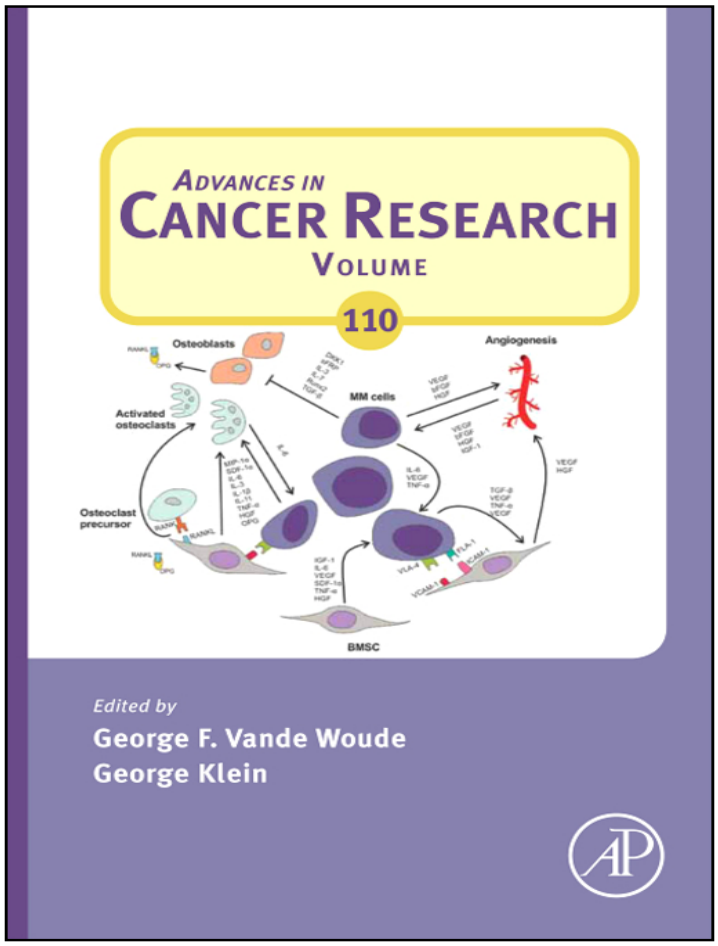

From Victor G. and Corces Jingping Yang, Chromatin Insulators: A Role in Nuclear Organization and Gene Expression. In: George F. Vande Woude and George Klein, editors: Advances in Cancer Research, Vol 110, Oxford: Academic Press; 2011, p. 43-76.
ISBN:978-0-12-386469-7

Chromatin Insulators: A Role in Nuclear Organization and Gene Expression

Jingping Yang and Victor G. Corces

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

Chromatin insulators are DNA–protein complexes with broad functions in nuclear biology. Based on the ability of insulator proteins to interact with each other, it was originally found that insulators form loops that bring together distant regions of the genome. Data from genome-wide localization studies indicate that insulator proteins can be present in intergenic regions as well as at the 5′, introns or 3′ of genes, suggesting a variety of roles for insulator loops in chromosome biology. Recent results suggest that insulators mediate intra- and interchromosomal interactions to affect transcription, imprinting, and recombination. Cells have developed mechanisms to control insulator activity by recruiting specialized proteins or by covalent modification of core components. It is then possible that insulator-mediated interactions set up cell-specific blueprints of nuclear organization that may contribute to the establishment of different patterns of gene expression during cell differentiation and development. As a consequence, disruption of insulator activity could result in the development of cancer or other disease states.

Advances in CANCER RESEARCH, Volume 110
0065-230X/10 $35.00
DOI: 10.1016/B978-0-12-386469-7.00003-7

I. INTRODUCTION

Patterns of transcription required for cell differentiation are initially established by specific transcription factors. The maintenance of these patterns of gene expression is then carried out by alterations in chromatin structure that are epigenetically inherited between cell generations. These changes in chromatin organization take place at the level of the 10 nm fiber and include covalent histone modifications, DNA methylation, and alterations induced by ATP-dependent remodeling complexes. In addition, recent evidence suggests that the higher order three-dimensional organization of the eukaryotic genome may also be critical for achieving proper spatial and temporal patterns of gene expression during development. The factors and processes involved in the establishment, maintenance, and regulation of specific states of nuclear organization are largely unknown, but insulators are emerging as likely candidates to play this crucial role (Bushey *et al.*, 2008; Phillips and Corces, 2009).

Insulators are DNA–protein complexes experimentally defined by their ability to block enhancer–promoter interactions and/or to serve as barriers against the spreading of the silencing effects of heterochromatin. Not all sequences described as insulators display these two properties; Felsenfeld and colleagues have proposed to classify these sequences as ''enhancer blocking'' or ''barrier'' insulators depending on whether they interfere with enhancer–promoter interactions or antagonize spreading of heterochromatin under experimental conditions (Gaszner and Felsenfeld, 2006). Barrier insulators appear to function by recruiting enzymes involved in histone modifications that stop the spreading of silencing marks (Dhillon *et al.*, 2009; Huang *et al.*, 2007) and will not be considered in this review. Instead, we concentrate our discussion on the function of enhancer-blocking insulators with special emphasis on recent data supporting a role in nuclear organization. These insulators were originally discovered in *Drosophila*, which has been a particularly good model system in which to analyze insulator function. Several different insulators have been identified in this organism whereas vertebrates appear to mostly rely on the CTCF insulator (Phillips and Corces, 2009). We conclude by proposing that the primary role of insulators may not be to regulate enhancer–promoter interactions or heterochromatin spreading. Rather, insulators may mediate intra- and interchromosomal interactions with the primary goal of organizing the eukaryotic genome into epigenetically heritable states. This insulator-mediated organization may be important to regulate DNA function at multiple levels, including transcription initiation, elongation, and DNA recombination. Miss-regulation of insulator

function may lead to alterations in gene expression and the development of disease states and cancer.

II. INSULATOR STRUCTURE: CORE COMPONENTS

There are several types of insulators in *Drosophila* that have been studied in detail. They include the scs and scs′ sequences originally discovered flanking the heat shock *hsp70* locus (Kellum and Schedl, 1992; Zhao *et al.*, 1995), the *gypsy* insulator first found in the *gypsy* retrotransposon (Geyer and Corces, 1992; Hoover *et al.*, 1993), the Fab-7/Fab-8/Mcp insulators located in the bithorax complex (Gyurkovics *et al.*, 1990; Karch *et al.*, 1994), and the SF1 insulator described in the Antennapedia complex (Belozerov *et al.*, 2003). Each of these insulators consists of a DNA sequence and a specific DNA binding protein that interacts with this sequence; additional factors that interact with the DNA binding protein have been characterized for some of these insulators (Fig. 1).

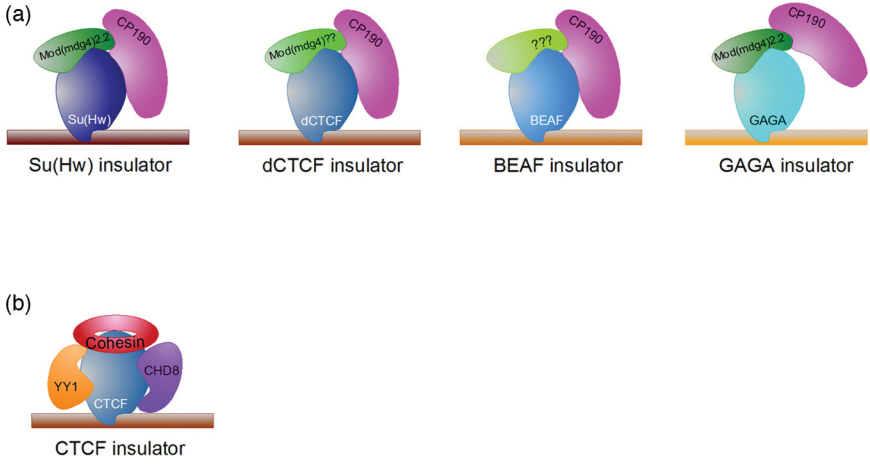

Fig. 1 Diagram showing the structure of different *Drosophila* and vertebrate insulators. (A) Each *Drosophila* insulator subclass contains a different binding protein that may define the specific function of the corresponding subclass. All insulators share the common protein CP190, although the role of this protein in the function of the GAGA insulator has not been demonstrated experimentally. In addition, all subclasses may also have one Mod(mdg4) isoform. The gypsy/Su(Hw) insulator contains Mod(mdg4)2.2. The dCTCF insulator lacks this isoform but contains a different variant of Mod(mdg4) (T. Gerasimova, E. Lei, and V. Corces, unpublished observations). It is not known whether the BEAF insulator contains a Mod(mdg4) variant but GAGA has been shown to interact with Mod (mdg4)2.2. (B) Structure of the vertebrate CTCF insulator.

In the case of the scs insulator, the DNA binding protein component is Zeste-White 5 (ZW5), which is a zinc finger protein required for cell viability. Null mutations in the *zw5* gene are recessive lethal, but hypomorphic alleles display a variety of pleiotropic effects on wing, bristle, and eye development (Gaszner *et al.*, 1999). The *scs′* sequences interact with a protein called Boundary Element Associated Factor 32 (BEAF 32). The *BEAF 32* gene encodes two different proteins named BEAF 32A and BEAF 32B, which are present at hundreds of sites on *Drosophila* polytene chromosomes (Zhao *et al.*, 1995). The two isoforms differ at the N-terminal DNA binding domain (BED finger domain). The common C-terminal region is involved in protein–protein interactions between the two isoforms. Analysis of mutations in the *BEAF 32* gene shows that BEAF 32B is required for viability whereas BEAF 32A mutations do not show significant phenotypic defects. Expression of a dominant negative form of BEAF 32 results in changes in chromosome structure and cell viability (Gilbert *et al.*, 2006; Roy *et al.*, 2007a, 2007b).

The *gypsy* insulator contains binding sites for Suppressor of Hairy-wing [Su(Hw)], which is a 12 zinc finger DNA binding protein. Mutations in the *su(Hw)* gene cause female sterility but do not result in lethality (Harrison *et al.*, 1993). Su(Hw) interacts with two other components of the *gypsy* insulator, Mod(mdg4)2.2 and CP190 (Gause *et al.*, 2001; Ghosh *et al.*, 2001). Mod(mdg4)2.2 does not bind to DNA directly but interacts with Su(Hw) through its carboxy-terminal domain. In addition, Mod (mdg4)2.2 contains a BTB domain in the N-terminal domain that mediates homo- and heteromultimerization with other insulator components. The *mod(mdg4)* gene encodes approximately 29 different isoforms that arise by alternative *cis*- and *trans*-splicing (Buchner *et al.*, 2000; Labrador and Corces, 2003; Labrador *et al.*, 2001); null mutations in the gene result in lethality, but mutations affecting the Mod(mdg4)2.2 isoform are viable and show defects in *gypsy* insulator function (Gerasimova *et al.*, 1995). CP190 also contains a BTB domain as well as 3 zinc fingers and it interacts with both Su(Hw) and Mod(mdg4)2.2. The BTB domains of Mod(mdg4) 2.2 and CP190 also mediate homo-, and multimerization of these two proteins. CP190 binds DNA with low affinity and specificity but it does not interact directly with insulator sequences present in the *gypsy* retrotransposon, where it is recruited through interactions with Su(Hw) and Mod(mdg4)2.2 instead. Mutations in the *CP190* gene are lethal (Pai *et al.*, 2004).

