1997; Furlan-Magaril et al., 2011; Valadez-Graham et al., 2004), and the imprinted *Igf2/H19* locus (Bell and Felsenfeld, 2000; Hark et al., 2000; Kanduri et al., 2000); studies using reporter constructs in different cell types have suggested that CTCF functions as an insulator protein that can block the ability of enhancers to activate promoters when placed between them in reporter assays (Recillas-Targa et al., 2002). Subsequent work revealed a role for CTCF in the mediation of enhancer-promoter interactions (Guo et al., 2015), alternative splicing (Marina et al., 2016; Shukla et al., 2011), recombination (Hu et al., 2015) and DNA repair (Han et al., 2017). These different functions of CTCF are presumably a reflection of its role, together with that of cohesin, in regulating the formation of chromatin loops and, hence, in controlling three-dimensional (3D) chromatin organization (see Box 2 and Fig. 2). (Fudenberg et al., 2016; Haarhuis et al., 2017; Nora et al., 2017; Sanborn et al., 2015).

Although this general role for CTCF in 3D chromatin organization has been studied in great detail (reviewed by Merkenschlager and Nora, 2016; Ong and Corces, 2014), the

[1]Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, 04510 Ciudad de México, México. [2]Department of Biology, Emory University, Atlanta, GA 30322, USA.

*Author for correspondence (vgcorces@gmail.com)

V.G.C., 0000-0001-5140-4337

precise role played by CTCF during organismal development has remained poorly explored. Here, we review evidence linking CTCF to the control of developmental processes (summarized in Table 1). We first provide an overview of how CTCF functions to regulate genome 3D organization and how this role affects gene expression. We then discuss the roles of CTCF in the development and differentiation of various cell and tissue types, ranging from embryonic stem cells (ESCs) to neural, cardiac and muscle cells. We conclude with an integrative view of CTCF as an important determinant of cell lineage specification during vertebrate development.

Mechanisms of CTCF function

In mammals, most transcriptional activity occurs inside chromatin loops that are bound at their base by CTCF ('CTCF loop anchors', see Fig. 2). Sequences adjacent to such loop anchors are enriched in active histone modifications, RNA polymerase II (RNAPII), housekeeping genes and transcription start sites (Tang et al., 2015), suggesting that CTCF chromatin loops could represent topological structures within which transcription can take place. In support of this, functional experiments that remove or invert CTCF-binding sites result in changes in gene expression that can be direct or indirect, i.e. the disruption of a chromatin loop can cause changes in the transcriptional regulation of the genes contained inside the CTCF loop or affect the transcription of nearby genes originally located outside the loop (Dowen et al., 2014; Guo et al., 2015; Hanssen et al., 2017; Narendra et al., 2015). Alternatively, the disruption of a specific CTCF site may untether regulatory sequences from their target promoter, resulting in transcriptional downregulation. Therefore, the role of CTCF in transcription, either locally or genome wide, is complex, and it is difficult to predict *a priori* the effect of a particular CTCF genomic site on transcription or the subset of genes that can change their transcriptional status due to loss of the site (de Wit et al., 2015; Kubo et al., 2017 preprint; Nora et al., 2017).

The full spectrum of genes that can be transcriptionally regulated by CTCF also seems to depend on the cell type being analyzed. For example, inactivation of the *Ctcf* gene in primary mouse embryonic fibroblasts results in the mis-regulation of 698 genes (Busslinger et al., 2017), whereas similar experiments in post-mitotic embryonic and postnatal neurons results in changes in the expression of about 400 and 800 genes, respectively (Hirayama et al., 2012; Sams et al., 2016). Likewise, removal of CTCF protein in mouse ESCs results in loss of chromatin loops and changes in the expression of hundreds of genes (Nora et al., 2017), whereas experiments using a different mouse ESC cell line and a similar CTCF depletion approach (Kubo et al., 2017 preprint) found a smaller effect of CTCF depletion on transcription. It is surprising that, despite the presence of CTCF at thousands of genomic sites, complete depletion of this protein does not result in more dramatic changes in gene expression. This could be explained, in part, by the activity of additional architectural proteins such as YY1 in establishing long-range interactions between enhancers and

1

CTCF is composed of an N-terminal domain, a central zinc-finger domain with 11 C2H2 zinc fingers (ZF) and a C-terminal domain. The zinc-finger domain is responsible for binding to a 15 bp core motif of DNA, employing ZFs 3-7, while the remaining ZFs can modulate CTCF-binding stability by interacting with adjacent DNA modules (Hashimoto et al., 2017; Nakahashi et al., 2013; Rhee and Pugh, 2011; Schmidt et al., 2012). All three domains of CTCF may also interact with other proteins (see Fig. 1) (Chernukhin et al., 2007; Delgado-Olguín et al., 2012; Ishihara et al., 2006; Lee et al., 2017; Uusküla-Reimand et al., 2016; Xiao et al., 2011, 2015) or RNA (Kung et al., 2015; Saldaña-Meyer et al., 2014; Sun et al., 2013), and are susceptible to post-translational modifications that could affect interactions with DNA or other proteins (Klenova et al., 2001; MacPherson et al., 2009; Yu et al., 2004).

CTCF binds to 40,000-80,000 sites in the mammalian genome, which are predominantly located in intergenic regions and introns, overlapping with regulatory sequences such as enhancers and promoters (Chen et al., 2012). CTCF occupancy across cell types is variable (Beagan et al., 2017; Chen et al., 2012; Martin et al., 2011; Maurano et al., 2015; Prickett et al., 2013; Wang et al., 2012). Cells originating from the same precursors tend to have a similar CTCF-binding landscape whereas cells from different lineages can have marked differences in CTCF occupancy (Prickett et al., 2013; Wang et al., 2012). DNA methylation can affect CTCF binding (Ayala-Ortega et al., 2016; Bell and Felsenfeld, 2000; Hark et al., 2000), possibly by regulating the affinity of CTCF for DNA (Hashimoto et al., 2017). However, the true extent to which DNA methylation directly affects CTCF binding is still controversial (Maurano et al., 2015).

promoters (Weintraub et al., 2017), such that, even in the absence of CTCF, many enhancer-promoter interactions remain and thus the expression of many genes is not affected. It would be of interest to deplete both YY1 and CTCF using the newly employed degron systems and to analyze the effect on genome organization and gene expression in the absence of both proteins. A second possibility is that CTCF and cohesin establish a basal topology of interactions between regulatory elements that is required for timely and dynamic control of gene transcription, e.g. during signal transduction (Oti et al., 2016; Wang et al., 2014). In this scenario, not all CTCF target genes may display a change in gene expression following CTCF depletion, and would only do so under specific cellular conditions. If this is the case, it would be interesting to investigate whether the disruption of CTCF chromatin loops impairs the transcriptional response of cells to specific stimuli. The study of CTCF during development thus offers a good starting point to understand how 3D genome organization integrates a plethora of stimuli that translate into the establishment of cell-specific transcription programs.

CTCF is required for early vertebrate development
In mice, depletion of CTCF from oocytes results in embryo lethality by the morula stage, whereas homozygous null mutant embryos fail to implant and die by the pre-implantation stage (E3.5) (Heath et al., 2008; Moore et al., 2012; Wan et al., 2008). In zebrafish, CTCF knockdown in the one-cell stage embryo results in lethality by 24 h post fertilization (Delgado-Olguín et al., 2011). In both organisms, lethality is accompanied by widespread apoptosis mediated by downregulation of *p53* and upregulation of *Puma* (*Bbc3*) both of which are direct targets of CTCF (Gomes and Espinosa, 2010; Moore et al., 2012; Saldaña-Meyer et al., 2014). These observations provide strong evidence of the importance of CTCF for very early development.

CTCF function is also crucial during postnatal and adult development (Gregor et al., 2013; Hori et al., 2017; Kemp et al., 2014; Sams et al., 2016). In particular, studies in mice have demonstrated that postnatal development of an organism is sensitive to CTCF dose (Gregor et al., 2013; Kemp et al., 2014; Marshall et al., 2017). For example, while CTCF heterozygous knockout (KO) mice are viable, a halved dose of CTCF pre-disposes the mice to develop spontaneous tumors in tissues with high rates of cell proliferation. The molecular mechanisms behind this susceptibility are unclear, although it is known that heterozygous CTCF KO mice show aberrant DNA hypermethylation at specific loci (Kemp et al., 2014), which could, over the course of a lifetime, predispose certain tissues to uncontrolled proliferation by epigenetic silencing of tumor-suppressor genes. CTCF has also been shown to be important for adult cognition (Gregor et al., 2013; Hori et al., 2017). It would thus be of interest to analyze heterozygous CTCF KO mice for cognitive defects, as humans with heterozygous mutations in *CTCF* are known to display intellectual disability (Gregor et al., 2013; Hori et al., 2017; Sams et al., 2016).

CTCF sets a ground state of genome organization in embryonic stem cells that is crucial for development
Embryonic stem cells (ESCs) have the ability to self-renew and differentiate, and this capacity to either maintain the pluripotent state or start a differentiation program has been associated with a unique chromatin landscape that is greatly influenced by CTCF. The functional relevance of CTCF in ESC biology relies on its ability to sustain cell viability, cell proliferation and the pluripotent state by regulating the expression of multiple genes (Balakrishnan et al., 2012; Dowen et al., 2014; Handoko et al., 2011; Ji et al., 2016; Nora et al., 2017; Phillips-Cremins et al., 2013; Tee and Reinberg, 2014). In line with this, it has been shown that CTCF binds 40,000-80,000 sites genome wide in mouse and human ESCs, with around 10% of them localizing to promoter regions and at least 7000 directly involved in chromatin looping interactions (Handoko et al., 2011; Ji et al., 2016; Teif et al., 2014). In mESCs and hESCs, pluripotency genes such as *Oct4* and *Nanog*, as well as lineage-specifying genes, are located inside chromatin loops termed insulated neighborhoods that are anchored by CTCF and cohesin (Dowen et al., 2014; Ji et al., 2016). This type of chromatin loop physically insulates genes and their regulatory sequences from the rest of the genome. For example, the *miR-290-295* gene cluster encodes a miRNA population important for ESC maintenance and survival (Kaspi et al., 2013; Lichner et al., 2011; Zheng et al., 2011). This cluster of miRNA genes and a super-enhancer are located inside a chromatin loop anchored by CTCF at both sides (Fig. 3A). The removal of a CTCF-binding site at one of the loop anchors by CRISPR-Cas9 results in a decrease in pri-miR transcripts, downregulation of endodermal markers and upregulation of *Pax6*, a target of mirR-290-295 involved in ectodermal differentiation (Dowen et al., 2014; Kaspi et al., 2013). In addition, the promoter region of *Nlpr2*, a gene originally positioned outside the chromatin loop, engages in long-range interactions with the super-enhancer. These interactions take place between a CTCF site located upstream of *Nlpr2* and a CTCF site next to the super-enhancer, resulting in transcriptional activation of the gene. In this scenario, the *miR-290-295* locus would be still located inside a larger loop, but the *Nlrp2* promoter will compete for the super-enhancer, thereby reducing *miR-290-295* transcription. This example highlights two important effects of perturbing CTCF binding. First, the removal of a single CTCF site at a chromatin loop anchor can perturb gene expression, in this case leading to premature expression of differentiation genes.
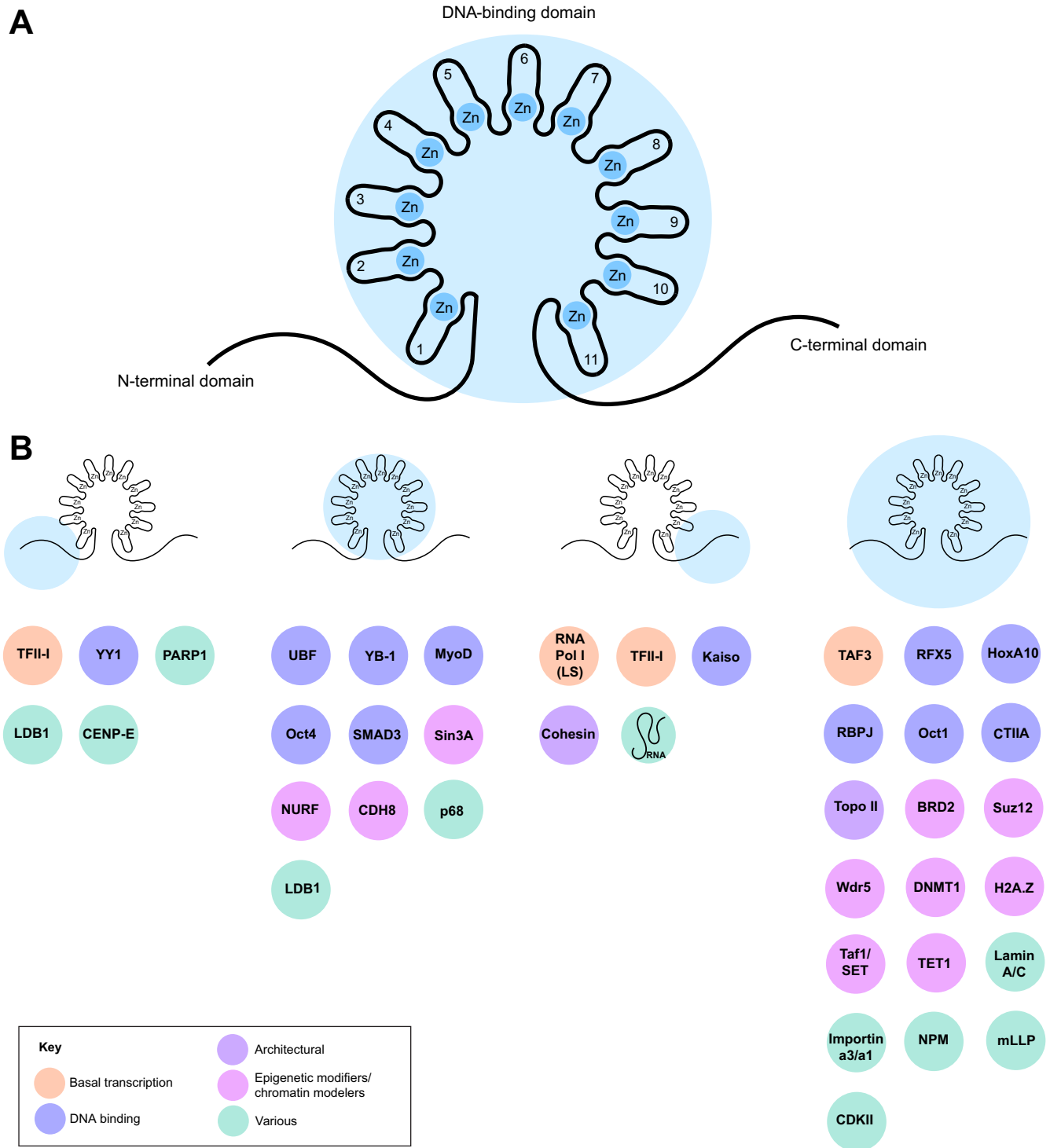
Fig. 1. CTCF interacts with a variety of proteins. (A) Domain structure of CTCF, highlighting the three major domains: the N-terminal domain, the central zinc-finger domain (containing Zn-fingers 1-11) and the C-terminal domain. (B) A variety of CTCF-interacting proteins are known to bind to specific domains of CTCF. Multiple proteins interact with the zinc-finger domain, whereas only RNAPII, cohesin, RNA, Kaiso and TFII-I interact with the C-terminal domain. Likewise, only a handful of proteins interact with the N-terminal domain. A number of additional proteins have been shown to interact with CTCF, although their binding has not been mapped to specific domains.