The bithorax complex of *Drosophila* contains an intricate collection of transcriptional regulatory sequences that orchestrate the complex spatiotemporal expression of the three genes present in the complex. The proper interplay between these regulatory sequences requires the function of

several insulators of which Fab-8 has been studied in most detail. Fab-8 sequences interact with the *Drosophila* homolog of the vertebrate CTCF insulator protein. *Drosophila* CTCF (dCTCF) has 11 zinc fingers. Mutations in *dCTCF* are lethal and show abdominal hometic phenotypes (Gerasimova *et al.*, 2007; Mohan *et al.*, 2007). dCTCF is also found in the Mcp and Fab-6 insulators present in the bithorax complex but not Fab-7 (Holohan *et al.*, 2007). Fab-7 may represent a fourth class of insulators that use the GAGA factor (GAF) as a DNA binding protein that also contains a BTB domain. Mutations in the *trl* gene, which encodes GAF, affect Fab-7 insulator activity (Schweinsberg *et al.*, 2004). In addition, GAF is present and required for the function of the SF1 insulator found in the Antennapedia complex (Belozerov *et al.*, 2003).

If the functions of all *Drosophila* insulators have converged into that of CTCF in vertebrates, one may expect to find some shared protein components among *Drosophila* insulators. This is indeed the case. The CP190 protein, first found in the *gypsy*/Su(Hw) insulator, also interacts with dCTCF (Gerasimova *et al.*, 2007; Mohan *et al.*, 2007). Genome-wide mapping of dCTCF and CP190 sites supports this conclusion (Bartkuhn *et al.*, 2009; Bushey *et al.*, 2009). These studies have also shown that BEAF and CP190 colocalize at hundreds of sites throughout the genome (Bushey *et al.*, 2009). These results suggest that the insulators defined by these three different DNA binding proteins, Su(Hw), dCTCF, and BEAF share the BTB domain-containing protein CP190 and may therefore use similar mechanisms to effect their insulator function. On the other hand, GAF does not appear to interact directly with CP190 but has been shown to interact with Su(Hw) and Mod(mdg4)2.2 (Melnikova *et al.*, 2004); since these two proteins can in turn interact with CP190, GAF insulators may act mechanistically like the other three types (Fig. 1).

Vertebrates appear to rely mostly on the widespread CTCF insulator (Wallace and Felsenfeld, 2007) and we will concentrate our discussion on this sequence. CTCF is a highly conserved protein containing an 11 zinc finger central DNA binding domain, displaying close to 100% homology between mouse, chicken, and human, embedded within slightly more divergent N- and C-termini (Ohlsson *et al.*, 2001). On the basis of its ability to bind to a wide range of variant sequences as well as specific coregulatory proteins through combinatorial use of different zinc fingers, CTCF was originally described as a multivalent factor (Filippova *et al.*, 1996). This structural feature may explain the ability of vertebrate cells to fulfill all insulator functions with just one insulator whereas *Drosophila* cells require multiple insulators. CTCF homozygous knockout mice exhibit early embryonic lethality prior to implantation, highlighting the critical importance of CTCF in diverse cellular processes (Splinter *et al.*, 2006).

Interestingly, the Mod(mdg4) and CP190 proteins have not been conserved in vertebrates. Instead, the vertebrate CTCF protein interacts with cohesin, which has recently emerged as a critical partner of CTCF in mediating inter- and intrachromosomal interactions necessary for transcription and insulator function (Fig. 1). Cohesin is a ring-shaped complex that holds chromatids together between the S and M phases of the cell cycle. In vertebrates, cohesins are present at CTCF target sequences in the genome and this localization depends on CTCF (Parelho *et al.*, 2008; Rubio *et al.*, 2008; Stedman *et al.*, 2008; Wendt *et al.*, 2008). Cohesin is necessary for nearly all the intrachromosomal interactions attributed to CTCF, such as those observed at the *H19/Igf2*, *β-globin*, *IFNG*, and *APO* loci; knockdown of cohesin impairs chromatin interactions at these loci (Hadjur *et al.*, 2009; Hou *et al.*, 2010; Mishiro *et al.*, 2009; Nativio *et al.*, 2009). These observations have led to the proposal that the ring-like structure of the cohesin complex can mediate interactions between distant genomic sites by a mechanism similar to that employed to maintain sister chromatids together.

III. INSULATOR STRUCTURE: ACCESSORY FACTORS

CTCF has been found to interact with other nuclear factors in addition to cohesins, although the broad relevance of some of these interactions to general aspects of CTCF function is unclear. For example, CTCF interacts *in vivo* with CHD8, which is a member of the chromodomain helicase (CHD) family and has an SNF2-like helicase/ATPase domain. CHD8 has been shown to be present at some CTCF target sites, including the *H19/Igf2* Imprinting Control Region (ICR), the promoter regions of the *BRCA1* and *c-myc* genes, and the 5′HS5 insulator of the *β-globin* locus in human cells (Ishihara *et al.*, 2006). Like CTCF, CHD8 is also required for the insulator activity of the *H19/Igf2* ICR. Loss of CHD8 leads to expression of *Igf2* from the maternal and paternal alleles (Ishihara *et al.*, 2006). CHD8 together with CTCF can affect aspects of chromatin structure such as CpG DNA methylation and histone acetylation. Knockdown of CHD8 leads to hypermethylation of CpGs at CTCF binding sites in the promoter regions of the *BRCA1* and *c-myc* genes (Ishihara *et al.*, 2006). In addition to CHD8, CTCF can also interact *in vitro* with SIN3A, which recruits a histone deacetylase activity necessary for the silencing function of CTCF (Lutz *et al.*, 2000); this result has not been confirmed by *in vivo* experiments. These observations suggest that CTCF may have context-dependent functions that are mediated by different protein partners.

CTCF has been also found to interact with transcription factors such as YY1, which is a zinc finger protein capable of activating or repressing transcription depending on the promoter context. Like CTCF, YY1 also binds to DNA in a methylation-sensitive manner. YY1 can bind the *Peg3* ICR and regulate its parent of origin-dependent expression. Insulator assays with deletion constructs of YY1 binding sites indicate that the region functions as a methylation-sensitive insulator like the *H19/Igf2* ICR (Kim *et al.*, 2003). YY1 is a required cofactor for CTCF in processes such as X-chromosome inactivation (XCI). Both YY1 and CTCF bind at *Tsix* DNA, and a deficiency of YY1 leads to aberrant *Tsix* and *Xist* expression, aberrant XCI patterns, and results in a deficit of male and female embryos. YY1 and CTCF together function as *Tsix* transcriptional activators in ES cells (Donohoe *et al.*, 2007). It is not yet clear whether YY1 alone can act as an insulator or only contributes to the insulator activity of CTCF as a cofactor. In addition to YY1, CTCF can interact with other transcription factors such as YB1, Oct4, Kaiso, and thyroid hormone receptor (TR). CTCF interacts with YB1 *in vitro* and *in vivo* and functions as a corepressor at the *myc* promoter (Chernukhin *et al.*, 2000). CTCF interacts with Kaiso *in vivo* and both bind to the chicken HS4 insulator. Kasio inhibits the enhancer-blocking activity of CTCF in transgenic assays (Defossez *et al.*, 2005). CTCF interacts with Oct4 and controls X-chromosome pairing during X-chromosome inactivation (Donohoe *et al.*, 2009). A subset of CTCF binding sites is found next to TR response elements in the mouse *c-myc* and the human *APP* genes. Functional TR binding is required for enhancer-blocking activity (Weth *et al.*, 2010). Intriguingly, CTCF has been also found to interact with RNA polymerase II. The largest subunit of Pol II associates with CTCF target sites *in vivo* and this interaction is dependent on the presence of intact CTCF binding sequences. A single CTCF binding site in the *c-myc* gene (N-Myc-CTCF), but not a mutant site deficient for CTCF binding, is sufficient to activate transcription from a promoterless reporter gene in stably transfected cells (Chernukhin *et al.*, 2007). This suggests that CTCF may recruit Pol II and activate transcription. The extent of such a role for CTCF is unclear, since genome-wide studies of CTCF sites suggest that only a subset of CTCF sites are located at promoters in the human genome (Kim *et al.*, 2007).

The function of *Drosophila* and vertebrate insulators appears to rely in part on their ability to interact with the nuclear lamina. *Drosophila* insulator proteins fractionate with components of the nuclear lamina, suggesting a direct or indirect interaction between the two (Byrd and Corces, 2003). Furthermore, the dTopors protein, which is located in the nuclear periphery, interacts with both Mod(mdg4)2.2 and Lamin, and mutations

in the *Lamin Dm0* gene result in disruption of insulator activity (Capelson and Corces, 2005). It appears that interaction with a nuclear substrate is also required for the proper function of the CTCF insulator. For example, it has been shown that CTCF interacts with nucleophosmin, a nuclear matrix protein that is concentrated at the surface of the nucleolus and associates with CTCF at the chicken HS4 insulator in the *β-globin* locus. Consistent with this distribution, transgenic copies of reporter genes containing the chicken HS4 insulator show preferential localization to the outer part of the nucleolus, and this localization depends on the presence of an intact CTCF binding site (Yusufzai *et al.*, 2004). In addition to nucleophosmin, other nuclear matrix proteins such as Lamin may also be important for CTCF function. In human cells, many of the Lamin B1-associated domains (LADs) in the genome are flanked by CTCF (Guelen *et al.*, 2008). Lamin A also works together with CTCF to establish nucleolar localization. The D4Z4 human subtelomeric repeats localize telomeres to the nuclear periphery and display insulator activity in transgenic cell lines. CTCF and Lamin A both bind to D4Z4 and are required for proper perinuclear localization. Reduction of either protein results in a decrease in the localization of telomeres at the nuclear periphery and impairs insulator function (Ottaviani *et al.*, 2009a, 2009b). It is thus possible that the interaction between CTCF and other nuclear matrix proteins contributes to chromatin localization and organization in the nucleus in a variety of cell processes.

These results suggest that *Drosophila* and vertebrate insulators may have a variety of functions in nuclear biology. Many of these functions can be explained by the ability of CTCF and *Drosophila* insulator proteins to mediate interactions between two or more chromosomal locations. Different *Drosophila* insulators may have specialized roles in effecting distinct nuclear functions based on the presence of particular protein components. In vertebrates, it is not clear whether CTCF acts by different mechanisms that are context dependent and vary based on its association with the different partners described above. It is nevertheless possible that these different functions have a common requirement for the ability of CTCF to mediate interactions between distant sites in the genome; this may be also the case in some of the more classical roles characteristic of standard transcription activators and repressors. Although the DNA recognition sequence of CTCF appears to be conserved at different genomic locations (Kim *et al.*, 2007), the distinct context-dependent roles of this protein must be mediated by the different protein partners with which it associates. The nature of the determinants responsible for these context-dependent interactions is unknown and should be an important issue for future investigation.

IV. GENOMIC DISTRIBUTION OF INSULATORS

The existence of several *Drosophila* insulator subclasses with different DNA binding proteins but sharing some functional components raises the question of whether they all have the same role in the regulation of gene expression or whether there is a functional specialization in their tasks. The possibility of such specialization is highlighted by results showing differential localization of insulator subclasses with respect to genomic landmarks. For example, Su(Hw) and dCTCF are preferentially excluded from exonic regions (mostly 5′ and 3′UTRs), with only 8%, 16%, and 17% of sites found within exons, respectively, whereas BEAF sites are enriched in UTRs (Bushey *et al.*, 2009). When the location of these proteins is compared with the location of genes, few Su(Hw) binding sites are found in the 1 kb regions flanking genes. However, dCTCF and BEAF show a distribution that is highly skewed toward the 5′-end of genes and is enriched in the first 200 bp just upstream of the transcription start site (TSS). Insulator proteins also show a compartmentalized distribution in relation to the level of gene expression. For example, 83% of dCTCF sites and 89% of BEAF sites at the 5′-end of genes localize to genes that are highly expressed. However, Su(Hw) binding sites are most often found near genes with low expression levels. Finally, different insulator proteins appear to associate with genes involved in different cellular processes. Genes containing dCTCF in the 200 bp region upstream of their TSS are mostly involved in development, whereas genes containing BEAF in this region are mostly involved in metabolism. Both dCTCF and BEAF are enriched near or at genes involved in cell cycle, whereas Su(Hw)-containing genes show little significant clustering based on biological process (Bushey *et al.*, 2009; Emberly *et al.*, 2008; Jiang *et al.*, 2009). These observations suggest a division of labor among *Drosophila* insulators, both with respect to gene function as well as specific aspects of cell function. Given the fact that all insulators share CP190 and perhaps Mod(mdg4), it is likely that all of them use the same mechanism to perform their function, namely bring together different regions of the genome. Nevertheless, *Drosophila* cells appear to use a variety of DNA binding insulator proteins to recruit other insulator components to mediate these interactions. Therefore, the specific outcome of these interactions may be determined by where in the genome the binding sites for each of these proteins are localized. In the case of Su(Hw) and a subset of dCTCF sites, their localization in intergenic regions suggests that their role may be to form loops that may represent independent functional domains. The rest of dCTCF sites and BEAF sites are located around promoter regions and their function may rely on the same type of interactions to bring these

regions of genes to specific nuclear compartments such as transcription factories.

Since *Drosophila* has several insulators with distinct localization patterns with respect to genomic features, it is interesting to contrast this information with the genome-wide localization of CTCF in vertebrates. The distribution of CTCF in mouse embryonic stem cells and in various human cell lines has been studied using ChIP-chip and ChIP-seq approaches (Barski *et al.*, 2007; Chen *et al.*, 2008; Cuddapah *et al.*, 2009; Jothi *et al.*, 2008; Kim *et al.*, 2007). The number of sites uncovered varies depending on the approach, from ca. 13,800 in IMR90 human fibroblasts using ChIP-chip (Kim *et al.*, 2007) to ca. 39,600 in mouse ES cells using ChIP-seq (Chen *et al.*, 2008). In general, the distribution of CTCF sites on each chromosome correlates with gene density. However, different from transcription factors, CTCF binding sites are generally located an average of 48 kb away from promoters. Approximately 46% of the sites are located in intergenic regions, consistent with the classical role of insulators in establishing chromatin domains, whereas ca. 20% of the sites display promoter proximal localization. As in *Drosophila*, a significant number of the sites fall within genes, with 22% in introns and 12% in exons in human fibroblast cells (Kim *et al.*, 2007). Analyses in resting human $CD4^+$ T cells have found a similar genomic distribution of CTCF sites, with 45% intergenic, 7% 5′UTR, 3% exonic, 29% intronic, 2% 3′UTR, and 13% within 5 kb of the TSS. Although CTCF sites generally correlate with gene density, it is possible to find patterns in their distribution, other than location with respect to genes, suggestive of their various roles in gene expression. For example, CTCF depleted domains, which exhibit lower-than-average CTCF density, tend to include clusters of related gene families and genes that are transcriptionally coregulated. In addition, some CTCF sites are located in intergenic regions at the transition of chromatin domains with different epigenetic status. For example, there are 793 CTCF sites flanking H3K27me3 domains in $CD4^+$ T cells and this occupancy may be cell type specific (Cuddapah *et al.*, 2009). CTCF can also bind 5–10 kb outside LADs, which contain low gene density and low expressing genes, representing a repressive chromatin environment; out of 1344 LADs found, 333 have CTCF binding at least at one side of the border (Guelen *et al.*, 2008). In all three cases, the distribution of CTCF agrees with a classical insulator role for this protein in the establishment or maintenance of functional transcriptional domains. On the other hand, genes located in CTCF enriched domains, which have higher than average CTCF binding, often have multiple alternative promoters (Kim *et al.*, 2007). These and other sites located

close or within genes may play roles in chromosome biology different from those classically assigned to insulators.

The genome occupancy of CTCF sites is similar in different cell types but a significant fraction appears to be cell type specific. For example, 40–60% of CTCF sites are common among $CD4^+$ T, HeLa, and Jurkat cells (Cuddapah *et al.*, 2009) and the rest of the sites occupied by CTCF vary between the different cell types. Similarly, around 30% of CTCF sites are different in human IMR90 fibroblasts and U937 erythroid progenitor cells (Kim *et al.*, 2007). Therefore, as is the case in *Drosophila*, a number of CTCF sites in mammals are cell type specific and may play a role in establishing patterns of gene expression required for cell differentiation.

V. INSULATORS MEDIATE INTRA- AND INTERCHROMOSOMAL INTERACTIONS

All *Drosophila* insulators, with perhaps the exception of *scs*/ZW5 for which data are unavailable, share the BTB domain proteins CP190 and Mod(mdg4), although the specific isoform of Mod(mdg4) appears to be different in different insulators. The BTB domains of these two proteins as well as GAF can interact with each other in various *in vitro* or *in vivo* assays, suggesting that insulator proteins may mediate intra- and interchromosomal interactions among insulator sites throughout the genome. Various types of observations support this conclusion. For example, Su (Hw), Mod(mdg4)2.2, dCTCF, and CP190 show a punctuate distribution in the nuclei of diploid cells. These sites, called insulator bodies, appear to contain multiple individual insulator sequences and their morphology is disrupted by mutations in insulator components (Gerasimova *et al.*, 2000, 2007; Pai *et al.*, 2004). Furthermore, FISH experiments have shown that DNA sequences contained between two insulators form a loop, which becomes two smaller loops when a new insulator is inserted in the middle of the DNA (Byrd and Corces, 2003). Although *Drosophila* insulator bodies are present throughout the nucleus, they seem to localize preferentially in the nuclear periphery. This localization may be mediated by the protein dTopors, which has been shown to interact with Su(Hw) and Mod (mdg4)2.2 as well as with Lamin Dm0 (Capelson and Corces, 2005, 2006). Therefore, dTopors may serve as an anchor to attach insulator sites to the nuclear lamina or nuclear matrix (Capelson and Corces, 2005). In addition to this type of information based on FISH and immunolocalization experiments, results from 3C experiments have shown that scs and scs′ sequences are present in close proximity in the nucleus forming

a loop of the intervening sequence. These contacts may be mediated by CP190 or may be direct, since ZW5 and BEAF 32 have been shown to interact both *in vitro* and *in vivo* in *Drosophila* embryos (Blanton *et al.*, 2003).

The role of insulators in mediating intra- and interchromosomal interactions has been better established in vertebrates using 3C, 4C, and FISH approaches in a number of different systems (Fig. 2). The *β-globin* locus contains several CTCF sites that are conserved in mice and humans. These sites are located within two DNase I hypersensitive regions that flank the *β-globin* locus: the 5′HS5 located in the locus control region (LCR) and

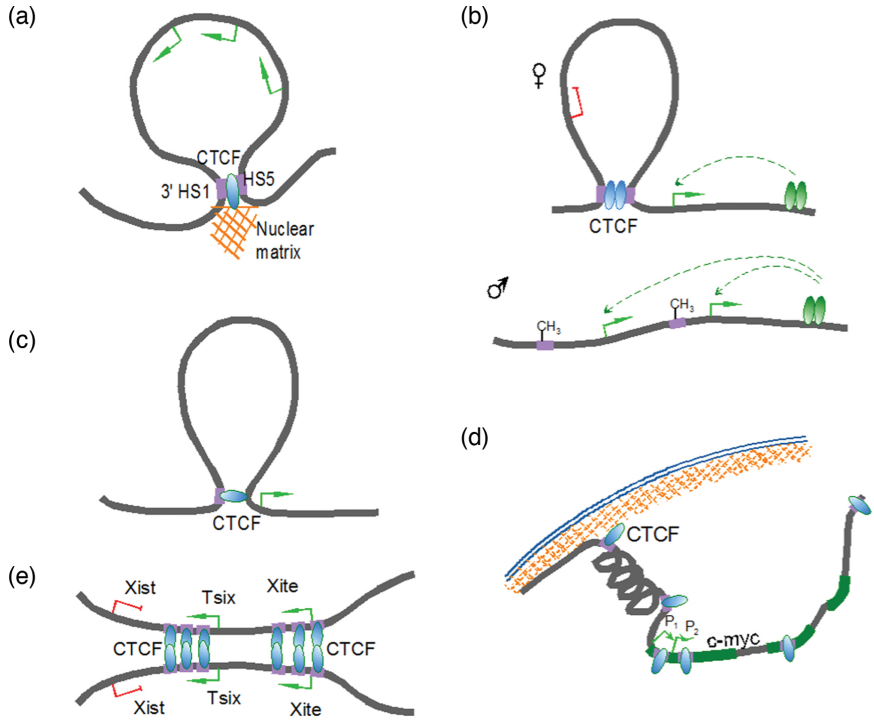

Fig. 2 Mechanisms by which CTCF affects gene expression by mediating intra- and interchromosomal interactions. (A) Anchoring loops via direct attachment to subnuclear structures such as the nucleolus and/or nuclear matrix. (B) Monoallelic gene expression via allele-specific contacts between multiple imprinted regulatory elements. (C) Transcriptional regulation via contacts between intergenic enhancers and promoter-proximal regulatory elements. (D) Global nuclear organization via demarcation of LADs. (E) X-chromosome inactivation or monoallelic gene expression via interchromosomal contacts between regulatory elements in *trans*.

the downstream 3′HS1; additional CTCF sites are located further away at either end of the locus. The 5′HS5 and 3′HS1 sites interact with each other as well as with the more distal CTCF sites flanking the locus, and these interactions are severely weakened in the absence of the CTCF protein. The interactions with 5′HS5 or 3′HS1 are cell type specific but they are not required for proper transcription of the *β-globin* gene. These interactions exist in both erythroid cells, in which the *β-globin* gene is transcribed, and in erythroid progenitor and fibroblast cells in which the gene is not expressed, but not in nonexpressing brain cells (Hou *et al.*, 2010; Splinter *et al.*, 2006; Tolhuis *et al.*, 2002). Mutation of 3′HS1 destabilizes the long-range interactions among these CTCF sites but mutation or deletion of 3′HS1 or 5′HS5 can neither affect the expression kinetics nor levels of the *β-globin* genes in erythroid cells (Farrell *et al.*, 2000; Splinter *et al.*, 2006). Nevertheless, a number of CTCF sites present outside of the *β-globin* locus interact with each other but not with 5′HS5 or 3′HS1, and some of these interactions are both cell type specific and transcription relevant. These interactions can only be detected in either erythroid cells or nonerythroid cells. Knockdown of CTCF, which leads to a global reduction of both groups of interactions, can negatively affect *β-globin* gene transcription (Hou *et al.*, 2010). It is possible that these interactions are required for *β-globin* expression, although it is also possible that the observed changes in transcription are due to indirect effects of CTCF knockdown. Therefore, it appears that CTCF-mediated interactions involving 5′HS5, 3′HS1, and other CTCF sites create a three-dimensional organization of the *β-globin* locus before the commitment to the erythroid lineage that is not necessary or sufficient for the activation of the globin genes. The fact that CTCF-mediated interactions among these sites are not observed in brain cells suggests that this organization may play an earlier role during differentiation after the commitment to ectodermal or mesodermal fates. New interactions established later during erythroid differentiation and involving a different set of more distally located CTCF sites may be involved in the establishment of a global architecture that is actually more directly responsible for the expression of the *β-globin* locus.