Second, the nature of the dysregulated gene can amplify the transcriptional defects associated with the removal of a CTCF binding site, as in this case where CTCF controls the expression of miRNAs that, in turn, regulate the levels of a cell type-specific transcription factor (Ayala-Ortega et al., 2016; Dowen et al., 2014).

Insulated neighborhoods can also promote the maintenance of transcriptional repression (Fig. 3B). For example, the removal of one CTCF anchor from a loop containing Polycomb-repressed genes is sufficient to induce aberrant gene expression (Dowen et al., 2014).

Genome-wide depletion of CTCF from mESCs results in loss of chromatin loops and changes in gene expression of 400 and 5000 genes after 24 and 96 h of depletion, respectively (Nora et al., 2017). Early downregulated genes display a preferential enrichment for CTCF binding 60 bp upstream of their transcription start site (TSS) and, intriguingly, CTCF motif orientation is concordant with the direction of transcription, a trend that is also observed in human cell lines (Tang et al., 2015). In contrast, 80% of upregulated genes in CTCF depleted mESCs do not have CTCF at their promoter region and may be stimulated by enhancers located in adjacent loop domains normally separated by CTCF (Nora et al., 2017). Restoration of CTCF levels after acute removal rescues transcription defects in downregulated genes but not in upregulated genes, with some of them showing a constant increase in transcription (Nora et al., 2017). This is reminiscent of a positive-feedback mechanism that could have crucial consequences in early embryonic development.

CTCF can also influence gene expression in ESCs by regulating the transcription of epigenetic factors such as WDR5, which is a core member of the MLL complex. In ESCs, CTCF occupies the *Wdr5* promoter and is necessary for its transcription. Overexpression of WDR5 is able to rescue the proliferative defects seen in CTCF-deficient ESCs, suggesting an important role for CTCF in ESC maintenance via direct regulation of TrxG components (Wang et al., 2017). CTCF also interacts with chromatin remodeling complexes such as NURF, with transcription factors like Oct4, and with the transcription initiator factor TFIID subunit 3 (TAF3) (Donohoe et al., 2009; Liu et al., 2011; Qiu et al., 2015) in ESCs, and perturbing these associations can have important functional consequences (Donohoe et al., 2009; Liu et al., 2011; Qiu et al., 2015). The interaction of CTCF with

TAF3 is interesting because, despite being a general transcription factor, TAF3 regulates lineage commitment in ESCs and co-occupies distal promoter regions bound by CTCF, with CTCF being necessary for TAF3 recruitment.

CTCF plays a central role in brain development and neural function
CTCF levels vary during brain development, with the highest level of expression occurring in the embryonic brain and decreasing from birth to adulthood (Beagan et al., 2017; Sams et al., 2016). CTCF protein levels in adults are higher in the amygdala, hippocampus, cerebellum and cortex than in other organs such as the kidney and heart (Sams et al., 2016). In addition, CTCF is highly enriched in neurons but not in astrocytes and oligodendrocytes, suggesting a specialized role for CTCF in neuronal physiology in adult organisms. Furthermore, analyses of the genome-wide occupancy of CTCF in the mouse brain reveal a high number of brain-specific, CTCF-bound sites that could be involved in regulating the expression of neural genes (Prickett et al., 2013). These data suggest a potential function for CTCF in brain development but, importantly, also hint at the existence of a regulatory program to control the expression of CTCF during development. Removal of the *Ctcf* gene from neural cells during either embryonic or postnatal development results in lethality (Hirayama et al., 2012; Sams et al., 2016; Watson et al., 2014), while depletion of CTCF from telencephalic tissue at E8.5 results in microcephaly, apoptosis by upregulation of *Puma*, cell loss, and embryonic lethality. The specific inactivation of CTCF in neural precursor cells (NPCs) at E11 also induces massive apoptosis, hypocellularity and lethality at birth by asphyxiation (Watson et al., 2014). In addition, CTCF regulates the pool of NPCs by restricting premature neurogenesis. These data suggest that CTCF is required for the survival, proliferation and controlled differentiation of NPCs during early embryonic development.
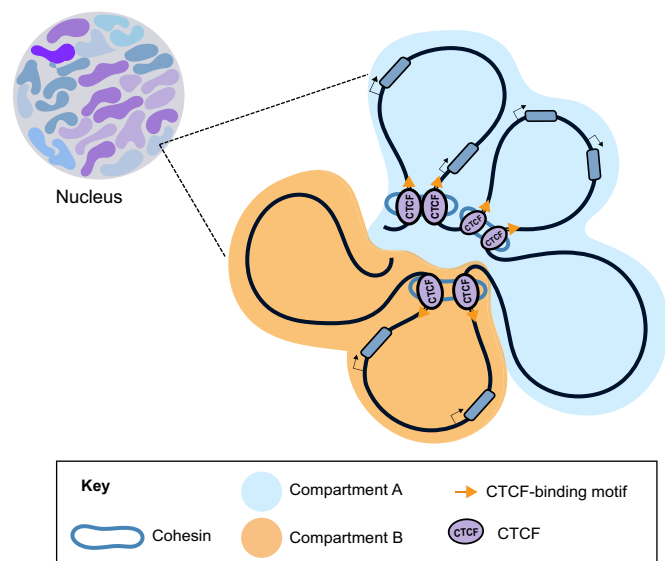
Fig. 2. The organization of the genome inside the nucleus. Eukaryotic chromosomes are organized within the three-dimensional space of the nucleus. They occupy positions in the nucleus termed chromosome territories, each of which can be further organized into two types of domains: compartmental domains, which reflect transcriptional and chromatin states; and loop domains, which depend on the occupancy of architectural proteins such as CTCF and cohesin. See also Box 2.

4

Table 1. Developmental defects associated with CTCF depletion

Cell type, structure or stage of inactivation	Organism	Experimental strategies to remove CTCF	Phenotypic details	References
Embryonic	Mouse	CTCF mutant alleles generated by homologous recombination or the Cre-LoxP system. Homozygous CTCF null mouse ($Ctcf^{-/-}$).	Lethal; CTCF-null embryos die after E3.5. Mutant embryos fail to outgrow; the ICM and trophectoderm do not develop. Increased apoptosis after E3.5 is coincident with loss of maternal CTCF transcript.	Moore et al. (2012) Heath et al. (2008)
Embryonic	Mouse	CTCF mutant alleles generated by homologous recombination and lack all coding exons. Isogenic heterozygous mouse ($Ctcf^{+/-}$).	Heterozygous mice show increased susceptibility to developing spontaneous benign and malignant uterine tumors, histiocytic sarcomas and diploid T cell and T cell infiltrating B cell lymphomas. Increased susceptibility to developing a wide range of tumors after exposure to ionizing radiation and carcinogenic chemicals. Defects in E-cadherin expression. $Ctcf^{+/-}$ mice show DNA methylation instability.	Kemp et al. (2014)
Embryonic	Zebrafish	CTCF knockdown with morpholino oligonucleotides after injection into the one-cell stage embryo.	Somite disorganization. Reduced muscle fibers. Loss of muscle heavy chain (MHC) marker. Reduction of myogenin in specific muscles. Defects in expression of genes involved in hematopoiesis, muscle development and the Wnt pathway.	Delgado-Olguín et al. (2011)
Embryonic	*D. melanogaster*	Amorphic alleles created by mobilization of P elements close to *CTCF* and imprecise excision.	Flies homozygous for amorphic alleles die at late pupal stages. Some flies are able to eclose but they die after 12 to 24 h. Male and females present homeotic transformations of the abdominal segments, while males also show rotated genitalia. Loss of *Abd-B* expression. The lethal phenotype can be rescued by a short pulse of *CTCF* during larvae development but not during the pupal stage.	Gerasimova et al. (2007) Mohan et al. (2007)
T cell	Mouse	Conditional CTCF knockout using Cre-LoxP. *Ctcf* floxed mice were crossed with *Lck-Cre* mice. *Lck* is expressed early during thymocyte development.	Mutant mice show reduced thymic ells and defects in differentiation of $\alpha\beta$ T cells. Blocked cell cycle progression due to increased expression of p21 and p27.	Heath et al. (2008)
Neuron	Mouse	Conditional CTCF knockout using the Cre-LoxP system. *Ctcf* floxed mice were crossed with *Foxg1-Cre* mice. *Foxg1* is expressed in the telencephalic neuroephithelium, the basal ganglia, olfactory bulbs and anterior retina.	Lethal. Pups found dead at birth. CTCF depletion from the forebrain at around E8.5 causes extensive apoptosis, which results in a small telencephalon at E11.5 and complete loss of cortex, hippocampal hem, basal ganglia, lens and anterior retina in mutant embryos at E13.5. Increase in *Puma* expression.	Watson et al. (2014)
		Conditional CTCF knockout using the Cre-LoxP system. *Ctcf* floxed mice were crossed with *Nestin-Cre* mice. *Nestin* is expressed after pre-plate formation.	Neonatal lethality likely due to asphyxiation at birth. Deletion of CTCF in neural progenitor cells at E11 results in an increase in apoptotic cells in the cortex, basal ganglia and hippocampal hem at E14. CTCF-null cells show reduced proliferative capacity. Premature differentiation of mutant apical progenitors. Increase in *p53* and *Puma* expression.	Watson et al. (2014)
		Conditional CTCF knockout using the Cre-LoxP system. *Ctcf* floxed mice were crossed with *Nex-Cre* mice. *Nex* is expressed in postmitotic cortical and hippocampal projection neurons.	Around 30% of mutant mice die within the first day and present with feeding defects. Surviving mice show growth retardation by postnatal day (P) 7 and abnormal behavior, and most of them die by P27. Loss of barrel structure in CTCF-knockout cortex. Strong changes in the expression of Pcdh genes. Dendritic length is reduced as well as dendritic intersections and spines by P14. CTCF-mutant neurons fail to form the appropriate number of synapses.	Hirayama et al. (2012)

DEVELOPMENT

Table 1. Continued

Cell type, structure or stage of inactivation	Organism	Experimental strategies to remove CTCF	Phenotypic details	References
Neuron (continued)	Mouse (continued)	Conditional CTCF knockout using the Cre-LoxP system. *Ctcf* floxed mice were crossed with *Camk2a-Cre* mice. *Camk2a* is highly expressed by P5, specifically in forebrain excitatory postmitotic neurons.	CTCF knockout in hippocampal neurons. Mutant mice display a shortened lifespan. Massive apoptosis detected specifically in the hippocampus by 14 weeks of age. Decrease of spine density in pyramidal cells of the CA1 and defects in long-term potentiation. Mutant mice at 8-10 weeks of age show deficits in cued learning, impaired spatial memory and defects in social recognition. Defects in Pcdh gene expression and learning-associated genes *Arc*, *Bdnf* and *Reln*.	Sams et al. (2016)
Heart	Mouse	Conditional CTCF knockout using the Cre-LoxP system. *Ctcf* floxed mice were crossed with *Nkx2.5-Cre* mice. *Nkx2.5* is expressed in cardiomyocyte precursors in the cardiac crescent by E7.5-E8.	Lethal by stage E12.5. Between E10.5 and E11.5, mutant embryos present with disorganization of the interventricular septum and thinning of the myocardial wall. At E12.5, mutant embryos presented with pericardial edema and defects in cardiac chambers. No defects in apoptosis or cell proliferation. Disorganized mitochondria in CTCF mutant cardiomyocytes at E10.5	Gomez-Velazquez et al. (2017)
Limb	Mouse	Conditional CTCF knockout using the Cre-LoxP system. Mice carrying CTCF-floxed alleles were crossed with *Prx1-Cre* mice. *Prx1* is expressed in early limb bud mesenchyme.	Depletion of CTCF in limbs by E10.75 results in truncation of forelimbs and hindlimbs detectable by E11.5. At E14.0, mutant forelimbs lacked all skeletal elements, including scapula, humerus, radius, ulna and digits. Increase in apoptosis in mutant mesenchymal cells by E11.5, particularly in areas of active proliferation.	Soshnikova et al. (2010)
Oocytes	Mouse	CTCF knockdown; transgenic mice with a transgene that express a RNA hairpin against CTCF mRNA in growing oocytes.	Loss of CTCF in growing oocytes results in hypermethylation of the H19 differentially methylated domain, defects in gene expression and meiosis. Transgenic females have fewer offspring. Zygotes derived from mutant oocytes show increased zygote lethality. Zygotes from transgenic female mice fail to develop to blastocysts. CTCF-depleted embryos arrested at the morula stage show increased apoptosis.	Fedoriw et al. (2004) Wan et al. (2008)
Spermatocytes	Mouse	Conditional CTCF knockout using the Cre-LoxP system. Mice carrying CTCF-floxed allele were crossed with *Stra8*-iCre transgenic mice. *Stra8* drives expression of the Cre recombinase in spermatogonia and in pre-leptotene spermatocytes.	CTCF knockout mice have smaller testes and are infertile (sperm count reduced by 90%). Formation of elongated spermatids is defective because of elimination by apoptosis. Elongated spermatid heads show defects in nuclear morphology and chromatin compaction. Chromatin structures protrude out of elongated spermatid heads. Deficient incorporation of protamine 1 (PRMD1) into chromatin. Defects in histone retention. Downregulation of spermiogenesis-related genes.	Hernández-Hernández et al. (2016)

CTCF is also important for the survival of post-mitotic neurons (Hirayama et al., 2012; Sams et al., 2016). The conditional removal of CTCF from post-mitotic cortical and hippocampal neurons in mice at E11.5 results in growth retardation, abnormal behavior and reduced lifespan, with mice dying by 5 weeks after birth (Hirayama et al., 2012). In this context, CTCF was found to be important for synapse formation and dendrite development during postnatal development, as mutant neurons in the cortex and hippocampus show a dramatic decrease in dendritic length and arborization, as well as dendritic spine formation. This results in decreased ability to form an appropriate number of synapses, which could affect learning and memory. These defects are evident at postnatal day 14 (P14) but not at P7, suggesting that the effect of CTCF on neuron morphology may be restricted to a particular window of adult development. CTCF also contributes to cognitive functions in adult mice. Specific removal of CTCF from postnatal excitatory neurons results in profound defects in memory and learning, as well as reduced life span, with mutant embryos dying by 14 weeks of age (Sams et al., 2016). In this case, only subtle defects in spine density are observed by 10 weeks of age, in contrast to the strong defects
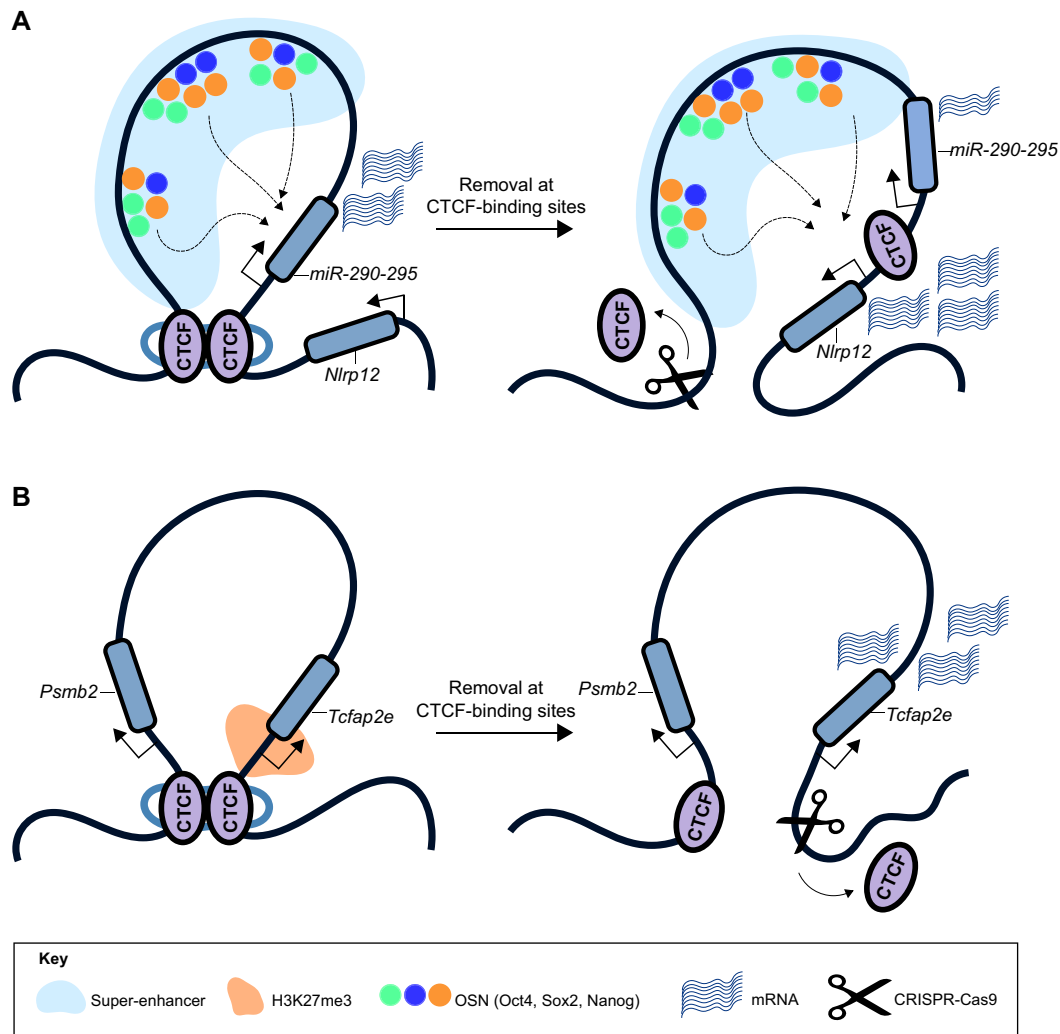
Fig. 3. CTCF promotes the establishment of insulated neighborhoods to maintain proper gene expression in ESCs. (A) CTCF forms loop domains in ESCs to insulate pluripotency genes (such as those encoding Oct4, Sox2 and Nanog; OSN) from sequences located outside of the loops. An ESC super-enhancer and the *miR-290-295* locus are located inside a chromatin loop anchored by CTCF. This topological structure restricts the stimulatory potential of the super-enhancer to the genes contained inside the loop. Removal (e.g. via CRISPR-Cas9-mediated approaches) of the CTCF-binding site at one of the loop anchors results in long-range interactions with a gene originally located outside the loop (*Nlrp12*) and a decrease in the transcription of the original target genes. (B) CTCF can also maintain the epigenetic silencing of Polycomb-repressed lineage-specifying genes through chromatin looping. For example, the disruption of looping (via the deletion of one of the CTCF sites) leads to ectopic expression of the *Tcfap2e* gene.

described when removing CTCF from postmitotic neurons during embryonic development (Hirayama et al., 2012). These phenotypic differences could be the result of a specific regulatory program directed by CTCF at different developmental time points during neuronal maturation, suggesting that the effect of CTCF on cell physiology is highly dependent on the developmental stage and, possibly, on the chromatin organization of its target genes.

In line with these phenotypes, it has been shown that CTCF controls the expression of several genes that are crucial for correct neural development and function (Hirayama et al., 2012; Sams et al., 2016). For example, CTCF depletion from post-mitotic cells, either during embryonic development or in postnatal neurons, results in profound downregulation of the stochastically expressed isoforms of the protocadherin genes (Pcdh), which are essential for the generation of neural diversity (Hirayama and Yagi, 2013). Remarkably, the dendritic arborization defects observed in CTCF mutant neurons are similar to those observed following the knockout of Pcdh genes (Garrett et al., 2012), further suggesting a

functional link between CTCF and Pcdh gene expression. Studies in neural cell lines and mouse brain tissue suggest that the molecular mechanisms underlying this link reside mainly in the ability of CTCF to mediate looping interactions along the Pcdh locus (Guo et al., 2012, 2015; Monahan et al., 2012). The Pcdh locus contains three clusters of genes termed *Pcdha*, *Pcdhb* and *Pcdhg*, each coding for variable exons. Each of the variable exons has an upstream promoter sequence that binds CTCF, while enhancer elements located along the locus direct the expression of Pcdh genes specifically in neural cells. The binding of CTCF to the enhancer elements in neural cells promotes long-range chromatin looping interactions with the promoter sequences of variable exons in an orientation-dependent manner, resulting in stochastic expression of Pcdh genes (Guo et al., 2015). Accordingly, inverting the orientation of one CTCF site located at an enhancer element disrupts long-range interactions with target promoters and results in loss of expression of stochastically transcribed isoforms (Guo et al., 2015). However, it is possible that CTCF affects the transcription of

DEVELOPMENT

Pcdh genes indirectly. Indeed, it is known that CTCF physically interacts with mLLP, which is a permeable nuclear protein that regulates dendritic and spine growth, to promote the expression of Pcdh genes and the amyloid β precursor (*App*) gene (Yu et al., 2016).

CTCF is also essential for the transcription of learning-inducible genes such as *Arc* and *Bdnf*, as well as for the repression of *Hdac3* and *Ppc1*, which are known suppressors of memory formation (Sams et al., 2016). In the case of the *Arc* and *Bdnf* genes, CTCF may promote long-range interactions between their promoters and potential enhancer sequences, as absence of this protein results in loss of chromatin contacts of the *Arc* and *Bdnf* promoters with regions enriched for RNAPII in hippocampal neurons. CTCF is also important for brain development in humans (Gregor et al., 2013). For example, individuals with monogenic mutations in *CTCF* that reduce transcript levels show profound intellectual disability, microcephaly, and feeding and heart defects, which resemble the phenotypes observed in brain-specific *Ctcf* mutant mice (Gregor et al., 2013; Hirayama et al., 2012). CTCF has also been shown to regulate the expression of genes implicated in neurodegenerative diseases, such as Huntington's disease (De Souza et al., 2016) and Spinocerebellar ataxia type 7 (Sopher et al., 2011), psychiatric diseases such as schizophrenia (Juraeva et al., 2014), and neurodevelopmental disorders such as autism (Meguro-Horike et al., 2011). CTCF could also be indirectly involved in mediating the defects in genome organization and gene expression observed in α-thalassemia X-linked intellectual disability (ATRX) and Rett syndromes (Kernohan et al., 2014).

CTCF is involved in cardiovascular development
The mammalian heart is the first organ to function in the embryo and its development involves complex regulatory processes that instruct the differentiation and assembly of cells (Bruneau, 2013; Delgado-Olguín et al., 2012). A recent study has revealed that CTCF plays key roles during this process. The conditional removal of CTCF from cardiac progenitors in mouse embryos results in embryonic lethality by E12.5 (Gomez-Velazquez et al., 2017). Hearts from mutant animals display abnormalities that are not due to defects in cell proliferation or survival but instead arise due to defective differentiation and maturation, probably because of the downregulation of multiple cardiac transcription factors as well as changes in chromatin long-range interactions (Gomez-Velazquez et al., 2017). For example, loss of expression of *Irx4* is accompanied by changes in the chromatin interaction landscape of its promoter region. In addition, key genes whose products participate in important signaling pathways such as BMP, Notch, EGF and TGFβ, also show changes in expression.

CTCF is also important for the development of the vasculature. Genetic inactivation of CTCF in endothelial progenitor cells results in embryonic lethality by E10.5, whereas heterozygous mice display no obvious defects (Roy, 2016). Mutant mice have narrow cerebral vessels and defects in the yolk sac and placenta vasculature. Importantly, vascular defects are not due to problems with cell proliferation or survival but may be attributable to the upregulation of vascular regulatory genes such as *Kdr*, *Vegfa* and *Erg* (Roy, 2016). In line with these findings, it has been demonstrated that CTCF binds to a chromatin insulator sequence located in proximity to the *Vegf* promoter region and insulates *Vegf* from the stimulatory effect of nearby enhancers, restraining its angiogenic potential (Tang et al., 2011). In fact, CTCF deficiency leads to increased and uncontrolled angiogenesis *in vitro* and *in vivo* due to increased expression of VEGF (Tang et al., 2011). This observation has

significant implications for understanding cancer because solid tumors are typically highly vascularized. Furthermore, an important observation stemming from both the heart and vascular phenotypes is that defects are not due to loss of proliferation or survival, which are features of other CTCF knockouts in structures such as the brain and limbs. Thus, it may be possible that CTCF in cardiac and endothelial progenitor cells is important for the control of cell type-specific transcriptional programs, acting either via the regulation of master transcription factors or by modulating signaling pathways.