The requirement of CTCF-mediated interactions for gene expression has been studied in several other loci. At the cytokine interferon-*γ* (*IFNG*) locus there are three conserved CTCF binding sites located upstream, downstream and within the gene 1.5 kb from the TSS in human and mouse cells. The CTCF sites display cell type-specific interactions in specialized T helper 1 (Th1) cells and these interactions parallel differences in CTCF occupancy (Hadjur *et al.*, 2009; Sekimata *et al.*, 2009). These CTCF sites also interact with the enhancers present in the locus, although

the enhancers do not have CTCF binding sites. The three-dimensional conformation of the locus and robust *Ifng* expression in Th1 cells are dependent on the presence of CTCF (Sekimata *et al.*, 2009). It is possible that interactions among the CTCF sites help recruit the enhancers and activate gene expression at the locus. The apolipoprotein (APO) gene (APOA1/C3/A4/A5) cluster in humans is another interesting example of CTCF-mediated loops that are required for proper gene expression. Results from 3C experiments indicate that CTCF sites in this region interact with each other in Hep3B cells. These interactions lead to the formation of two transcribed loops with APOC3/A4/A5 in one loop and APOA1 in the other. Consistent with the model based on 3C data, CTCF knockdown leads to a decrease of APOC3/A4/A5 but an increase of APOA1 expression (Mishiro *et al.*, 2009). Major histocompatibility class II (MHC-II) genes are also regulated by intrachromosome interactions mediated by CTCF. The XL9 element contains CTCF sites and is located in the intergenic region between the MHC-II genes HLA-DRB1 and HLA-DQA1. Knockdown of CTCF diminishes expression of these genes. 3C experiments detect an interaction between XL9 and proximal promoter elements of these two MHC-II genes. These interactions also depend on the class II transactivator (CIITA), which can associate with CTCF (Majumder *et al.*, 2008). The chromatin conformation mediated by the association between CIITA and CTCF is required for transcription of the MHC-II genes. In addition to its role in processes that result in increased transcription of genes, CTCF may also mediate chromatin high order structure that results in gene silencing. For example, 3C experiments suggest that silenced Hox genes in human cells are spatially clustered via a specific three-dimensional architecture of the locus; CTCF is present at the sites of contact and appears to be responsible for the establishment and/or maintenance of this architecture (Ferraiuolo *et al.*, 2010).

The three-dimensional arrangement of the chromatin fiber created by CTCF-mediated interactions also plays an important role in imprinted gene expression at the *H19/Igf2* locus. The ICR immediately upstream of *H19* contains CTCF sites that are crucial for the imprinted expression of these two genes. Results from 3C experiments demonstrate that these CTCF sites can mediate allele-specific chromosome interactions that seem to control the accessibility of the *Igf2* promoter to the shared enhancer. On the maternal allele of both mouse and human cells the ICR is not methylated and CTCF binds to DNA. In the mouse, the ICR interacts with CTCF sites present in the upstream DMR1 and downstream MAR3 sites that flank the *Igf2* gene. The three-dimensional structure formed as a consequence of these interactions keeps *Igf2* in an enclosed domain (Kurukuti *et al.*, 2006; Murrell *et al.*, 2004). In human cells, the ICR

interacts with a CTCF Downstream Site (CTCF DS) located downstream of the shared enhancer; this interaction creates a loop that encloses the enhancer (Nativio *et al.*, 2009). In both mouse and human maternal chromosomes, the enhancer is unable to interact with the *Igf2* promoter. In mouse, the 142^{*} allele has a mutated CTCF binding site in the ICR, and in mice that inherit this allele from their mother, the ICR interacts with DMR2 instead of DMR1 and MAR3; this is accompanied by biallelic expression of *Igf2*. On the other hand, there is no effect if 142^{*} is paternally inherited (Kurukuti *et al.*, 2006; Pant *et al.*, 2003). On the paternal chromosome, the ICR is methylated and CTCF does not bind. In mouse cells the ICR now interacts with the DMR2 site located downstream of *Igf2* and allows the enhancer to interact with the *Igf2* promoter (Kurukuti *et al.*, 2006; Murrell *et al.*, 2004) whereas in human cells the ICR cannot interact with the CTCF DS (Nativio *et al.*, 2009). These results suggest that imprinted expression of the *H19/Igf2* locus is dependent on a specific three-dimensional organization of the region mediated by interactions between CTCF sites. Other studies suggest that CTCF-mediated interactions are not limited to the locus but also extend to CTCF sites located in other chromosomes. Using 3C or 4C, the *H19* ICR on chromosome 7 has been shown to interact with genes on different chromosomes, such as *Wsb1/Nf1* on chromosome 11, *Abcg2* on chromosome 6, and *Osbpl1a* on chromosome 18 (Ling *et al.*, 2006; Zhao *et al.*, 2006). The majority of these interchromosomal interactions take place primarily with the maternally inherited *H19* ICR. Knockdown of CTCF or mutation of CTCF sites on the *H19* ICR in the maternal chromosome abrogates these associations, suggesting that these interactions are CTCF dependent (Ling *et al.*, 2006; Zhao *et al.*, 2006). Interestingly, imprinted loci are strongly overrepresented among the regions involved in interchromosome interactions with the *H19* ICR. The clustering of these imprinting genes by interchromosome interactions, which is termed ''imprinting interactome,'' may facilitate the regulation of these genes in *trans*. When the CTCF site is mutated in the maternal *H19* ICR, the interactions are abrogated and the transcription of these imprinted genes is also altered; the paternal expression of the *Wsb1/Nf1* gene is reduced by 50-fold, expression of *Impact* is reduced by 50% and transcription of *Osbpl1a* is increased by 2.5-fold (Ling *et al.*, 2006; Zhao *et al.*, 2006). In addition, the asynchronous replication timing of these genes in spermatogonia is switched from late to early (Sandhu *et al.*, 2009). Therefore, CTCF-mediated interchromosomal interactions are important for the nonallelic regulation of the epigenetic status of multiple genes in *trans*. One important advantage for such regulation is that the interactions can be reprogrammed during germline development when the epigenetic states of imprinted domains are

reprogrammed. In the testis, when the maternal allele is turned to the paternal mode, the interactome can be observed in spermatogonia, but not in spermatocytes and round spermatids where the reprogramming is complete (Sandhu *et al.*, 2009).

Another interesting example of CTCF-mediated interchromosome interactions takes place during X-chromosome pairing, which is important for X inactivation. To achieve the mutually exclusive designation of active X (Xa) or inactive X (Xi), it is necessary for the two X chromosomes to communicate in *trans* through homologous pairing. Pairing depends on a 15-kb region within the *Tsix* and *Xite* loci. ChIP experiments show that CTCF binds both *Tsix* and *Xite* elements in female embryonic stem cells. CTCF knockdown reduces the frequency of X–X pairing to background levels in wild-type embryonic stem cells and embryoid bodies (Xu *et al.*, 2007). It is not clear whether CTCF mediates the pairing directly or by recruiting other factors; Oct4, which can associate with CTCF, is also needed for this pairing (Donohoe *et al.*, 2009).

It is possible that CTCF-mediated interactions have a broader role than just regulation of transcription. One interesting example of the diverse roles that CTCF may play in nuclear biology is that of V(D)J recombination, a process that is highly regulated during B-cell development. IgH rearrangement in pro-B cells begins with D_H to J_H rearrangement followed by rearrangement of a V_H gene segment to D_HJ_H. The complexity of the mechanisms controlling V_H to D_HJ_H rearrangement is in part due to the organization of the genes involved in the process. In mice, there are more than 100 V_H genes spanning a 2.5 Mb region; the 96 V_κ genes cover 3.1 Mb whereas the J genes occupy a 2 kb region. How do all the V genes access the small J cluster in the *Igh* and *Igκ* loci? It has been recently shown that CTCF is present across the *Igh*, *Igκ*, and *Igλ* loci and plays a critical role in bringing these different loci together in the nucleus and that the cohesin subunit Rad21, which interacts with CTCF in a cell lineage-specific manner, plays a regulatory role in the process (Degner *et al.*, 2009). Although CTCF does not display obvious cell lineage-specific binding, as does Rad21, it is indispensable for the cell-lineage-specific contraction of the *Igh* locus. Preliminary results using 3D-FISH indicate that *Igh* locus contraction is decreased in pro-B cells treated with CTCF shRNA (Degner-Leisso and Feeney, 2010).

Taken together, these results suggest that CTCF can mediate long-range intra- and interchromosome interactions at different loci throughout the genome either by interactions with other CTCF sites or by association with other factors. The result of these interactions is a specific three-dimensional arrangement of the chromatin that can have different effects on

chromosome biology, depending on the nature of the sequences brought together by CTCF.

VI. MECHANISMS OF REGULATION OF INSULATOR FUNCTION

If insulators mediate inter- or intrachromosomal interactions that result in the formation of chromatin loops, which in turn may be attached to the nuclear matrix, it is possible that the resulting structures determine a particular pattern of nuclear organization that may be important for gene expression. For example, it is possible that as cells differentiate, insulator-mediated changes in nuclear organization precede or accompany cell differentiation and may be crucial in the establishment and/or maintenance of specific patterns of gene expression. If this is the case, cells must possess mechanisms to regulate insulator activity in order to establish distinct nuclear architectures that are cell fate specific.

Evidence for the existence of mechanisms to control insulator function comes in part from genome-wide studies of insulator protein localization in *Drosophila* cell lines of different tissue origin. Studies in Kc cells, which have a neural origin, indicate that there are 3747 Su(Hw), 2266 dCTCF, 2995 BEAF, and 5272 CP190 sites where these proteins are present in the genome. Of these, 47% of Su(Hw), 62% of dCTCF, and 71% of BEAF sites colocalize with CP190 sites (Bushey *et al.*, 2009). Since CP190 is required for insulator function, this observation suggests that cells may control the activity of these various insulators by regulating the recruitment of CP190. In addition, comparison of the genomic location of different insulator proteins in Kc and Mbn2 cells (a hematopoietic cell line), reveals that while many sites are constant, a fraction of the localization sites for each of the four insulator proteins is different between the two lines. For example, 18% of Su(Hw) sites in Kc cells and 5% of Su(Hw) sites in Mbn2 cells are cell type specific. This is also the case for dCTCF, for which 18% of sites in Kc cells and 37% in Mbn2 cells are cell unique, whereas the number of cell type-specific BEAF sites is 11% in Kc cells and 11% in Mbn2 cells. In the case of CP190, which is found at all three insulator subclasses, 17% of sites present in Kc cells and 14% in Mbn2 cells were found to be cell type specific (Bushey *et al.*, 2009). These results suggest that cells may regulate insulator activity by controlling the recruitment of the DNA binding proteins to their target sites in the genome in addition to controlling the recruitment of CP190.

Several proteins have been characterized in *Drosophila* that may play a role in regulating insulator function (Fig. 3). dTopors, in addition to serving as an attachment point for insulators to the nuclear lamina, it has E3 ubiquitin ligase activity. This activity is required for proper insulator function. Its substrate has not been clearly identified but Su(Hw) is a likely candidate, since overexpression of dTopors enzymatic activity reverses the effect of *mod(mdg4)* mutations on the ability of Su(Hw) to interact with chromatin (Capelson and Corces, 2005). In addition, modification of Mod(mdg4)2.2 and CP190 by sumoylation inhibits insulator function. Disruption of the SUMO conjugation pathway improves the enhancer-blocking function of a partially active insulator, indicating that SUMO modification acts to negatively regulate the activity of the *gypsy* insulator. Sumoylation does not affect the ability of CP190 or Mod(mdg4) 2.2 to bind chromatin (Capelson and Corces, 2006). Interestingly, dTopors inhibits sumoylation of Mod(mdg4)2.2 and CP190. Therefore, this protein may have a double effect on insulator function by ubiquitinating some insulator components and inhibiting the sumoylation of others. A second candidate protein with a possible role in regulating insulator function is the Rm62 RNA helicase. Insulator activity decreases in the presence of mutations in components of the RNAi machinery; insulator function is restored by mutations in Rm62. These observations have led to a model suggesting that insulator bodies contain RNA whose synthesis requires RNAi proteins. Rm62 may interact with this RNA to decrease insulator function (Lei and Corces, 2006).

dTopors and Rm62 have been only shown to affect the function of the *gypsy*/Su(Hw) insulator but their potential role in regulating the activity of other insulator subclasses has not been tested. There are currently no characterized mechanisms that control the activity of the dCTCF insulator. O-Glycosylation of BEAF can be detected in *Drosophila* embryonic cells in a domain of the protein that is required for association with the nuclear matrix; however, it is not clear whether glycosylation is required for *scs′* insulator function (Pathak *et al.*, 2007). A second possible candidate to regulate the BEAF insulator is the DREF protein. DREF has been characterized as a transcription factor that shares binding sites with BEAF 32. It is possible that DREF regulates BEAF binding through competition for the same DNA sequences (Hart *et al.*, 1999).