CTCF promotes limb development
Limb development is accompanied by changes in gene expression that are tissue and time specific (Andrey et al., 2017), and defects in this spatio-temporal regulation of gene expression can cause limb malformations (Franke et al., 2016; Lupiáñez et al., 2015; Symmons et al., 2016; Will et al., 2017). During mouse limb development, promoters of genes involved in limb development engage in static and stage-specific contacts with enhancer sequences that are enriched for the binding of CTCF and cohesin, suggesting that CTCF could have a role in the establishment of a regulatory landscape upon which fine control of gene expression can take place (Andrey et al., 2017). For example, conditional removal of CTCF from developing mouse limbs results in limb truncation with massive apoptosis and changes in the expression of hundreds of genes, including *Shh* (Soshnikova et al., 2010). Some of the changes in gene expression upon CTCF removal could be attributable to defects in CTCF-mediated looping interactions. This has been shown for the *Wnt6/Ihh*, *Epha4* and *Pax3* genes in mouse and human, which are located consecutively in the linear genome but are arranged into different chromatin loop domains (Lupiáñez et al., 2015). *Epha4*, for example, is located inside a CTCF loop, and long-range interactions between enhancer sequences and the *Epha4* promoter are insulated from neighboring domains by CTCF. Structural rearrangements encompassing the CTCF boundaries of the *Epha4* domain result in the gain of ectopic interactions between genes in the adjacent domains and the *Epha4* enhancers, which results in ectopic transcriptional activation of neighboring genes and limb malformations (Lupiáñez et al., 2015). This work has provided compelling evidence for how the integrity of CTCF loop domains provides a framework for the establishment of specific regulatory interactions and how their disruption results in critical consequences for development.

CTCF is important for the regulation of Hox genes
Hox genes encode transcription factors that direct anterior-posterior patterning in all bilaterians as well as the formation of secondary structures such as limbs and external genitalia (Lonfat and Duboule, 2015; Mallo et al., 2010). Hox genes often exist in gene clusters, and their order of expression along the body axis correlates with their position within the cluster, with the 3′ genes expressed more anteriorly and the 5′ genes expressed in the posterior part of the embryo later in development (Mallo et al., 2010). The presence of CTCF at several sites along the Hox loci, both in *D. melanogaster* and mammals (Holohan et al., 2007; Narendra et al., 2015; Soshnikova et al., 2010), as well as its ability to mediate enhancer-promoter interactions, suggests that it could contribute to the temporal control of Hox gene expression by establishing chromatin domains. Indeed, a number of studies have now shown that CTCF regulates Hox gene expression in various contexts.

In *Drosophila*, CTCF depletion early in development results in a homeotic phenotype and decreased expression of the homeotic gene *Abd-B* (Mohan et al., 2007). The conditional removal of CTCF from

developing mouse limbs results in changes in the expression of some posterior HoxD genes but no activation of anterior ones (Soshnikova et al., 2010). In this context, it has recently been shown that the HoxD cluster, which is rich in CTCF binding sites, acts as a dynamic border between two topological domains that contain enhancer sequences controlling the spatiotemporal transcription of the HoxD genes (Rodríguez-Carballo et al., 2017). The precise location of this border seems to be dependent on the cell type or tissue under analysis, and ultimately on the expression of specific HoxD genes. Removal of the whole-gene cluster, which encompasses most of the CTCF-binding sites, is necessary to abrogate the border and to merge the neighboring domains (Rodríguez-Carballo et al., 2017). This means that the CTCF sites located in the HoxD cluster are capable of promoting long-range contacts with convergent CTCF sites at enhancer sequences in neighboring domains. Upon removal of most of the CTCF sites, the two domains outside of the HoxD cluster merge together.

The targeted removal of specific CTCF-binding sites along the HoxA locus has clarified the role of individual binding sites for this protein in the transcriptional regulation of Hox genes (Narendra et al., 2015). ESC-derived motor neurons exposed to retinoic acid (RA) express the rostral part of the HoxA cluster (Hoxa1-Hoxa6) while Hoxa7-Hoxa13 remain repressed. The transition between active and inactive domains is localized between the Hoxa5 to Hoxa7 genes, where two highly conserved and constitutive CTCF-binding sites are located. One of these sites is located at the intergenic region between transcriptionally active Hoxa5 and Hoxa6 (C5|6), and the other is between the active Hoxa6 and the inactive Hoxa7 (C6|7). Importantly, CTCF bound at the C5|6 site establishes at least three strong, long-range interactions with CTCF sites outside the 3′ side of the cluster in a convergent orientation (Narendra et al., 2016). Removal of 9 bp within the core motif of the C5|6 site results in loss of CTCF binding, and strong upregulation of Hoxa7 and, to a lesser extent, of Hoxa9. These transcriptional effects are also accompanied by gain of H3K4me3 at both genes and the establishment of new long-range interactions with the active side of the domain (Hoxa1-Hoxa6) (Narendra et al., 2015). This effect was also recapitulated in the HoxC locus, causing homeotic transformations typical of Hox deregulation in mice (Narendra et al., 2016). The effects observed upon removal of CTCF-binding sites could be attributed to the formation of new long-range interactions between convergent CTCF sites that may bring enhancers located outside the HoxA domain (Langston et al., 1997; Woltering et al., 2014) into close proximity to promoters of the HoxA genes, limiting stimulatory activity to the Hox genes located inside the new CTCF loop. Therefore, in response to developmental signals, a single CTCF-binding site may form long-range interactions that help to establish domains of active transcription.

The association of CTCF with sites within a Hox locus can also be dynamic and modulated by epigenetic mechanisms, such as DNA methylation, or by RA signaling (Ishihara et al., 2016; Min et al., 2016). In mouse embryonic fibroblasts, the expression of both the anterior and posterior HoxC genes correlates with the binding of CTCF between the Hoxc12 and Hoxc11 genes (C12|11). This CTCF-binding site is methylation sensitive and in a hypomethylated state is permissive for CTCF association, resulting in long-range interactions between the posterior and the anterior parts of the locus, and the expression of posterior HoxC genes (Min et al., 2016). In a second example, the differentiation of human NT2/D1 cells to neural cells by treatment with RA was shown to induce the expression of anterior Hoxa1-Hoxa5 genes and the eviction of CTCF from a binding site located between the Hoxa5 and Hoxa4

genes; this CTCF site has a RA response element that may compete with CTCF for binding and may function as an enhancer blocker (Ishihara et al., 2016).

Together, these findings highlight the contribution of CTCF to the proper expression of Hox genes during development by promoting long-range contacts with enhancer sequences that stimulate the transcription of target Hox genes. At the same time, CTCF-mediated loops can insulate enhancer elements from inducing the transcription of Hox genes that have to remain inactive in a particular cell type.

The role of CTCF in regulating left-right asymmetry during organ formation

Pitx2 is a transcription factor that directs the transcriptional programs responsible for the left-right asymmetry of internal organs (Shiratori et al., 2001). In the dorsal mesentery of chicken and mouse embryos, Pitx2 regulates the molecular pathways required for looping and vascularization of the gut (Mahadevan et al., 2014). A recent study has shown that Pitx2 is transcribed in the left dorsal mesentery and its transcriptional repressor, the lncRNA Playrr, is repressed, while both loci are in close spatial proximity (Welsh et al., 2015). In contrast, in the right dorsal mesentery, Pitx2 and Playrr are no longer in spatial proximity and Playrr is expressed while Pitx2 is repressed. Furthermore, Hi-C data from mESCs, which mirror the pattern of gene expression seen in the left dorsal mesentery, suggest that Pitx2 and Playrr are located inside a TAD (see Box 2) but exist within adjacent sub-TADs; CTCF demarcates the borders of both sub-TADs (Welsh et al., 2015), suggesting that these sub-TADs are equivalent to the CTCF loop domains defined previously (Rao et al., 2014). Prominent CTCF-binding sites are located at the Pitx2 locus and upstream of Playrr, and knockdown of CTCF results in loss of spatial proximity between these loci, giving rise to a configuration similar to the one observed in the right dorsal mesentery where Pitx2 is inactive (Welsh et al., 2015). It is not known whether the loss of proximity after CTCF removal results in Playrr transcriptional activation and Pitx2 repression, but two observations are worth pointing out. First, the Pitx2 protein is important for the spatial proximity between Pitx2 and Playrr. Second, Pitx2 protein is present at the anchors of CTCF loop domains for both Pitx2 and Playrr. This is supported by observations of genome-wide association between Pitx2 and CTCF based on the finding of a Pitx2-binding motif at CTCF-binding sites (Chen et al., 2012). Based on these observations, one could speculate that CTCF establishes interactions in the left dorsal mesentery that are further stabilized by Pitx2 by its co-binding to CTCF sites. The loss of either protein thus results in loss of the 3D organization of the Pitx2-Playrr locus. As the function of Pitx2 in the control of organ asymmetry is conserved in non-bilaterian organisms, and given that CTCF is a bilaterian innovation (Heger et al., 2012; Watanabe et al., 2014), it is possible that the CTCF-dependent regulation of Pitx2 may be a bilaterian-specific strategy.

The regulation of myogenesis by CTCF

Myogenesis is regulated by the master transcription factor MyoD and additional muscle-specific transcription factors that promote a muscle-specific gene expression program (Tapscott, 2005; Weintraub et al., 1989). CTCF physically interacts with MyoD, an interaction that increases MyoD affinity for some target promoters by an as yet unknown mechanism (Delgado-Olguín et al., 2011). Interestingly, muscle differentiation is also accompanied by the upregulation of CTCF, although the role of this is still unclear (Delgado-Olguín et al., 2011). The functional relevance of CTCF in

DEVELOPMENT

muscle development is evident in zebrafish, where CTCF is enriched in somites – the tissues in which myogenic precursors are determined. CTCF depletion in developing zebrafish embryos induces somite disorganization, loss of myogenic markers and reduced muscle fibers, as well as changes in the expression of genes related to muscle development and hematopoiesis (Delgado-Olguín et al., 2011). In line with this, CTCF knockdown in the mouse myoblast cell line C2C12 impairs myogenic differentiation and results in transcriptional downregulation of the myogenin gene (Battistelli et al., 2014; Delgado-Olguín et al., 2011), again suggesting a role for CTCF in the transcriptional activation of muscle-specific genes.

CTCF can also induce chromatin looping interactions that insulate genes that are important for the myogenic differentiation program. For example, it has been shown that the promoter of the Cdk inhibitor *p57* (*Cdkn1c*) physically interacts with the imprinting control region KvDMR1, which is located more than 150 kb away. This looping interaction is important for *p57* repression prior to differentiation and depends on the binding of CTCF and Rad21 to both regulatory elements (Battistelli et al., 2014). During skeletal muscle differentiation, *p57* is induced and this correlates with disruption of the chromatin loop, loss of Rad21 and the binding of MyoD to the KvDMR1 element (Battistelli et al., 2014; Busanello et al., 2012). In this context, it has been suggested that MyoD binding causes the displacement of cohesin, possibly by physically interacting with CTCF, which induces destabilization of the chromatin loop and allows the expression of *p57*. Importantly, CTCF binding remains unchanged at interacting regions. Instead, the dynamic association of Rad21 and MyoD determines whether a chromatin loop is formed (Battistelli et al., 2014). Therefore, CTCF shows a complex regulatory behavior during myogenesis by promoting the expression of some muscle-specific genes or by establishing a chromatin loop that constrains the expression of other genes.

CTCF is involved in retinal cell differentiation

Several transcription factors direct the development of the retina. Of these, Pax6 has emerged as a critical regulator that is modulated by CTCF. Pax6 and CTCF display non overlapping patterns of expression during retinal development in chick embryos, suggesting antagonistic roles for these proteins during eye formation (Canto-Soler et al., 2008). Indeed, CTCF overexpression in mouse embryos results in under-developed eyes, small lenses and reduced populations of cells in the retina, lens and cornea; this phenotype is similar to the one observed in *Pax6* mutants (Hill et al., 1992; Li et al., 2004). These CTCF-dependent defects are accompanied by a marked decrease in Pax6 expression, and studies have indeed demonstrated that CTCF binds to a sequence located upstream of the *Pax6* promoter and downstream of an enhancer element that is active in eye-derived cells (Li et al., 2004). Therefore, CTCF could act as an enhancer blocker in the *Pax6* regulatory region. The developmentally restricted expression patterns of both CTCF and Pax6 in the developing eye suggest that additional signals may affect the expression of these genes, although their nature is not known. Activation of the EGF pathway in rabbit corneal epithelial cells induces CTCF upregulation, which results in repression of *Pax6* and its target genes, as well as enhanced cell proliferation (Li and Lu, 2005; Tsui et al., 2016), but whether a similar mechanism exists in other species is unclear.

CTCF contributes to the generation of immune cell diversity

One of the most remarkable properties of the immune system is its ability to differentiate between self and foreign antigens, with T

lymphocytes (T-cells) playing a major role in this event. During normal development, T-cells are exposed to a vast number of tissue-restricted antigens that are expressed, processed and presented on the surface of medullary thymic epithelial cells (mTECs) in the thymus, thereby allowing the negative selection of these cells and preventing organ autoimmunity. The expression of these tissue-restricted antigens is regulated by the autoimmune regulator AIRE, which is a transcription factor largely responsible for inducing the expression of around 4000 genes specifically in mTECs (Sansom et al., 2014). In various mouse cell types, the AIRE locus is flanked on each side by two CTCF sites, suggesting that CTCF binding at this locus is relatively cell type invariant (Herzig et al., 2017). However, in mTECs, CTCF is evicted from the AIRE TSS in parallel with the binding of cell type-specific transcription factors to the promoter region and an upstream enhancer, suggesting that the presence of CTCF correlates with AIRE repression. In line with this, it has been shown that CTCF knockdown experiments result in AIRE transcription (Herzig et al., 2017). Whether AIRE is located inside a chromatin loop formed by the two flanking CTCF-binding sites remains to be determined, but overall these data suggest that CTCF flanking the AIRE locus is important either for transcriptional repression of AIRE or to insulate the AIRE promoter from active nearby enhancers that could activate AIRE transcription in cell types other than mTECs. The role of CTCF in autoimmunity is further supported by its role in regulating the expression of major histocompatibility complex II (MHC-II), both in humans and mice (Majumder and Boss, 2011; Majumder et al., 2008, 2014).

CTCF, via its regulation of immunoglobulins and T-cell receptor genes, is also crucial for the recognition of foreign antigens. A fundamental feature of the adaptive immune response is the specific recognition of a foreign antigen, which is achieved by B and T cells expressing specific receptors. These receptors originate from seven antigen receptor genes, four T-cell receptor (TCR) genes (*Tcrg*, *Tcrd*, *Tcrb* and *Tcra*) and three B-cell immunoglobulin genes (*Igh*, *Igk* and *Igl*). During development, B and T cells make use of these genes to generate an almost infinite variety of specific receptors that are created by the somatic rearrangement of consecutive variable (V), diversity (D) and joining (J) coding gene segments at each locus, in a process known as V(D)J recombination (Proudhon et al., 2015). This process is developmentally regulated, highly ordered and, as indicated by recent studies, involves CTCF. For example, CTCF binds to more than one hundred sites in the V_H region of the mouse *Igh* locus (Choi et al., 2013); the distribution of CTCF at this locus is B-cell lineage-specific and it remains stable during B-cell maturation. Importantly, the depletion of CTCF results in loss of locus compaction, which in turn affects the proximity of the V_H gene segment to the DJ_H gene segment, and biases the V(D)J recombination processes (Gerasimova et al., 2015).

The binding of CTCF to specific regions of the *Igh* locus has also been shown to be important for locus-specific 3D organization. The intergenic control region 1 (IGCR1) located in the V_H-D_H interval of the *Igh* locus contains two binding sites for CTCF. The binding of CTCF to these sites mediates chromatin looping interactions between regions of the V_H domain with IGCR1, as well as interactions between IGCR1 and the 3′ end of the *Igh* locus. Accordingly, the removal of these CTCF-binding sites results in disruption of ordered and lineage-specific V_H-to-DJ_H joining, with each of the two CTCF-binding sites contributing in a differential way to the control of V(D)J recombination (Guo et al., 2011; Lin et al., 2015). CTCF is also important for constraining V(D)J recombination to specific loci. This recombination process is known

DEVELOPMENT

to be dependent on the activity of lymphoid specific recombinases (RAG1/2). Although these RAGs recognize sequence motifs present along the genome, their activity becomes restricted to the immunoglobulin and TCR loci. This coordinated and on-target activity of RAGs is restricted by chromatin loops flanked by convergent CTCF-binding sites. Notably, IGCR1 is a crucial element for folding the locus in at least two chromosomal domains where regulated recombination occurs (Hu et al., 2015).

CTCF also mediates inter-chromosomal interactions between genes that are important for T-cell development (Kim et al., 2014). In naïve T cells, the locus control region (LCR) of the *Th2* cytokine gene on chromosome 11 physically interacts with the promoter of the *Il17a* gene on chromosome 1. This interaction modulates the transcription of *Il17a* and is dependent on the association of CTCF and the transcription factor Oct1 with the regulatory elements of both loci (Kim et al., 2014). Interestingly, CTCF and Oct1 physically interact, and the binding of both proteins to the *Th2* and *Il17a* genes is lost during differentiation into Th2 or Th17 cells, which also correlates with the loss of inter-chromosomal interactions between Th2 and IL17. Despite the presence of binding sites for both CTCF and Oct1 at the Th2 and IL17 loci, knockdown of either one affects the binding of the other at both loci, suggesting that Oct1- and CTCF-dependent inter-chromosomal interactions are contingent on mutual recruitment, which could enhance the physical interaction between co-bound loci located in different chromosomes.

CTCF is involved in the development of gametes
In mouse oocytes, CTCF depletion causes changes in gene expression, mild meiotic defects and embryonic lethality (Wan et al., 2008), probably owing to loss of the CTCF maternal contribution. Likewise, the conditional inactivation of CTCF in mouse spermatocytes results in smaller testes and infertility (Hernández-Hernández et al., 2016). Mutant spermatids, which are eliminated through apoptosis, show defects in nuclear morphology, widespread changes in gene expression, and protrusion of de-compacted chromatin. In line with this, CTCF inactivation results in dramatic defects in histone retention and a profound reduction of the protamine PRM1 (Cho et al., 2001). These data suggest that CTCF is important for chromatin compaction during spermiogenesis, in part by regulating PRM1 protein levels. Remarkably, Hi-C analysis has shown that, despite the high level of chromatin compaction, compartmental and CTCF loop domains are clearly present in mature mouse sperm (Jung et al., 2017). In line with this, genome-wide mapping of CTCF occupancy in mature sperm has shown that CTCF sites present at loop anchors also contain cohesin and are preferentially arranged in a convergent orientation, a finding that in turn confirms that sperm chromatin follows the same rules of 3D organization as diploid somatic cells. Interestingly, chromatin 3D organization is highly similar between sperm and ESCs in terms of compartmental and CTCF loop domains (Jung et al., 2017). This observation, in conjunction with the shared CTCF-binding sites between sperm and ESCs, as well as the identification of enhancers and super-enhancers in sperm that are also present in ESCs, suggest a role for the paternal chromosome in instructing early zygotic transcription (Teperek et al., 2016).

Conclusions and perspectives
Here, we have presented evidence supporting a role for CTCF in cell differentiation and, hence, development, highlighting a crucial function of individual CTCF-binding sites in the developmental control of gene expression (Fig. 4). Together, these studies reveal

that the removal or inversion of specific CTCF sites disrupts normal long-range chromatin interactions, in some cases directly affecting the contacts between enhancers and promoters. In other cases, the phenotypic effect of disruption of CTCF sites is a manifestation of the insulating properties of CTCF. Genes located inside a CTCF loop may not change their expression after deletion of a CTCF anchor, but new looping interactions with neighboring CTCF sites result in stimulation of adjacent genes located outside the original loop. These observations suggest that, despite the recruitment of cell type-specific transcription factors to enhancer sequences, these factors can stimulate the transcription of non-cognate promoters due to their inclusion in the new CTCF loop, suggesting that the specificity of an enhancer element might be dictated by the topological constraints imposed by the chromatin looping mediated by CTCF and cohesin. In these cases, the phenotypic consequences of changes in chromatin topology will depend on the nature of the affected gene. CTCF-binding sites are therefore not all equivalent at the functional level, i.e. the transcriptional consequences of the removal of a CTCF binding site at a particular locus are context dependent.

Developmental signals can also instruct a CTCF site to regulate gene expression by means of establishing functional interactions with either transcription factors or co-factors. The physical interaction of CTCF with cell type-specific transcription factors such as Oct1, Oct4, MyoD and LDB1, or its association with effectors of major signaling factors such as RBPJ or Smad proteins, suggests that a ubiquitous DNA-binding protein such as CTCF can diversify its regulatory potential by cooperating with different regulatory proteins. In this regard, an intriguing observation is the frequent physical association of cell type-specific transcription factors or co-factors with the zinc-finger domain of CTCF, which could potentially modulate CTCF binding but also could promote CTCF recruitment to additional sites or even modify cohesin dynamics at that site. Fine mapping of such physical interactions, coupled with functional assays in which critical residues involved in CTCF interactions with particular proteins are mutated, could reveal the functional significance of these physical interactions.

Based on the findings presented here, we propose that the progressive gain of genome organization during early development is accompanied by the binding of CTCF to the genome, which, together with cohesin and other architectural proteins such as YY1, can result in the establishment of loop domains. During differentiation, the CTCF-binding landscape is dynamic, with some binding sites present in most cell types while others are occupied in several lineages and yet others can be found in only a specific cell type. Importantly, it appears to be not only the presence but also the levels of CTCF at a specific site that determine the frequency of interactions and, therefore, its effect on gene expression. Ubiquitous CTCF-bound sites might represent central locations for genome organization by chromatin looping that affect the expression of housekeeping genes. Some of these stable loops may form a topological regulatory framework upon which rapid control of gene expression can take place, e.g. via the activation of transcriptionally poised or inactive genes. CTCF-binding sites may also mediate the formation of cell type-specific long-range interactions that regulate the transcription of specific genes. In this scenario, CTCF in association with specific transcription factors and in response to post-translational modifications can mediate looping among newly active enhancer sequences. Thus, by mediating long-range interactions, CTCF can establish loops that can insulate a regulatory sequence from nearby non-target genes and it can promote transcription by shortening the distance separating
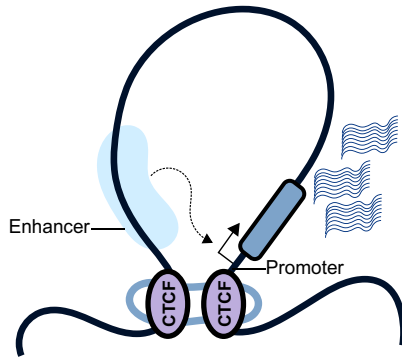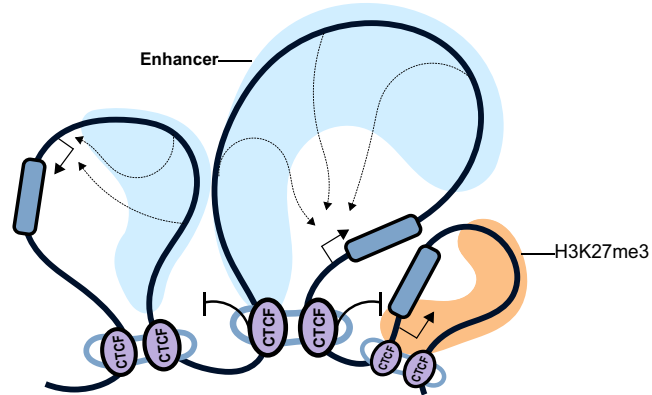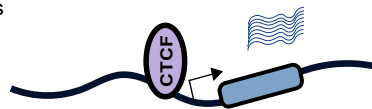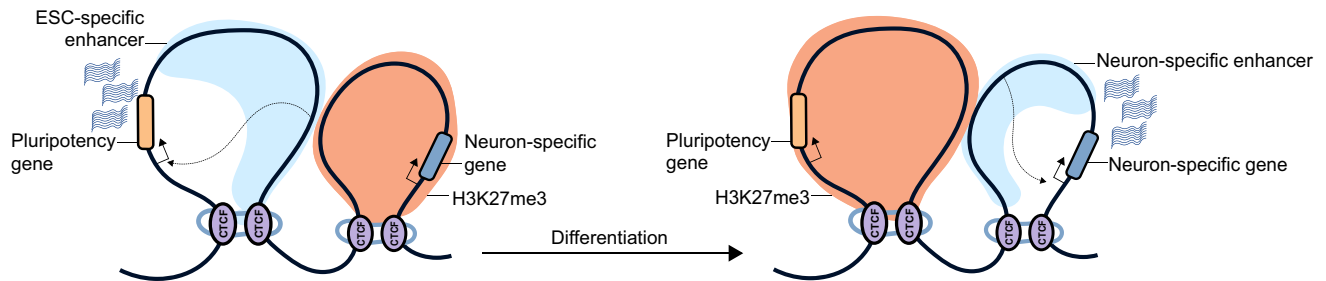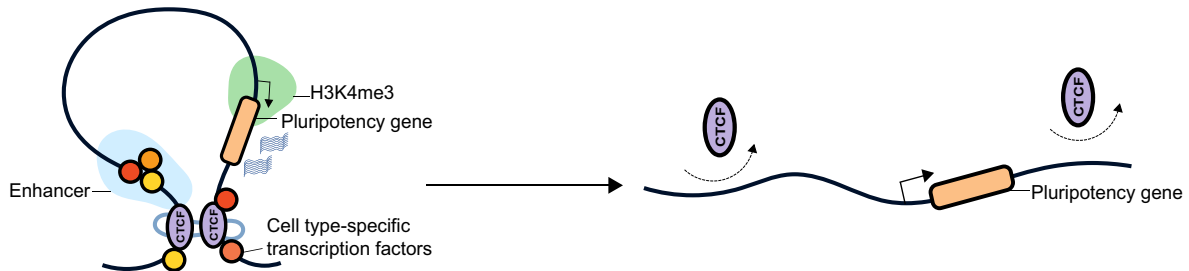
Fig. 4. Modes of CTCF function during cell differentiation and development. (A-C) CTCF can promote the physical interaction between enhancers and promoters (A), insulate genes from neighboring regulatory signals by chromatin looping (B) or mediate transcriptional activation by binding at the promoter regions of certain genes (C). (D) During cell 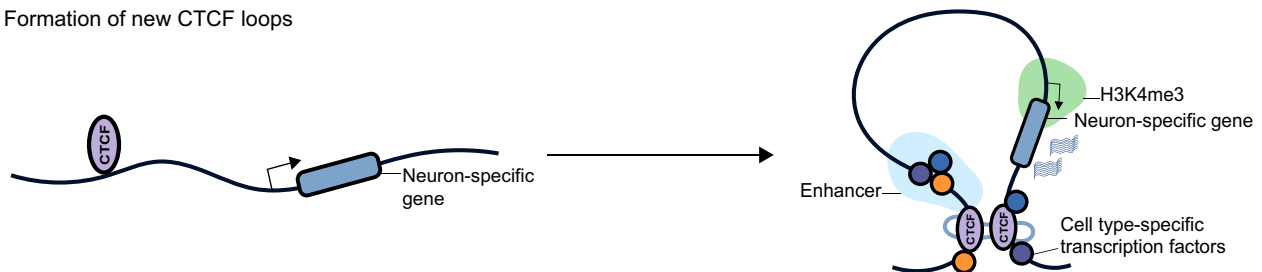differentiation, CTCF can also promote the establishment of a basal, stable topology upon which regulatory transactions can take place. Some chromatin loops can be formed early in development and remain unchanged during differentiation, although genes inside could be subject to differential gene expression owing to the binding of specific transcription factors. (E,F) In addition, some chromatin loops can be lost (E) or formed *de novo* by recruitment of CTCF and perhaps cell type-specific transcription factors (F).

enhancers and promoters. The binding of CTCF to some sites close to promoters can also protect against DNA methylation, therefore protecting the associated gene from transcriptional repression.