In vertebrates, the CTCF insulator can also be regulated at different levels. As in *Drosophila*, one strategy is to control the interaction between CTCF and its DNA target sequence. The best understood mechanism to control CTCF occupancy is through DNA methylation on CpG dinucleotides within and around the CTCF binding site. This has been well studied in the *H19/Igf2* locus, where CTCF only binds to the unmethylated

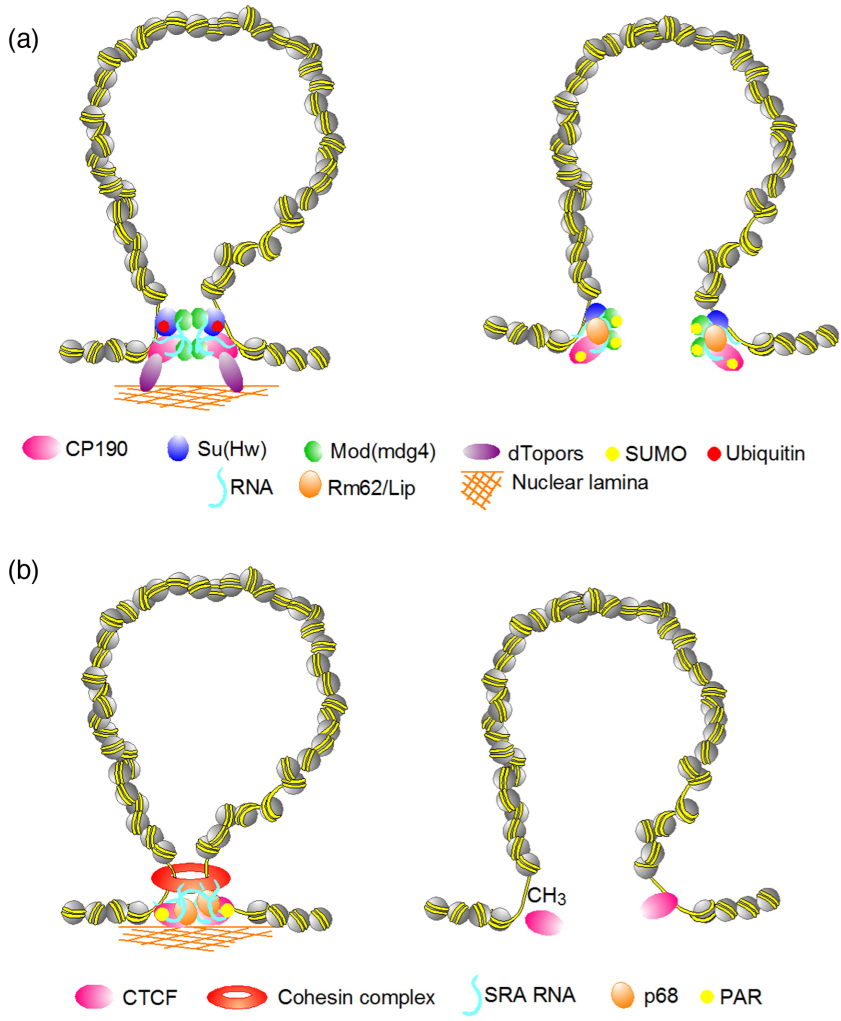

Fig. 3 Similarities in the regulatory mechanisms of insulator function in *Drosophila* and vertebrates. (A) Regulation of insulator activity in *Drosophila*. The left panel shows two active insulators coming together to make a functional loop and the right panel displays two inactive insulators unable to form a loop. At an active insulator site, dTopors is present, Rm62/Lip is not present, Su(Hw) is ubiquitinated, Mod(mdg4)2.2 and CP190 are not sumoylated and dTopors serves as a bridge to the nuclear lamina/matrix. At inactive insulator sites, dTopors is absent and Su(Hw) is not ubiquitinated, whereas Mod(mdg4) 2.2 and CP190 are sumoylated. Rm62/Lip is present and bound to RNA. Under these conditions, the two insulator sites cannot interact to form a loop. Absence of dTopors also precludes interactions with the nuclear lamina. (B) Regulation of insulator activity in vertebrates. The left panel shows two active CTCF insulators coming together to make a

maternal allele but does not bind to the methylated paternal allele (Bell and Felsenfeld, 2000; Hark *et al.*, 2000). CTCF binding to methylation-free sites is also documented for other genes, including both imprinted and nonimprinted loci, such as Rasgrf1, KvDMR1, GRB10, INK/ARF, and DM1 (Filippova *et al.*, 2001; Fitzpatrick *et al.*, 2007; Hikichi *et al.*, 2003; Rodriguez *et al.*, 2010; Yoon *et al.*, 2005). Therefore, regulation of CTCF binding via DNA methylation may be a general strategy to control the function of this insulator. In addition, CTCF function can be regulated by posttranslational modification. CTCF can be phosphorylated at the C-terminus by the protein kinase CK2. Mutation of the phosphorylated residue in CTCF does not alter its nuclear localization or DNA binding *in vitro*, but enhances repression of the *c-myc* promoters. A phosphomimetic mutant behaves in the opposite fashion (El-Kady and Klenova, 2005; Klenova *et al.*, 2001). Thus, phosphorylation may affect the interaction of CTCF with other proteins and modulate its activity.

CTCF can also be covalently modified by poly(ADP-ribosyl)ation. CTCF interacts with poly(ADP-ribose)polymerase (PARP)-1 both *in vivo* and *in vitro* (Guastafierro *et al.*, 2008). Both proteins colocalize at the *Igf2/H19* ICR and overlap genome wide at more than 140 CTCF target sites in the mouse genome, including both imprinted and nonimprinted loci (Yu *et al.*, 2004). Poly(ADP-ribosyl)ation does not affect the ability of CTCF to bind DNA but it is required for its insulator function. For example, mutation of the poly(ADP-ribosyl)ation site in CTCF or treatment with the PARP inhibitor ABA, can compromise imprinting of the *H19/Igf2* locus, resulting in biallelic expression of *Igf2*, but CTCF binding to the *H19/Igf2* ICR is not affected (Farrar *et al.*, 2010; Yu *et al.*, 2004). In addition to poly(ADP-ribosyl)ation of CTCF, PARP-1 can also modify itself. The modified PARP-1 can inhibit DNMT1 activity and control DNA methylation patterns (Guastafierro *et al.*, 2008). This may in turn control the binding of CTCF. Thus, poly(ADP-ribosyl)ation may affect the function of CTCF by modulating its interaction with DNA and with other proteins.

Another mechanism to regulate insulator function that is shared by *Drosophila* and vertebrates is the use of homologous proteins that can

Fig. 3 (*cont.*) functional loop and the right panel displays two inactive CTCF insulators unable to form a loop. At an active insulator site, CTCF is present and poly(ADP-ribosyl)ated and, directly or indirectly, interacts with the nuclear matrix/lamina; in addition, the SRA RNA and p68 are present while the cohesin complex brings together the two DNA molecules forming the base of the loop. At inactive insulator sites, CTCF is either not bound because of the DNA is methylated or it is bound but not modified by poly(ADP-ribosyl)ation; cohesin, SRA RNA, and/or p68 are absent. Under these conditions the two insulator sites cannot interact to form a loop.

compete for DNA binding. Like DREF and BEAF in *Drosophila*, there is also a homologue of CTCF in vertebrates named CTCFL or BORIS. These two proteins share homology of the central 12 zinc finger DNA binding domain and they can recognize the same DNA binding sequences (Loukinov *et al.*, 2002). BORIS is normally present only in the testis during germ-line development where it can stimulate the histone-methyltransferase activity of PRMT7 and contribute to the efficient DNA methylation of BORIS/CTCF sites in male germ line (Jelinic *et al.*, 2006). Therefore, BORIS can play a role in writing the epigenetic marks that will be read by CTCF. Usually, BORIS is expressed in a mutually exclusive manner with CTCF and only influences CTCF function by its effects in the germ line (Loukinov *et al.*, 2002). However, BORIS can be aberrantly expressed under particular conditions; for example, reduction of CTCF in normally BORIS-negative human fibroblasts results in derepression of BORIS expression (Renaud *et al.*, 2007). Abnormally expressed BORIS may affect the activity of CTCF by various mechanisms. First, BORIS can compete with CTCF for binding sites and lead to alterations in gene expression; although BORIS shares the zinc finger domain with CTCF, the two proteins are different in the N- and C-terminal regions that constitute approximately two-thirds of the full-length amino acid sequences of these proteins and could recruit different functional partners. For example, the transcription factor Sp1 can interact with BORIS but not CTCF (Hong *et al.*, 2005; Kang *et al.*, 2007). Second, the two proteins differ in their ability to bind methylated DNA. In contrast to CTCF, the binding of BORIS to its target site is methylation-independent *in vitro* and methylation-preferential *in vivo* at the *H19/Igf2* ICR (Nguyen *et al.*, 2008). As a consequence, BORIS could bind to methylated CTCF/BORIS sites, where CTCF is not normally present, and affect the expression of adjacent genes. Third, aberrantly expressed BORIS together with ubiquitously present PRMT7 might change the methylation status of previously nonmethylated sites and evict CTCF. The binding of BORIS and the different transcription factors recruited to genes harboring the CTCF/BORIS sites could lead to totally different expression patterns from those observed when only CTCF is present in the cell (Hong *et al.*, 2005). Not surprisingly then, aberrant expression of BORIS can lead to disease and cancer as we discuss below.

Protein partners of CTCF may be also indispensable for the establishment and/or maintenance of chromosome interactions mediated by this protein and it is therefore plausible that regulation of these factors could change CTCF activity (Fig. 3). Knockdown of CHD8 has been shown to impair imprinting of *Igf2* (Ishihara *et al.*, 2006). Depletion of cohesin components without affecting CTCF expression can also lead to

disruption of chromatin interactions and changes in the expression of genes under CTCF control (Hadjur *et al.*, 2009; Hou *et al.*, 2010; Mishiro *et al.*, 2009; Nativio *et al.*, 2009). In the *Igh* locus, although the binding of CTCF is not altered during B-cell differentiation, the recruitment of cohesin is cell type specific and appears to be responsible for the hypothesized role of CTCF in V(D)J recombination (Degner *et al.*, 2009). It is therefore possible that control of the expression or posttranslational modification of CTCF partners will regulate the different cellular activities of CTCF. The complexity of mechanisms by which CTCF activity can be regulated is showcased by a recent report showing that RNA helicase p68 (DDX5) and its associated noncoding RNA, steroid receptor RNA activator (SRA), bind to CTCF, and are both essential *in vivo* for insulator function at the *H19/Igf2* ICR. Contrary to the inhibitory role of the RNA helicase Rm62 on insulator function in *Drosophila*, p68 is needed to positively regulate insulator function in vertebrate cells. Knockdown of p68 decreases the insulator activity of the *H19/Igf2* ICR and increases the expression of *Igf2* as previously observed in CTCF knockdown cells. In *Drosophila* cells, Rm62 interacts with CP190, while in vertebrates p68 is required for the recruitment of cohesin (Yao *et al.*, 2010). In spite of the opposite effect of the two helicases on insulator function, the conservation of this mechanism between insects and vertebrates is striking.