Overall, these findings suggest that the transcriptional outcomes of removing CTCF will be highly dependent on context, on its relative location within the 1D genome and on the underlying topology mediated by this protein. For example, in a particular context, CTCF removal may immediately affect genes for which CTCF directly promotes transcription. By contrast, the removal of CTCF from loops containing transcriptionally poised genes in a particular cell type may not result in transcriptional activation because of the absence of a signal that is able to promote transcription in that cell type. However, the presence of an active enhancer sequence nearby could result in activation of transcription. Furthermore, as CTCF binding may signal cohesin to stop at specific sites during loop extrusion, the loss of CTCF could result in cohesin failing to stop at the sites originally occupied by CTCF, and instead continuing to extrude until a loop is formed at a new anchor. In this scenario, regulatory sequences such as enhancers may not be insulated by the original loop, resulting in transcriptional changes. Consequently, the effect of CTCF on gene expression programs will depend on the type of affected genes.

Results from several avenues of research have greatly informed our understating of the mechanisms by which CTCF exerts its regulatory functions and have highlighted how this protein can regulate a variety of cellular processes. However, we still have much to learn. The precise mechanisms by which CTCF integrates cell signals, for example, is still unclear. In addition, the modes by which CTCF regulates the expression of specific genes during differentiation are complex, and it is currently unknown how the interaction of CTCF with cell type-specific transcription factors or other architectural proteins impacts gene expression and genome organization. Detailed knowledge of these functional properties of CTCF will no doubt shed light on the processes controlling gene expression during development at the level of cell types, tissues and the whole organism.

References
Andrey, G., Schöpflin, R., Jerković, I., Heinrich, V., Ibrahim, D. M., Paliou, C., Hochradel, M., Timmermann, B., Haas, S., Vingron, M. et al. (2017). Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome Res.* **27**, 223-233.

Ay, F., Bunnik, E. M., Varoquaux, N., Bol, S. M., Prudhomme, J., Vert, J.-P., Noble, W. S. and Le Roch, K. G. (2014). Three-dimensional modeling of the P. falciparum genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Res.* **24**, 974-988.

Ayala-Ortega, E., Arzate-Mejía, R., Pérez-Molina, R., González-Buendía, E., Meier, K., Guerrero, G. and Recillas-Targa, F. (2016). Epigenetic silencing of miR-181c by DNA methylation in glioblastoma cell lines. *BMC Cancer* **16**, 226.

Balakrishnan, S. K., Witcher, M., Berggren, T. W. and Emerson, B. M. (2012). Functional and molecular characterization of the role of CTCF in human embryonic stem cell biology. *PLoS ONE* **7**, e42424.

Battistelli, C., Busanello, A. and Maione, R. (2014). Functional interplay between MyoD and CTCF in regulating long-range chromatin interactions during differentiation. *J. Cell Sci.* **127**, 3757-3767.

Beagan, J. A., Duong, M. T., Titus, K. R., Zhou, L., Cao, Z., Ma, J., Lachanski, C. V., Gillis, D. R. and Phillips-Cremins, J. E. (2017). YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res.* **27**, 1139-1152.

Bell, A. C. and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**, 482-485.

Bell, A. C., West, A. G. and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**, 387-396.

Bruneau, B. G. (2013). Signaling and transcriptional networks in heart development and regeneration. *Cold Spring Harb. Perspect. Biol.* **5**, a008292.

Busanello, A., Battistelli, C., Carbone, M., Mostocotto, C. and Maione, R. (2012). MyoD regulates p57 kip2 expression by interacting with a distant cis-element and modifying a higher order chromatin structure. *Nucleic Acids Res.* **40**, 8266-8275.

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

Canto-Soler, M. V., Huang, H., Romero, M. S. and Adler, R. (2008). Transcription factors CTCF and Pax6 are segregated to different cell types during retinal cell differentiation. *Dev. Dyn.* **237**, 758-767.

Chen, H., Tian, Y., Shu, W., Bo, X. and Wang, S. (2012). Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PLoS ONE* **7**, e41374.

Chernukhin, I., Shamsuddin, S., Kang, S. Y., Bergström, R., Kwon, Y.-W., Yu, W., Whitehead, J., Mukhopadhyay, R., Docquier, F., Farrar, D. et al. (2007). CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. *Mol. Cell. Biol.* **27**, 1631-1648.

Cho, C., Willis, W. D., Goulding, E. H., Jung-Ha, H., Choi, Y.-C., Hecht, N. B. and Eddy, E. M. (2001). Haploinsufficiency of protamine-1 or -2 causes infertility in mice. *Nat. Genet.* **28**, 82-86.

Choi, N. M., Loguercio, S., Verma-Gaur, J., Degner, S. C., Torkamani, A., Su, A. I., Oltz, E. M., Artyomov, M. and Feeney, A. J. (2013). Deep sequencing of the murine Igh repertoire reveals complex regulation of non-random V gene rearrangement frequencies. *J. Immunol. Baltim. Md 1950* **191**, 2393-2402.

Chung, J. H., Bell, A. C. and Felsenfeld, G. (1997). Characterization of the chicken beta-globin insulator. *Proc. Natl. Acad. Sci. USA* **94**, 575-580.

Crane, E., Bian, Q., McCord, R. P., Lajoie, B. R., Wheeler, B. S., Ralston, E. J., Uzawa, S., Dekker, J. and Meyer, B. J. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240-244.

Cremer, T., Cremer, M., Dietzel, S., Müller, S., Solovei, I. and Fakan, S. (2006). Chromosome territories–a functional nuclear landscape. *Curr. Opin. Cell Biol.* **18**, 307-316.

De Souza, R. A. G., Islam, S. A., McEwen, L. M., Mathelier, A., Hill, A., Mah, S. M., Wasserman, W. W., Kobor, M. S. and Leavitt, B. R. (2016). DNA methylation profiling in human Huntington's disease brain. *Hum. Mol. Genet.* **25**, 2013-2030.

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

Delgado-Olguín, P., Brand-Arzamendi, K., Scott, I. C., Jungblut, B., Stainier, D. Y., Bruneau, B. G. and Recillas-Targa, F. (2011). CTCF promotes muscle differentiation by modulating the activity of myogenic regulatory factors. *J. Biol. Chem.* **286**, 12483-12494.

Delgado-Olguín, P., Huang, Y., Li, X., Christodoulou, D., Seidman, C. E., Seidman, J. G., Tarakhovsky, A. and Bruneau, B. G. (2012). Epigenetic repression of cardiac progenitor gene expression by Ezh2 is required for postnatal cardiac homeostasis. *Nat. Genet.* **44**, 343-347.

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S. and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380.

Donohoe, M. E., Silva, S. S., Pinter, S. F., Xu, N. and Lee, J. T. (2009). The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature* **460**, 128-132.

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

Fedoriw, A. M., Stein, P., Svoboda, P., Schultz, R. M. and Bartolomei, M. S. (2004). Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. *Science* **303**, 238-240.

Franke, M., Ibrahim, D. M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., Kraft, K., Kempfer, R., Jerković, I., Chan, W.-L. et al. (2016). Formation of

DEVELOPMENT

13

new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265-269.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N. and Mirny, L. A. (2016). Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038-2049.

Furlan-Magaril, M., Rebollar, E., Guerrero, G., Fernández, A., Moltó, E., González-Buendía, E., Cantero, M., Montoliu, L. and Recillas-Targa, F. (2011). An insulator embedded in the chicken α-globin locus regulates chromatin domain configuration and differential gene expression. *Nucleic Acids Res.* **39**, 89-103.

Galazka, J. M., Klocko, A. D., Uesaka, M., Honda, S., Selker, E. U. and Freitag, M. (2016). Neurospora chromosomes are organized by blocks of importin alpha-dependent heterochromatin that are largely independent of H3K9me3. *Genome Res.* **26**, 1069-1080.

Garrett, A. M., Schreiner, D., Lobas, M. A. and Weiner, J. A. (2012). γ-protocadherins control cortical dendrite arborization by regulating the activity of a FAK/PKC/MARCKS signaling pathway. *Neuron* **74**, 269-276.

Gerasimova, T. I., Lei, E. P., Bushey, A. M. and Corces, V. G. (2007). Coordinated control of dCTCF and gypsy chromatin insulators in Drosophila. *Mol. Cell* **28**, 761-772.

Gerasimova, T., Guo, C., Ghosh, A., Qiu, X., Montefiori, L., Verma-Gaur, J., Choi, N. M., Feeney, A. J. and Sen, R. (2015). A structural hierarchy mediated by multiple nuclear factors establishes IgH locus conformation. *Genes Dev.* **29**, 1683-1695.

Goloborodko, A., Marko, J. F. and Mirny, L. A. (2016). Chromosome compaction by active loop extrusion. *Biophys. J.* **110**, 2162-2168.

Gomes, N. P. and Espinosa, J. M. (2010). Gene-specific repression of the p53 target gene PUMA via intragenic CTCF–Cohesin binding. *Genes Dev.* **24**, 1022-1034.

Gomez-Velazquez, M., Badia-Careaga, C., Lechuga-Vieco, A. V., Nieto-Arellano, R., Tena, J. J., Rollan, I., Alvarez, A., Torroja, C., Caceres, E. F., Roy, A. R. et al. (2017). CTCF counter-regulates cardiomyocyte development and maturation programs in the embryonic heart. *PLoS Genet.* **13**, e1006985.

Gregor, A., Oti, M., Kouwenhoven, E. N., Hoyer, J., Sticht, H., Ekici, A. B., Kjaergaard, S., Rauch, A., Stunnenberg, H. G., Uebe, S. et al. (2013). De novo mutations in the genome organizer CTCF cause intellectual disability. *Am. J. Hum. Genet.* **93**, 124-131.

Guo, C., Yoon, H. S., Franklin, A., Jain, S., Ebert, A., Cheng, H.-L., Hansen, E., Despo, O., Bossen, C., Vettermann, C. et al. (2011). CTCF-binding elements mediate control of V(D)J recombination. *Nature* **477**, 424-430.

Guo, Y., Monahan, K., Wu, H., Gertz, J., Varley, K. E., Li, W., Myers, R. M., Maniatis, T. and Wu, Q. (2012). CTCF/cohesin-mediated DNA looping is required for protocadherin α promoter choice. *Proc. Natl. Acad. Sci. USA* **109**, 21081-21086.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., Jung, I., Wu, H., Zhai, Y., Tang, Y. et al. (2015). CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162**, 900-910.

Haarhuis, J. H. I., van der Weide, R. H., Blomen, V. A., Yáñez-Cuna, J. O., Amendola, M., van Ruiten, M. S., Krijger, P. H. L., Teunissen, H., Medema, R. H., van Steensel, B. et al. (2017). The cohesin release factor WAPL restricts chromatin loop extension. *Cell* **169**, 693-707.e14.

Han, D., Chen, Q., Shi, J., Zhang, F. and Yu, X. (2017). CTCF participates in DNA damage response via poly(ADP-ribosyl)ation. *Sci. Rep.* **7**, 43530.

Handoko, L., Xu, H., Li, G., Ngan, C. Y., Chew, E., Schnapp, M., Lee, C. W. H., Ye, C., Ping, J. L. H., Mulawadi, F. et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.* **43**, 630-638.

Hanssen, L. L. P., Kassouf, M. T., Oudelaar, A. M., Biggs, D., Preece, C., Downes, D. J., Gosden, M., Sharpe, J. A., Sloane-Stanley, J. A., Hughes, J. R. et al. (2017). Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nat. Cell Biol.* **19**, 952-961.

Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M. and Tilghman, S. M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405**, 486-489.

Hashimoto, H., Wang, D., Horton, J. R., Zhang, X., Corces, V. G. and Cheng, X. (2017). Structural basis for the versatile and methylation-dependent binding of CTCF to DNA. *Mol. Cell* **66**, 711-720.e3.

Heath, H., de Almeida, C. R., Sleutels, F., Dingjan, G., van de Nobelen, S., Jonkers, I., Ling, K.-W., Gribnau, J., Renkawitz, R., Grosveld, F. et al. (2008). CTCF regulates cell cycle progression of αβ T cells in the thymus. *EMBO J.* **27**, 2839-2850.

Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E. and Wiehe, T. (2012). The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc. Natl. Acad. Sci. USA* **109**, 17507-17512.

Hernández-Hernández, A., Lilienthal, I., Fukuda, N., Galjart, N. and Höög, C. (2016). CTCF contributes in a critical way to spermatogenesis and male fertility. *Sci. Rep.* **6**, 28355.

Herzig, Y., Nevo, S., Bornstein, C., Brezis, M. R., Ben-Hur, S., Shkedy, A., Eisenberg-Bord, M., Levi, B., Delacher, M., Goldfarb, Y. et al. (2017). Transcriptional programs that control expression of the autoimmune regulator gene Aire. *Nat. Immunol.* **18**, 161-172.

Hill, R. E., Favor, J., Hogan, B. L. M., Ton, C. C. T., Saunders, G. F., Hanson, I. M., Prosser, J., Jordan, T., Hastie, N. D. and van Heyningen, V. (1992). Mouse Small eye results from mutation in a paired-like homeobox-containing gene. *Nature* **355**, 750.

Hirayama, T. and Yagi, T. (2013). Clustered protocadherins and neuronal diversity. *Prog. Mol. Biol. Transl. Sci.* **116**, 145-167.

Hirayama, T., Tarusawa, E., Yoshimura, Y., Galjart, N. and Yagi, T. (2012). CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. *Cell Rep.* **2**, 345-357.

Hnisz, D., Weintraub, A. S., Day, D. S., Valton, A.-L., Bak, R. O., Li, C. H., Goldmann, J., Lajoie, B. R., Fan, Z. P., Sigova, A. A. et al. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454-1458.

Holohan, E. E., Kwong, C., Adryan, B., Bartkuhn, M., Herold, M., Renkawitz, R., Russell, S. and White, R. (2007). CTCF genomic binding sites in Drosophila and the organisation of the Bithorax Complex. *PLoS Genet.* **3**, e112.

Hori, I., Kawamura, R., Nakabayashi, K., Watanabe, H., Higashimoto, K., Tomikawa, J., Ieda, D., Ohashi, K., Negishi, Y., Hattori, A. et al. (2017). CTCF deletion syndrome: clinical features and epigenetic delineation. *J. Med. Genet.*

Hou, C., Li, L., Qin, Z. S. and Corces, V. G. (2012). Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol. Cell* **48**, 471-484.

Hsieh, T.-H. S., Weiner, A., Lajoie, B., Dekker, J., Friedman, N. and Rando, O. J. (2015). Mapping nucleosome resolution chromosome folding in yeast by micro-C. *Cell* **162**, 108-119.

Hu, J., Zhang, Y., Zhao, L., Frock, R. L., Du, Z., Meyers, R. M., Meng, F., Schatz, D. G. and Alt, F. W. (2015). Chromosomal loop domains direct the recombination of antigen receptor genes. *Cell* **163**, 947-959.

Ishihara, K., Oshimura, M. and Nakao, M. (2006). CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol. Cell* **23**, 733-742.

Ishihara, K., Nakamoto, M. and Nakao, M. (2016). DNA methylation-independent removable insulator controls chromatin remodeling at the HOXA locus via retinoic acid signaling. *Hum. Mol. Genet.* **25**, 5383-5394.

Ji, X., Dadon, D. B., Powell, B. E., Fan, Z. P., Borges-Rivera, D., Shachar, S., Weintraub, A. S., Hnisz, D., Pegoraro, G., Lee, T. I. et al. (2016). 3D chromosome regulatory landscape of human pluripotent cells. *Cell Stem Cell* **18**, 262-275.

Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., Yen, C.-A., Schmitt, A. D., Espinoza, C. A. and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290-294.

Jung, Y. H., Sauria, M. E. G., Lyu, X., Cheema, M. S., Ausio, J., Taylor, J., Corces, V. G., Jung, Y. H., Sauria, M. E. G., Lyu, X. et al. (2017). Chromatin states in mouse sperm correlate with embryonic and adult regulatory landscapes. *Cell Rep.* **18**, 1366-1382.

Juraeva, D., Haenisch, B., Zapatka, M., Frank, J., GROUP Investigators, PSYCH-GEMS SCZ Working Group, Witt, S. H., Mühleisen, T. W., Treutlein, J., Strohmaier, J. et al. (2014). Integrated pathway-based approach identifies association between genomic regions at CTCF and CACNB2 and schizophrenia. *PLoS Genet.* **10**, e1004345.

Kanduri, C., Pant, V., Loukinov, D., Pugacheva, E., Qi, C.-F., Wolffe, A., Ohlsson, R. and Lobanenkov, V. V. (2000). Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr. Biol.* **10**, 853-856.

Kaspi, H., Chapnik, E., Levy, M., Beck, G., Hornstein, E. and Soen, Y. (2013). Brief report: miR-290-295 regulate embryonic stem cell differentiation propensities by repressing Pax6. *Stem Cells Dayt. Ohio* **31**, 2266-2272.

Kemp, C. J., Moore, J. M., Moser, R., Bernard, B., Teater, M., Smith, L. E., Rabaia, N. A., Gurley, K. E., Guinney, J., Busch, S. E. et al. (2014). CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer. *Cell Rep.* **7**, 1020-1029.

Kernohan, K. D., Vernimmen, D., Gloor, G. B. and Bérubé, N. G. (2014). Analysis of neonatal brain lacking ATRX or MeCP2 reveals changes in nucleosome density, CTCF binding and chromatin looping. *Nucleic Acids Res.* **42**, 8356-8368.

Kim, L. K., Esplugues, E., Zorca, C. E., Parisi, F., Kluger, Y., Kim, T. H., Galjart, N. J. and Flavell, R. A. (2014). Oct-1 regulates IL-17 expression by directing interchromosomal associations in conjunction with CTCF in T cells. *Mol. Cell* **54**, 56-66.

Klenova, E. M., Nicolas, R. H., Paterson, H. F., Carne, A. F., Heath, C. M., Goodwin, G. H., Neiman, P. E. and Lobanenkov, V. V. (1993). CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol. Cell. Biol.* **13**, 7612-7624.

Klenova, E. M., Chernukhin, I. V., El-Kady, A., Lee, R. E., Pugacheva, E. M., Loukinov, D. I., Goodwin, G. H., Delgado, D., Filippova, G. N., León, J. et al. (2001). Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. *Mol. Cell. Biol.* **21**, 2221-2234.

Kubo, N., Ishii, H., Gorkin, D., Meitinger, F., Xiong, X., Fang, R., Liu, T., Ye, Z., Li, B., Dixon, J. et al. (2017). Preservation of chromatin organization after acute loss of CTCF in mouse embryonic stem cells. *bioRxiv* 118737.

14

Kung, J. T., Kesner, B., An, J. Y., Ahn, J. Y., Cifuentes-Rojas, C., Colognori, D., Jeon, Y., Szanto, A., del Rosario, B. C., Pinter, S. F. et al. (2015). Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF. Mol. Cell 57, 361-375.

Langston, A. W., Thompson, J. R. and Gudas, L. J. (1997). Retinoic acid-responsive enhancers located 3′ of the Hox A and Hox B homeobox gene clusters. J. Biol. Chem. 272, 2167-2175.

Lee, J., Krivega, I., Dale, R. K. and Dean, A. (2017). The LDB1 complex co-opts CTCF for erythroid lineage-specific long-range enhancer interactions. Cell Rep. 19, 2490-2502.

Li, T. and Lu, L. (2005). Epidermal growth factor-induced proliferation requires down-regulation of Pax6 in corneal epithelial cells. J. Biol. Chem. 280, 12988-12995.

Li, T., Lu, Z. and Lu, L. (2004). Regulation of eye development by transcription control of CCCTC binding factor (CTCF). J. Biol. Chem. 279, 27575-27583.

Lichner, Z., Páll, E., Kerekes, A., Pállinger, E., Maraghechi, P., Bősze, Z. and Gócza, E. (2011). The miR-290-295 cluster promotes pluripotency maintenance by regulating cell cycle phase distribution in mouse embryonic stem cells. Differ. Res. Biol. Divers. 81, 11-24.

Lin, S. G., Guo, C., Su, A., Zhang, Y. and Alt, F. W. (2015). CTCF-binding elements 1 and 2 in the Igh intergenic control region cooperatively regulate V(D)J recombination. Proc. Natl. Acad. Sci. USA 112, 1815-1820.

Liu, Z., Scannell, D. R., Eisen, M. B. and Tjian, R. (2011). Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. Cell 146, 720-731.

Liu, C., Wang, C., Wang, G., Becker, C., Zaidem, M. and Weigel, D. (2016). Genome-wide analysis of chromatin packing in Arabidopsis thaliana at single-gene resolution. Genome Res. 26, 1057-1068.

Lobanenkov, V. V., Nicolas, R. H., Adler, V. V., Paterson, H., Klenova, E. M., Polotskaja, A. V. and Goodwin, G. H. (1990). 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.

Lonfat, N. and Duboule, D. (2015). Structure, function and evolution of topologically associating domains (TADs) at HOX loci. FEBS Lett. 589, 2869-2876.

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

MacPherson, M. J., Beatty, L. G., Zhou, W., Du, M. and Sadowski, P. D. (2009). The CTCF insulator protein is posttranslationally modified by SUMO. Mol. Cell. Biol. 29, 714-725.

Mahadevan, A., Welsh, I. C., Sivakumar, A., Gludish, D. W., Shilvock, A. R., Noden, D. M., Huss, D., Lansford, R. and Kurpios, N. A. (2014). The left-right Pitx2 pathway drives organ-specific arterial and lymphatic development in the intestine. Dev. Cell 31, 690-706.

Majumder, P. and Boss, J. M. (2011). Cohesin regulates major histocompatibility complex class II genes through interactions with MHC-II insulators. J. Immunol. Baltim. Md 1950 187, 4236-4244.

Majumder, P., Gomez, J. A., Chadwick, B. P. and Boss, J. M. (2008). 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.

Majumder, P., Scharer, C. D., Choi, N. M. and Boss, J. M. (2014). B cell differentiation is associated with reprogramming the CCCTC binding factor-dependent chromatin architecture of the murine MHC class II locus. J. Immunol. Baltim. Md 1950 192, 3925-3935.

Mallo, M., Wellik, D. M. and Deschamps, J. (2010). Hox genes and regional patterning of the vertebrate body plan. Dev. Biol. 344, 7-15.

Marina, R. J., Sturgill, D., Bailly, M. A., Thenoz, M., Varma, G., Prigge, M. F., Nanan, K. K., Shukla, S., Haque, N. and Oberdoerffer, S. (2016). TET-catalyzed oxidation of intragenic 5-methylcytosine regulates CTCF-dependent alternative splicing. EMBO J. 35, 335-355.

Marshall, A. D., Bailey, C. G., Champ, K., Vellozzi, M., O'Young, P., Metierre, C., Feng, Y., Thoeng, A., Richards, A. M., Schmitz, U. et al. (2017). CTCF genetic alterations in endometrial carcinoma are pro-tumorigenic. Oncogene 36, 4100-4110.

Martin, D., Pantoja, C., Fernández Miñán, A., Valdes-Quezada, C., Moltó, E., Matesanz, F., Bogdanović, O., de la Calle-Mustienes, E., Domínguez, O., Taher, L. et al. (2011). Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes. Nat. Struct. Mol. Biol. 18, 708-714.

Maurano, M. T., Wang, H., John, S., Shafer, A., Canfield, T., Lee, K. and Stamatoyannopoulos, J. A. (2015). Role of DNA methylation in modulating transcription factor occupancy. Cell Rep. 12, 1184-1195.

Meguro-Horike, M., Yasui, D. H., Powell, W., Schroeder, D. I., Oshimura, M., Lasalle, J. M. and Horike, S.-I. (2011). Neuron-specific impairment of inter-chromosomal pairing and transcription in a novel model of human 15q-duplication syndrome. Hum. Mol. Genet. 20, 3798-3810.

Merkenschlager, M. and Nora, E. P. (2016). CTCF and cohesin in genome folding and transcriptional gene regulation. Annu. Rev. Genomics Hum. Genet. 17, 17-43.

Min, H., Kong, K.-A., Lee, J.-Y., Hong, C.-P., Seo, S.-H., Roh, T.-Y., Bae, S. S. and Kim, M. H. (2016). CTCF-mediated chromatin loop for the posterior Hoxc gene expression in MEF cells. IUBMB Life 68, 436-444.

Mizuguchi, T., Fudenberg, G., Mehta, S., Belton, J.-M., Taneja, N., Folco, H. D., FitzGerald, P., Dekker, J., Mirny, L., Barrowman, J. et al. (2014). Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. Nature 516, 432-435.