VII. INSULATORS, CANCER, AND DISEASE

CTCF affects the expression of genes such as *hTERT* or *myc* that regulate cell cycle processes that are important for cell growth, differentiation, and apoptosis. An appropriate balance between these processes is essential for normal development, whereas an imbalance can lead to tumor development. *hTERT* is the human telomerase reverse transcriptase, a catalytic subunit and limiting factor for telomerase activity, which is required during the cell cycle. CTCF binds to sequences of the *hTERT* gene located in the first two exons and represses its expression. CTCF only binds to *hTERT* in cells where the gene is not transcribed but not in telomerase-expressing cells. Knockdown of CTCF derepresses *hTERT* gene expression in normal telomerase-negative cells (Renaud *et al.*, 2005). *hTERT* is a general cancer susceptibility locus, and its abnormal expression may disturb appropriate control of the cell cycle and induce aberrant cell growth (Johnatty *et al.*, 2010). CTCF also binds constitutively to two regulatory elements close to the *c-myc* P2 promoter, the P2 upstream *c-myc* insulator *e*lement (MINE or CTCF-N) and the CTCF-A

site located immediately downstream of the P2 promoter (Filippova *et al.*, 1996; Gombert and Krumm, 2009; Gombert *et al.*, 2003). Binding of CTCF can affect the expression of *c-myc*, although the molecular mechanism underlying this effect is not well understood. MINE has been shown to have CTCF-dependent enhancer-blocking activity and CTCF knockdown leads to an increase in *c-myc* protein levels (Gombert and Krumm, 2009; Gombert *et al.*, 2003; Torrano *et al.*, 2005). The c-myc protein is important for cells during the transition from proliferation to differentiation. Ectopic expression of CTCF in K562 cells leads to growth retardation and promotion of differentiation associated with increased *c-myc* expression, whereas CTCF knockdown significantly inhibits differentiation with reduced *c-myc* transcription (Torrano *et al.*, 2005). Thus, the loss of CTCF function may disturb the balance between cell growth, differentiation, and apoptosis due to its effects on the expression of cell growth regulators (Heath *et al.*, 2008). This observation has been extended to other genes whose expression depends on CTCF such as the tumor repressor retinoblastoma (Rb) gene. Mutations and deletions of the Rb gene have been associated with a number of inherited malignancies. CTCF binds *in vitro* and *in vivo* to the human Rb promoter and this binding is required for Rb expression. When the CTCF binding site is removed or mutated, reporter gene expression decreases (De La Rosa-Velázquez *et al.*, 2007). Some other genes encoding regulators of the cell cycle, such as *p19(ARF)*, *p16(INK4a)*, *PLK*, *BRCA1*, *p53*, and *p27* are also growth suppressors frequently silenced in cancer whose expression is also controlled by CTCF (Filippova, 2007). For example, *BRCA1* is expressed in normal cells but it is silenced in some cancer cells. CTCF binds at the promoter region of *BRCA1* only in expressing cells but not in tumor cells in which *BRCA1* is silenced (Darci *et al.*, 2004; Xu *et al.*, 2010). The *INK4B–ARF–INK4A* (*INK/ARF*) locus contains three tumor suppressor genes that are kept silenced by DNA methylation in different types of cancer. The *p16(INK4a)* tumor suppressor gene is a frequent target of epigenetic inactivation in cancers such as breast, lung, colorectal, and multiple myeloma. CTCF binds upstream of the *p16(INK4a)* promoter and the absence of binding is associated with silencing of *p16 (INK4a)* expression in breast cancer and multiple myeloma cells. Moreover, ablation of CTCF protein function from *p16(INK4a)*-expressing cells by shRNA results in epigenetic changes in the *p16(INK4a)* promoter and loss of transcription (Rodriguez *et al.*, 2010; Witcher and Emerson, 2009). Conditional overexpression of CTCF in B cells enhances expression of *p27*, *p21*, *p53*, and *p19(ARF)*, followed by inhibition of cell growth and induction of apoptosis, while knockdown of CTCF results in inhibition of these genes (Qi *et al.*, 2003). These results suggest that loss of

CTCF function can lead to the silencing of growth suppressor genes and contribute to carcinogenesis. Whether these effects are due to changes in the three-dimensional organization of the DNA or more local effects on transcription is unclear at this time.

CTCF also plays a role in trinucleotide repeat-associated diseases. The affected genes in myotonic dystrophy (DM1), spinocerebellar ataxia 7 (SCA7), SCA2, dentatorubral-pallidoluysian atrophy (DRPLA), and Huntington's (HD) contain CTCF sites on one or both sides of the CTG/CAG repeat-containing region (Filippova *et al.*, 2001). It has been reported that CTCF can control noncoding transcription of the repeats. For example, at DM1, the two CTCF sites flanking the CTG repeat can function as an insulator to restrict the antisense transcription and constrain the heterochromatin state to the repeats without affecting the nearby DMPK and SIX5 genes, whose silencing contributes to myotonic myopathy, cataracts, and cardiac-conduction defects in the disease (Cho *et al.*, 2005; Filippova *et al.*, 2001). In contrast, cells from affected individuals show a loss of CTCF binding, which is associated with expansion of the repeats, spread of heterochromatin, and regional CpG methylation (Cho *et al.*, 2005). Therefore, CTCF can *cis*-regulate the stability of the trinucleotide repeats. In transgenic mice carrying a SCA7 genomic fragment with a CTCF binding site, mutation or methylation of the DNA sequence at the binding site can promote repeat expansion (Libby *et al.*, 2008). However, the mechanism by which CTCF affects the stability of the repeats is not clear. A recent report suggests that CTCF can contribute to repeat stability through regulation of DNA replication. CTCF sites located between a replication origin and the repeats can slow or pause the progression of the replication machinery and enable safe passage of the replication fork. Without CTCF, the procession of the replication fork allows for slippage, hairpin formation, fork reversal, and other replication errors (Cleary *et al.*, 2010). Thus, loss of CTCF flanking the repeats may affect repeat instability. Aberrant methylation of CTCF binding sites may also lead to eviction of the protein, repeat expansion, and abnormal transcription of the locus.

CTCF can regulate the expression and epigenetic features of imprinted genes both in *cis* and in *trans* through intra- or interchromosomal interactions as discussed above. Almost all imprinted genes identified to date can be classified as regulators of embryonic growth, placental growth or adult metabolism (Jelinic and Shaw, 2007). It is then not surprising that loss of imprinting (LOI) can lead to various human cancers and disease. In fact, the earliest and most common alteration observed in human cancers is LOI, which has been documented in 100% of chronic myeloid leukemia, 80% of ovarian tumors, 70% of Wilm's tumors, 66% of colorectal cancer,

56% of Barrett's esophagus, 50% of renal-cell carcinomas, 50% of esophageal cancer, 47–85% of lung adenocarcinoma, and 30% of meningiomas. Many human diseases are also associated with altered expression of imprinted genes, including Beckwith–Wiedemann (BWS), Prader–Willi, Angelman, Silver–Russell, immunodeficiency syndrome (ICF), Rett, Albright hereditary osteodystrophy, and hydatidiform mole (Jelinic and Shaw, 2007). Although the role of CTCF on imprinting has only been studied in detail in the *H19/Igf2* locus, similar CTCF-associated mechanisms appear to also play a role in imprinting at the *Rasgrf1*, *DLK1/GTL2*, *Wsb1/Nf1*, and *KvDMR* loci (Fitzpatrick *et al.*, 2007; Ling *et al.*, 2006; Yoon *et al.*, 2005). This suggests a widespread function of CTCF in the regulation of genomic imprinting and, as a consequence, the development of imprinting-associated human cancer and other diseases.

Alterations of CTCF function that lead to disease development can have both genetic and epigenetic origins. The CTCF gene maps to the cancer-associated human chromosome locus 16q22.1, which is the smallest overlap region of a variety of deletions found in breast, prostate, ovarian, and Wilm's tumors. Since CTCF null mutations are lethal, most CTCF mutations involved in human disease appear to cluster in the zinc finger domain of the protein. Instead of causing a complete loss of function, mutations in this domain alter the binding ability of CTCF. For example, one of the mutations identified abrogates CTCF binding to a subset of target sites in certain genes involved in the regulation of cell proliferation (c-myc, ARF, PIM1, PLK, and Igf2) but does not alter binding to other regions of the genome, including the β-globin insulator, the lysozyme silencer, or the APP promoter (Filippova *et al.*, 2002). The selective loss of CTCF function caused by mutations in the zinc finger region may be tolerated because it does not affect cell viability, but results instead in transformation to a malignant phenotype. Since only selective changes of CTCF function can be tolerated in cells, it is not surprising that CTCF mutations are infrequently discovered in cancer or disease states. Instead, it is possible that mutations leading to changes of CTCF function could affect the regulatory roles of this protein rather than its structure. These alterations could be epigenetic, gene specific, and relatively tolerable. Most CTCF-mediated cancers may occur through aberrant methylation of CTCF binding sites, which results in loss of CTCF binding to the DNA. For example, abnormal methylation patterns of CTCF targets involved in *H19/Igf2* imprinting control have been shown to be associated with colorectal, Wilm's, and bladder tumors (Klenova *et al.*, 2002). The methylation pattern at the KvDMR locus was found to be altered in a child with BWS (Fitzpatrick *et al.*, 2007). This alteration of methylation at CTCF sites has also been observed at nonimprinted genes. For example, in mice

lung tumors, CpG sites harboring CTCF binding sequences are hypermethylated at the *INK/ARF* locus, resulting in the absence of CTCF binding and reduced expression of the tumor suppressor genes at this locus (Rodriguez *et al.*, 2010; Tam *et al.*, 2003). In some human tumors, the promoter of the *BRCA1* gene is methylated, CTCF is evicted, and *BRCA1* is silenced (Darci *et al.*, 2004; Xu *et al.*, 2010).

In addition to DNA methylation, other regulatory mechanisms of CTCF function can also be altered and lead to the occurrence of cancer and disease. For example, testicular and ovarian tumor cells exhibit the same methylation profile as normal cells, but the BORIS protein, which recognizes the same DNA binding sites as CTCF, is expressed abnormally. Both CTCF and BORIS bind to DNA in cancer cells but BORIS inhibits the function of CTCF and leads to expression of *hTERT* (Renaud et al., 2010). Telomerase activity is not detectable in most somatic cells of adult humans but is found in highly proliferative cells, such as germ cells and stem cells, and 85–95% of cancers. The expression of BORIS in normal cells is sufficient to allow *hTERT* transcription and to extend their lifespan *in vitro* (Renaud *et al.*, 2010). Reciprocal binding of CTCF and BORIS has also been observed at the NY-ESO-1 promoter and leads to derepression of this cancer-testis gene in lung tumors (Hong *et al.*, 2005). BORIS is also aberrantly expressed in 71% (41 of 58 cases) of breast tumors. High levels of BORIS correlate with high levels of progesterone receptor (PR) and estrogen receptor (ER). The link between BORIS and PR/ER was further confirmed by the ability of BORIS to activate the promoters of the *PR* and *ER* genes in reporter gene assays (D'Arcy *et al.*, 2008). Defects in other regulatory mechanisms that alter CTCF function can also lead to cancer. For example, alterations in the poly(ADP-ribosyl)ation pathway results in the absence of CTCF PARlation and reduction of *p16(INK4a)* and *Rassf1A* expression (Witcher and Emerson, 2009).

Understanding of the different mechanisms that alter CTCF function has opened new possibilities in the design of treatments for cancer or other diseases. For example, trichostatin A (TSA), an inhibitor of histone deacetylase activity, is a well-known antitumor agent that effectively and selectively induces arrest of tumor growth and apoptosis. hTERT appears to be one of the primary targets for TSA-induced apoptosis in cancer cells. TSA induces demethylation of CpGs present in the binding site of CTCF on the hTERT promoter, leading to repression of hTERT (Choi *et al.*, 2010). CTCF can contribute to the regulation of a variety of genes whose proper expression is required for normal cell differentiation. Both genetic and epigenetic changes of CTCF function can lead to the miss-expression of these genes, resulting in the development of a malignant phenotype.

VIII. FUTURE QUESTIONS

Chromatin insulators are important regulatory sequences present in the genome of most eukaryotes. Although they are defined experimentally based on their ability to affect enhancer–promoter interactions and interfere with the spreading of repressive signals from heterochromatin, their role appears to be more general. Intra- and interchromosomal interactions mediated by insulator proteins may establish a web of contacts between individual insulator sites that give rise to specific patterns of nuclear organization. Insulator-mediated nuclear structures may be regulatable by controlling the interactions between insulator DNA binding proteins and their cognate sequences. In addition, recruitment of insulator components involved in mediating interinsulator interactions may represent a second level of regulation of insulator function. These two levels of control may be the result of specific covalent modifications of insulator proteins. The specific outcome of interinsulator interactions is a consequence of the location of the particular insulator sequences with respect to specific genome features, and interference with enhancer–promoter interactions may be just one of these outcomes. It is unclear at this time whether the three-dimensional pattern of nuclear organization created by the insulator interactome is different in various cell types and whether this organization carries meaningful epigenetic information. Most studies to date have concentrated on the analysis of alterations of insulator function at the local level. These studies suggest that aberrant insulator function can lead to alterations in gene expression and, when the affected genes are involved in cell growth processes, to cancer. Nevertheless, it is possible that changes in insulator function cause more general effects on transcription. Understanding the role of nuclear organization in gene expression and cell differentiation remains the main issue for future investigation.

ACKNOWLEDGMENTS

We would like to thank Chunhui Hou for inspirational discussions. Work in the authors' laboratory is funded by National Institutes of Health Award GM35463.

REFERENCES

Barski, A., and Cuddapah, S., *et al.*, (2007). High-resolution profiling of histone methylations in the human genome. *Cell* **129**(4), 823–837.

Bartkuhn, M., and Straub, T., *et al*., (2009). Active promoters and insulators are marked by the centrosomal protein 190. *EMBO J.* **28**(7), 877–888.

Bell, A. C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**(6785), 482–485.

Belozerov, V. E., and Majumder, P., *et al*., (2003). A novel boundary element may facilitate independent gene regulation in the Antennapedia complex of *Drosophila*. *EMBO J.* **22** (12), 3113–3121.

Blanton, J., and Gaszner, M., *et al*., (2003). Protein:protein interactions and the pairing of boundary elements in vivo. *Genes Dev.* **17**(5), 664–675.

Buchner, K., and Roth, P., *et al*., (2000). Genetic and molecular complexity of the position effect variegation modifier mod(mdg4) in *Drosophila*. *Genetics* **155**(1), 141–157.

Bushey, AM., and Dorman, E. R., *et al*., (2008). Chromatin insulators: Regulatory mechanisms and epigenetic inheritance. *Mol. Cell* **32**(1), 1–9.

Bushey, A. M., and Ramos, E., *et al*., (2009). Three subclasses of a *Drosophila* insulator show distinct and cell type-specific genomic distributions. *Genes Dev.* **23**(11), 1338–1350.

Byrd, K., and Corces, V. G. (2003). Visualization of chromatin domains created by the gypsy insulator of *Drosophila*. *J. Cell Biol.* **162**(4), 565–574.

Capelson, M., and Corces, V. G. (2005). The ubiquitin ligase dTopors directs the nuclear organization of a chromatin insulator. *Mol. Cell* **20**(1), 105–116.

Capelson, M., and Corces, V. G. (2006). SUMO conjugation attenuates the activity of the gypsy chromatin insulator. *EMBO J.* **25**(9), 1906–1914.

Chen, X., and Xu, H., *et al*., (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**(6), 1106–1117.

Chernukhin, I. V., and Shamsuddin, S., *et al*., (2000). Physical and functional interaction between two pluripotent proteins, the Y-box DNA/RNA-binding factor, YB-1, and the multivalent zinc finger factor, CTCF. *J. Biol. Chem.* **275**(38), 29915–29921.

Chernukhin, I., and Shamsuddin, S., *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**(5), 1631–1648.

Cho, D. H., and Thienes, C. P., *et al*., (2005). Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol. Cell* **20**(3), 483–489.

Choi, J. -H., and Min, N. Y., *et al*., (2010). TSA-induced DNMT1 down-regulation represses hTERT expression via recruiting CTCF into demethylated core promoter region of hTERT in HCT116. *Biochem. Biophys. Res. Commun.* **391**(1), 449–454.

Cleary, J. D., and Tome, S., *et al*., (2010). Tissue- and age-specific DNA replication patterns at the CTG/CAG-expanded human myotonic dystrophy type 1 locus. *Nat. Struct. Mol. Biol.* **17**(99), 1079–1087.

Cuddapah, S., and Jothi, R., *et al*., (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* **19**(1), 24–32.

Darci, T. B., and Debora, N. M. -D., *et al*., (2004). DNA binding sites for putative methylation boundaries in the unmethylated region of the *BRCA1* promoter. *Int. J. Cancer* **111** (5), 669–678.

D'Arcy, V., and Pore, N., *et al*., (2008). BORIS, a paralogue of the transcription factor, CTCF, is aberrantly expressed in breast tumours. *Br. J. Cancer* **98**(3), 571–579.

Defossez, P. -A., and Kelly, K. F., *et al*., (2005). The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso. *J. Biol. Chem.* **280**(52), 43017–43023.

Degner-Leisso, S. C., and Feeney, A. J. (2010). Epigenetic and 3-dimensional regulation of V (D)J rearrangement of immunoglobulin genes. *Semin. Immunol.* 22(6), 346–352.

Degner, S. C., and Wong, T. P., *et al.*, (2009). Cutting edge: Developmental stage-specific recruitment of cohesin to CTCF sites throughout immunoglobulin loci during B lymphocyte development. *J. Immunol.* **182**(1), 44–48.

De La Rosa-Velázquez, I. A., and Rincón-Arano, H., *et al.*, (2007). Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. *Cancer Res.* **67**(6), 2577–2585.

Dhillon, N., and Raab, J., *et al.*, (2009). DNA polymerase epsilon, acetylases and remodellers cooperate to form a specialized chromatin structure at a tRNA insulator. *EMBO J.* **28**(17), 2583–2600.

Donohoe, M. E., and Silva, S. S., *et al.*, (2009). The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature* **460**(7251), 128–132.

Donohoe, M. E., and Zhang, L. -F., *et al.*, (2007). Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. *Mol. Cell* **25**(1), 43–56.

El-Kady, A., and Klenova, E. (2005). Regulation of the transcription factor, CTCF, by phosphorylation with protein kinase CK2. *FEBS Lett.* **579**(6), 1424–1434.

Emberly, E., and Blattes, R., *et al.*, (2008). BEAF regulates cell-cycle genes through the controlled deposition of H3K9 methylation marks into its conserved dual-core binding sites. *PLoS Biol.* **6**(12), 2896–2910.

Farrar, D., and Rai, S., *et al.*, (2010). Mutational analysis of the poly(ADP-ribosyl)ation sites of the transcription factor CTCF provides an insight into the mechanism of its regulation by poly(ADP-ribosyl)ation. *Mol. Cell. Biol.* **30**(5), 1199–1216.

Farrell, C. M., and Grinberg, A., *et al.*, (2000). A large upstream region is not necessary for gene expression or hypersensitive site formation at the mouse β-globin locus. *Proc. Natl. Acad. Sci. U.S.A.* **97**(26), 14554–14559.

Ferraiuolo, M. A., Rousseau, M., *et al.* (2010). The three-dimensional architecture of Hox cluster silencing. *Nucleic Acids Res* 38(21), 7472–7484.

Filippova, G. N. (2007). Genetics and epigenetics of the multifunctional protein CTCF. P. S. Gerald ed.), **Vol. 80**, pp. 337–360. Academic Press, New York.

Filippova, G. N., and Fagerlie, S., *et al.*, (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol. Cell. Biol.* **16**(6), 2802–2813.

Filippova, G. N., and Qi, C. -F., *et al.*, (2002). Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter its DNA-binding specificity. *Cancer Res.* **62**(1), 48–52.

Filippova, G. N., and Thienes, C. P., *et al.*, (2001). CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. *Nat. Genet.* **28**(4), 335–343.

Fitzpatrick, G. V., and Pugacheva, E. M., *et al.*, (2007). Allele-specific binding of CTCF to the multipartite imprinting control region KvDMR1. *Mol. Cell. Biol.* **27**(7), 2636–2647.

Gaszner, M., and Felsenfeld, G. (2006). Insulators: Exploiting transcriptional and epigenetic mechanisms. *Nat. Rev. Genet.* **7**(9), 703–713.

Gaszner, M., and Vazquez, J., *et al.*, (1999). The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction. *Genes Dev.* **13**(16), 2098–2107.

Gause, M., and Morcillo, P., *et al.*, (2001). Insulation of enhancer-promoter communication by a gypsy transposon insert in the *Drosophila* cut gene: Cooperation between suppressor of hairy-wing and modifier of mdg4 proteins. *Mol. Cell. Biol.* **21**(14), 4807–4817.

Gerasimova, T. I., and Byrd, K., *et al.*, (2000). A chromatin insulator determines the nuclear localization of DNA. *Mol. Cell* **6**(5), 1025–1035.

Gerasimova, T. I., and Gdula, D. A., *et al.*, (1995). A *Drosophila* protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. *Cell* **82**(4), 587–597.

Gerasimova, T. I., and Lei, E. P., *et al.*, (2007). Coordinated control of dCTCF and gypsy chromatin insulators in *Drosophila*. *Mol. Cell* **28**(5), 761–772.

Geyer, P. K., and Corces, V. G. (1992). DNA position-specific repression of transcription by a *Drosophila* zinc finger protein. *Genes Dev.* **6**(10), 1865–1873.

Ghosh, D., and Gerasimova, T. I., *et al.*, (2001). Interactions between the Su(Hw) and Mod (mdg4) proteins required for gypsy insulator function. *EMBO J.* **20**(10), 2518–2527.

Gilbert, M. K., and Tan, Y. Y., *et al.*, (2006). The *Drosophila* boundary element-associated factors BEAF-32A and BEAF-32B affect chromatin structure. *Genetics* **173**(3), 1365–1375.

Gombert, W. M., *et al.*, (2003). The c-myc insulator element and matrix attachment regions define the c-myc chromosomal domain. *Mol. Cell. Biol.* **23**, 9338–9348.

Gombert, W. M., and Krumm, A. (2009). Targeted deletion of multiple CTCF-binding elements in the human C-MYC gene reveals a requirement for CTCF in C-MYC expression. *PLoS ONE* **4**(7), e6109.

Guastafierro, T., and Cecchinelli, B., *et al.*, (2008). CCCTC-binding factor activates PARP-1 affecting DNA methylation machinery. *J. Biol. Chem.* **283**(32), 21873–21880.

Guelen, L., and Pagie, L., *et al.*, (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**(7197), 948–951.

Gyurkovics, H., and Gausz, J., *et al.*, (1990). A new homeotic mutation in the *Drosophila bithorax* complex removes a boundary separating two domains of regulation. *EMBO J.* **9** (8), 2579–2585.

Hadjur, S., and Williams, L. M., *et al.*, (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* **460**(7253), 410–413.

Hark, A. T., and Schoenherr, C. J., *et al.*, (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405**(6785), 486–489.