Mohan, M., Bartkuhn, M., Herold, M., Philippen, A., Heinl, N., Bardenhagen, I., Leers, J., White, R. A. H., Renkawitz-Pohl, R., Saumweber, H. et al. (2007). The Drosophila insulator proteins CTCF and CP190 link enhancer blocking to body patterning. EMBO J. 26, 4203-4214.

Monahan, K., Rudnick, N. D., Kehayova, P. D., Pauli, F., Newberry, K. M., Myers, R. M. and Maniatis, T. (2012). 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.

Moore, J. M., Rabaia, N. A., Smith, L. E., Fagerlie, S., Gurley, K., Loukinov, D., Disteche, C. M., Collins, S. J., Kemp, C. J., Lobanenkov, V. V. et al. (2012). Loss of maternal CTCF is associated with peri-implantation lethality of Ctcf null embryos. PLoS ONE 7, e34915.

Nakahashi, H., Kwon, K.-R. K., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A. et al. (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Rep. 3, 1678-1689.

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

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

Nichols, M. H. and Corces, V. G. (2015). A CTCF code for 3D genome architecture. Cell 162, 703-705.

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., Sedat, J. et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485, 381-385.

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

Ong, C.-T. and Corces, V. G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat. Rev. Genet. 15, 234-246.

Oti, M., Falck, J., Huynen, M. A. and Zhou, H. (2016). CTCF-mediated chromatin loops enclose inducible gene regulatory domains. BMC Genomics 17, 252.

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

Pope, B. D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D. L., Wang, Y., Hansen, R. S., Canfield, T. K. et al. (2014). Topologically associating domains are stable units of replication-timing regulation. Nature 515, 402-405.

Prickett, A. R., Barkas, N., McCole, R. B., Hughes, S., Amante, S. M., Schulz, R. and Oakey, R. J. (2013). Genome-wide and parental allele-specific analysis of CTCF and cohesin DNA binding in mouse brain reveals a tissue-specific binding pattern and an association with imprinted differentially methylated regions. Genome Res. 23, 1624-1635.

Proudhon, C., Hao, B., Raviram, R., Chaumeil, J. and Skok, J. A. (2015). Long range regulation of V(D)J recombination. Adv. Immunol. 128, 123-182.

Qiu, Z., Song, C., Malakouti, N., Murray, D., Hariz, A., Zimmerman, M., Gygax, D., Alhazmi, A. and Landry, J. W. (2015). Functional interactions between NURF and Ctcf regulate gene expression. Mol. Cell. Biol. 35, 224-237.

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S. et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665-1680.

Rao, S. S. P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K.-R., Sanborn, A. L., Johnstone, S. E., Bascom, G. D., Bochkov, I. D. et al. (2017). Cohesin loss eliminates all loop domains. Cell 171, 305-320.e24.

Recillas-Targa, F., Pikaart, M. J., Burgess-Beusse, B., Bell, A. C., Litt, M. D., West, A. G., Gaszner, M. and Felsenfeld, G. (2002). Position-effect protection and enhancer blocking by the chicken -globin insulator are separable activities. Proc. Natl. Acad. Sci. USA 99, 6883-6888.

Rhee, H. S. and Pugh, B. F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell 147, 1408-1419.

Rodríguez-Carballo, E., Lopez-Delisle, L., Zhan, Y., Fabre, P. J., Beccari, L., El-Idrissi, I., Huynh, T. H. N., Ozadam, H., Dekker, J. and Duboule, D. (2017). The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. Genes Dev. 31, 2264-2281.

Rowley, M. J., Nichols, M. H., Lyu, X., Ando-Kuri, M., Rivera, I. S. M., Hermetz, K., Wang, P., Ruan, Y. and Corces, V. G. (2017). Evolutionarily conserved principles predict 3D chromatin organization. *Mol. Cell.* **67**, 1-16.

Roy, A. R., Ahmed, A., DiStefano, P. V., Chi, L., Khyzha, N., Galjart, N., Wilson, M. D., Fish, J. E. and Delgado-Olguin, P. (2018). The transcriptional regulator CCCTC-binding factor limits oxidative stress in endothelial cells. *J. Biol. Chem.* (in press).

Saldaña-Meyer, R., González-Buendía, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F. and Reinberg, D. (2014). CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev.* **28**, 723-734.

Sams, D. S., Nardone, S., Getselter, D., Raz, D., Tal, M., Rayi, P. R., Kaphzan, H., Hakim, O. and Elliott, E. (2016). Neuronal CTCF is necessary for basal and experience-dependent gene regulation, memory formation, and genomic structure of BDNF and arc. *Cell Rep.* **17**, 2418-2430.

Sanborn, A. L., Rao, S. S. P., Huang, S.-C., Durand, N. C., Huntley, M. H., Jewett, A. I., Bochkov, I. D., Chinnappan, D., Cutkosky, A., Li, J. et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. USA* **112**, E6456-E6465.

Sansom, S. N., Shikama-Dorn, N., Zhanybekova, S., Nusspaumer, G., Macaulay, I. C., Deadman, M. E., Heger, A., Ponting, C. P. and Holländer, G. A. (2014). Population and single-cell genomics reveal the Aire dependency, relief from Polycomb silencing, and distribution of self-antigen expression in thymic epithelia. *Genome Res.* **24**, 1918-1931.

Schmidt, D., Schwalie, P. C., Wilson, M. D., Ballester, B., Gonçalves, A., Kutter, C., Brown, G. D., Marshall, A., Flicek, P. and Odom, D. T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148**, 335-348.

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N. A., Huber, W., Haering, C., Mirny, L. et al. (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51-56. advance online publication.

Shiratori, H., Sakuma, R., Watanabe, M., Hashiguchi, H., Mochida, K., Sakai, Y., Nishino, J., Saijoh, Y., Whitman, M. and Hamada, H. (2001). Two-step regulation of left-right asymmetric expression of Pitx2: initiation by nodal signaling and maintenance by Nkx2. *Mol. Cell* **7**, 137-149.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R. and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74-79.

Sopher, B. L., Ladd, P. D., Pineda, V. V., Libby, R. T., Sunkin, S. M., Hurley, J. B., Thienes, C. P., Gaasterland, T., Filippova, G. N. and La Spada, A. R. (2011). CTCF regulates ataxin-7 expression through promotion of a convergently transcribed, antisense noncoding RNA. *Neuron* **70**, 1071-1084.

Soshnikova, N., Montavon, T., Leleu, M., Galjart, N. and Duboule, D. (2010). Functional analysis of CTCF during mammalian limb development. *Dev. Cell* **19**, 819-830.

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

Sun, S., Del Rosario, B. C., Szanto, A., Ogawa, Y., Jeon, Y. and Lee, J. T. (2013). Jpx RNA activates Xist by evicting CTCF. *Cell* **153**, 1537-1551.

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

Tang, M., Chen, B., Lin, T., Li, Z., Pardo, C., Pampo, C., Chen, J., Lien, C.-L., Wu, L., Ai, L. et al. (2011). Restraint of angiogenesis by zinc finger transcription factor CTCF-dependent chromatin insulation. *Proc. Natl. Acad. Sci. USA* **108**, 15231-15236.

Tang, Z., Luo, O. J., Li, X., Zheng, M., Zhu, J. J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B. et al. (2015). CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* **163**, 1611-1627.

Tapscott, S. J. (2005). The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. *Development* **132**, 2685-2695.

Tee, W.-W. and Reinberg, D. (2014). Chromatin features and the epigenetic regulation of pluripotency states in ESCs. *Development* **141**, 2376-2390.

Teif, V. B., Beshnova, D. A., Vainshtein, Y., Marth, C., Mallm, J.-P., Höfer, T. and Rippe, K. (2014). Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development. *Genome Res.* **24**, 1285-1295.

Teperek, M., Simeone, A., Gaggioli, V., Miyamoto, K., Allen, G. E., Erkek, S., Kwon, T., Marcotte, E. M., Zegerman, P., Bradshaw, C. R. et al. (2016). Sperm is epigenetically programmed to regulate gene transcription in embryos. *Genome Res.* **26**, 1034-1046.

Tsui, S., Wang, J., Wang, L., Dai, W. and Lu, L. (2016). CTCF-mediated and Pax6-associated gene expression in corneal epithelial cell-specific differentiation. *PLoS ONE* **11**, e0162071.

Uusküla-Reimand, L., Hou, H., Samavarchi-Tehrani, P., Rudan, M. V., Liang, M., Medina-Rivera, A., Mohammed, H., Schmidt, D., Schwalie, P., Young, E. J. et al. (2016). Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. *Genome Biol.* **17**, 182.

Valadez-Graham, V., Razin, S. V. and Recillas-Targa, F. (2004). CTCF-dependent enhancer blockers at the upstream region of the chicken -globin gene domain. *Nucleic Acids Res.* **32**, 1354-1362.

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

Wan, L.-B., Pan, H., Hannenhalli, S., Cheng, Y., Ma, J., Fedoriw, A., Lobanenkov, V., Latham, K. E., Schultz, R. M. and Bartolomei, M. S. (2008). Maternal depletion of CTCF reveals multiple functions during oocyte and preimplantation embryo development. *Development* **135**, 2729-2738.

Wang, H., Maurano, M. T., Qu, H., Varley, K. E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R. et al. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* **22**, 1680-1688.

Wang, H., Zang, C., Taing, L., Arnett, K. L., Wong, Y. J., Pear, W. S., Blacklow, S. C., Liu, X. S. and Aster, J. C. (2014). NOTCH1–RBPJ complexes drive target gene expression through dynamic interactions with superenhancers. *Proc. Natl. Acad. Sci. USA* **111**, 705-710.

Wang, F., Han, J., Wang, L., Jing, Y., Zhu, Z., Hui, D., Wang, Z., Wang, Y., Dong, Y. and Tan, T. (2017). CCCTC-binding factor (CTCF) transcriptionally targets Wdr5 to mediate somatic cell reprogramming. *Stem Cells Dev.* **26**, 743-750.

Watanabe, H., Schmidt, H. A., Kuhn, A., Höger, S. K., Kocagöz, Y., Laumann-Lipp, N., Özbek, S. and Holstein, T. W. (2014). Nodal signalling determines biradial asymmetry in Hydra. *Nature* **515**, 112-115.

Watson, L. A., Wang, X., Elbert, A., Kernohan, K. D., Galjart, N. and Bérubé, N. G. (2014). Dual effect of CTCF loss on neuroprogenitor differentiation and survival. *J. Neurosci. Off. J. Soc. Neurosci.* **34**, 2860-2870.

Weintraub, H., Tapscott, S. J., Davis, R. L., Thayer, M. J., Adam, M. A., Lassar, A. B. and Miller, A. D. (1989). Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. *Proc. Natl. Acad. Sci. USA* **86**, 5434-5438.

Weintraub, A. S., Li, C. H., Zamudio, A. V., Sigova, A. A., Hannett, N. M., Day, D. S., Abraham, B. J., Cohen, M. A., Nabet, B., Buckley, D. L. et al. (2017). YY1 is a structural regulator of enhancer-promoter loops. *Cell* **171**, 1573-1588.e28.

Welsh, I. C., Kwak, H., Chen, F. L., Werner, M., Shopland, L. S., Danko, C. G., Lis, J. T., Zhang, M., Martin, J. F. and Kurpios, N. A. (2015). Chromatin architecture of the Pitx2 locus requires CTCF- and Pitx2-dependent asymmetry that mirrors embryonic gut laterality. *Cell Rep.* **13**, 337-349.

Will, A. J., Cova, G., Osterwalder, M., Chan, W.-L., Wittler, L., Brieske, N., Heinrich, V., de Villartay, J.-P., Vingron, M., Klopocki, E. et al. (2017). Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). *Nat. Genet.* **49**, 1539-1545.

Woltering, J. M., Noordermeer, D., Leleu, M. and Duboule, D. (2014). Conservation and divergence of regulatory strategies at Hox loci and the origin of tetrapod digits. *PLoS Biol.* **12**, e1001773.

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

Xiao, T., Wongtrakoongate, P., Trainor, C. and Felsenfeld, G. (2015). CTCF recruits centromeric protein CENP-E to the pericentromeric/centromeric regions of chromosomes through unusual CTCF-binding sites. *Cell Rep.* **12**, 1704-1714.

Yu, N.-K., Kim, H. F., Shim, J., Kim, S., Kim, D. W., Kwak, C., Sim, S.-E., Choi, J.-H., Ahn, S., Yoo, J. et al. (2016). A transducible nuclear/nucleolar protein, mLLP, regulates neuronal morphogenesis and synaptic transmission. *Sci. Rep.* **6**, 22892.

Yu, W, Ginjala, V., Pant, V., Chernukhin, I., Whitehead, J., Docquier, F., Farrar, D., Tavoosidana, G., Mukhopadhyay, R., Kanduri, C. et al. (2004). Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. *Nat. Genet.* **36**, 1105-1110.

Zheng, G. X. Y., Ravi, A., Calabrese, J. M., Medeiros, L. A., Kirak, O., Dennis, L. M., Jaenisch, R., Burge, C. B. and Sharp, P. A. (2011). A latent pro-survival function for the Mir-290-295 cluster in mouse embryonic stem cells. *PLoS Genet.* **7**, e1002054.