Harrison, DA., and Gdula, D. A., *et al.*, (1993). A leucine zipper domain of the suppressor of Hairy-wing protein mediates its repressive effect on enhancer function. *Genes Dev.* **7**(10), 1966–1978.

Hart, C. M., and Cuvier, O., *et al.*, (1999). Evidence for an antagonistic relationship between the boundary element-associated factor BEAF and the transcription factor DREF. *Chromosoma* **108**(6), 375–383.

Heath, H., and Ribeiro de Almeida, C., *et al.*, (2008). CTCF regulates cell cycle progression of alphabeta T cells in the thymus. *EMBO J.* **27**(21), 2839–2850.

Hikichi, T., and Kohda, T., *et al.*, (2003). Imprinting regulation of the murine Meg1/Grb10 and human GRB10 genes; roles of brain-specific promoters and mouse-specific CTCF-binding sites. *Nucleic Acids Res.* **31**(5), 1398–1406.

Holohan, E. E., and Kwong, C., *et al.*, (2007). CTCF genomic binding sites in *Drosophila* and the organisation of the bithorax complex. *PLoS Genet.* **3**(7), e112.

Hong, J. A., and Kang, Y., *et al.*, (2005). Reciprocal binding of CTCF and BORIS to the NY-ESO-1 promoter coincides with derepression of this cancer-testis gene in lung cancer cells. *Cancer Res.* **65**(17), 7763–7774.

Hoover, K. K., and Chien, A. J., *et al.*, (1993). Effects of transposable elements on the expression of the forked gene of *Drosophila melanogaster*. *Genetics* **135**(2), 507–526.

Hou, C., and Dale, R., *et al.*, (2010). Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl. Acad. Sci. U.S.A.* **107**(8), 3651–3656.

Huang, S., and Li, X., *et al.*, (2007). USF1 recruits histone modification complexes and is critical for maintenance of a chromatin barrier. *Mol. Cell. Biol.* **27**(22), 7991–8002.

Ishihara, K., and Oshimura, M., *et al*., (2006). CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol. Cell* **23**(5), 733–742.

Jelinic, P., and Shaw, P. (2007). Loss of imprinting and cancer. *J. Pathol.* **211**(3), 261–268.

Jelinic, P., and Stehle, J. -C., *et al*., (2006). The testis-specific factor CTCFL cooperates with the protein methyltransferase PRMT7 in H19 imprinting control region methylation. *PLoS Biol.* **4**(11), e355.

Jiang, N., and Emberly, E., *et al*., (2009). Genome-wide mapping of BEAF binding sites in *Drosophila* links BEAF to transcription. *Mol. Cell. Biol.* **29**(13), 3556–3568.

Johnatty, S. E., and Beesley, J., *et al*., (2010). Evaluation of candidate stromal epithelial cross-talk genes identifies association between risk of serous ovarian cancer and *TERT*, a cancer susceptibility "Hot-Spot". *PLoS Genet.* **6**(7), e1001016.

Jothi, R., and Cuddapah, S., *et al*., (2008). Genome-wide identification of in vivo protein–DNA binding sites from ChIP-Seq data. *Nucleic Acids Res.* **36**(16), 5221–5231.

Kang, Y., and Hong, J. A., *et al*., (2007). Dynamic transcriptional regulatory complexes including BORIS, CTCF and Sp1 modulate NY-ESO-1 expression in lung cancer cells. *Oncogene* **26**(30), 4394–4403.

Karch, F., o, M., and Galloni., *et al*., (1994). Mcp and Fab-7: Molecular analysis of putative boundaries of cis-regulatory domains in the bithorax complex of *Drosophila melanogaster*. *Nucleic Acids Res.* **22**(15), 3138–3146.

Kellum, R., and Schedl, P. (1992). A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol. Cell. Biol.* **12**(5), 2424–2431.

Kim, T. H., and Abdullaev, Z. K., *et al*., (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**(6), 1231–1245.

Kim, J., and Kollhoff, A., *et al*., (2003). Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, Peg3. *Hum. Mol. Genet.* **12**(3), 233–245.

Klenova, E. M., and Chernukhin, I. V., *et al*., (2001). Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. *Mol. Cell. Biol.* **21**(6), 2221–2234.

Klenova, E. M., and Morse, H. C., *et al*., (2002). The novel BORIS + CTCF gene family is uniquely involved in the epigenetics of normal biology and cancer. *Semin. Cancer Biol.* **12** (5), 399–414.

Kurukuti, S., Tiwari, V. K., *et al*., (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc. Natl. Acad. Sci. U.S.A.***103**(28), 10684–10689.

Labrador, M., and Corces, V. G. (2003). Extensive exon reshuffling over evolutionary time coupled to trans-splicing in *Drosophila*. *Genome Res.* **13**(10), 2220–2228.

Labrador, M., and Mongelard, F., *et al*., (2001). Protein encoding by both DNA strands. *Nature* **409**, 1000.

Lei, E. P., and Corces, V. G. (2006). RNA interference machinery influences the nuclear organization of a chromatin insulator. *Nat. Genet.* **38**(8), 936–941.

Libby, R. T., and Hagerman, K. A., *et al*., (2008). CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: A novel basis for mutational hot spot determination. *PLoS Genet.* **4**(11), e1000257.

Ling, J. Q., and Li, T., *et al*., (2006). CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science* **312**(5771), 269–272.

Loukinov, D. I., Pugacheva, E., *et al*., (2002). BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. *Proc. Natl. Acad. Sci. U.S.A.***99**(10), 6806–6811.

Lutz, M., and Burke, L. J., *et al*., (2000). Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Res.* **28**(8), 1707–1713.

Majumder, P., and Gomez, J. A., *et al*., (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**(4), 785–798.

Melnikova, L., and Juge, F., *et al*., (2004). Interaction between the GAGA factor and Mod (mdg4) proteins promotes insulator bypass in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **101**(41), 14806–14811.

Mishiro, T., and Ishihara, K., *et al*., (2009). Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *EMBO J.* **28**(9), 1234–1245.

Mohan, M., and Bartkuhn, M., *et al*., (2007). The *Drosophila* insulator proteins CTCF and CP190 link enhancer blocking to body patterning. *EMBO J.* **26**(19), 4203–4214.

Murrell, A., and Heeson, S., *et al*., (2004). Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nat. Genet.* **36**(8), 889–893.

Nativio, R., and Wendt, K. S., *et al*., (2009). Cohesin is required for higher-order chromatin conformation at the imprinted *IGF2-H19* locus. *PLoS Genet.* **5**(11), e1000739.

Nguyen, P., and Cui, H., *et al*., (2008). CTCFL/BORIS Is a methylation-independent DNA-binding protein that preferentially binds to the paternal H19 differentially methylated region. *Cancer Res.* **68**(14), 5546–5551.

Ohlsson, R., and Renkawitz, R., *et al*., (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet.* **17**(9), 520–527.

Ottaviani, A., and Rival-Gervier, S., *et al*., (2009a). The *D4Z4* macrosatellite repeat acts as a CTCF and A-type Lamins-dependent insulator in facio-scapulo-humeral dystrophy. *PLoS Genet.* **5**(2), e1000394.

Ottaviani, A., and Schluth-Bolard, C., *et al*., (2009b). Identification of a perinuclear positioning element in human subtelomeres that requires A-type Lamins and CTCF. *EMBO J.* **28**(16), 2428–2436.

Pai, C. Y., and Lei, E. P., *et al*., (2004). The centrosomal protein CP190 is a component of the gypsy chromatin insulator. *Mol. Cell* **16**(5), 737–748.

Pant, V., and Mariano, P., *et al*., (2003). The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the H19 imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. *Genes Dev.* **17**(5), 586–590.

Parelho, V., and Hadjur, S., *et al*., (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132**(3), 422–433.

Pathak, R. U., and Rangaraj, N., *et al*., (2007). Boundary element-associated factor 32B connects chromatin domains to the nuclear matrix. *Mol. Cell. Biol.* **27**(13), 4796–4806.

Phillips, J. E., and Corces, V. G. (2009). CTCF: Master weaver of the genome. *Cell* **137**(7), 1194–1211.

Qi, C. F., and Martensson, A., *et al*., (2003). CTCF functions as a critical regulator of cell-cycle arrest and death after ligation of the B cell receptor on immature B cells. *Proc. Natl. Acad. Sci. U. S. A.* **100**(2), 633–638.

Renaud, S., and Loukinov, D., *et al*., (2005). CTCF binds the proximal exonic region of hTERT and inhibits its transcription. *Nucleic Acids Res.* **33**(21), 6850–6860.

Renaud, S., and Loukinov, D., *et al*., (2010). BORIS/CTCFL-mediated transcriptional regulation of the hTERT telomerase gene in testicular and ovarian tumor cells. *Nucleic Acids Res.***39**(3), 862–873

Renaud, S., and Pugacheva, E. M., *et al*., (2007). Expression of the CTCF-paralogous cancer-testis gene, brother of the regulator of imprinted sites (BORIS), is regulated by three

alternative promoters modulated by CpG methylation and by CTCF and p53 transcription factors. *Nucleic Acids Res.* **35**(21), 7372–7388.

Rodriguez, C., and Borgel, J., *et al.*, (2010). CTCF is a DNA methylation-sensitive positive regulator of the INK/ARF locus. *Biochem. Biophys. Res. Commun.* **392**(2), 129–134.

Roy, S., and Gilbert, M. K., *et al.*, (2007a). Characterization of BEAF mutations isolated by homologous recombination in *Drosophila*. *Genetics* **176**(2), 801–813.

Roy, S., and Tan, Y. Y., *et al.*, (2007b). A genetic screen supports a broad role for the *Drosophila* insulator proteins BEAF-32A and BEAF-32B in maintaining patterns of gene expression. *Mol. Genet. Genom.* **277**(3), 273–286.

Rubio, E. D., and Reiss, D. J., *et al.*, (2008). CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. U. S. A.* **105**(24), 8309–8314.

Sandhu, K. S., and Shi, C., *et al.*, (2009). Nonallelic transvection of multiple imprinted loci is organized by the H19 imprinting control region during germline development. *Genes Dev.* **23**(22), 2598–2603.

Schweinsberg, S., and Hagstrom, K., *et al.*, (2004). The enhancer-blocking activity of the Fab-7 boundary from the *Drosophila bithorax* complex requires GAGA-factor-binding sites. *Genetics* **168**(3), 1371–1384.

Sekimata, M., and Pérez-Melgosa, M., *et al.*, (2009). CCCTC-binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon-[gamma] locus. *Immunity* **31**(4), 551–564.

Splinter, E., and Heath, H., *et al.*, (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* **20**(17), 2349–2354.

Stedman, W., and Kang, H., *et al.*, (2008). Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *EMBO J.* **27**(4), 654–666.

Tam, A. S., and Devereux, T. R., *et al.*, (2003). Perturbations of the Ink4a/Arf gene locus in aflatoxin B1-induced mouse lung tumors. *Carcinogenesis* **24**(1), 121–132.

Tolhuis, B., and Palstra, R. -J., *et al.*, (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* **10**(6), 1453–1465.

Torrano, V., and Chernukhin, I., *et al.*, (2005). CTCF regulates growth and erythroid differentiation of human myeloid leukemia cells. *J. Biol. Chem.* **280**(30), 28152–28161.

Wallace, J. A., and Felsenfeld, G. (2007). We gather together: Insulators and genome organization. *Curr. Opin. Genet. Dev.* **17**(5), 400–407.

Wendt, KS., and Yoshida, K., *et al.*, (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**(7180), 796–801.

Weth, O., and Weth, C., *et al.*, (2010). Modular insulators: Genome wide search for composite CTCF/thyroid hormone receptor binding sites. *PLoS ONE* **5**(4), e10119.

Witcher, M., and Emerson, B. M. (2009). Epigenetic silencing of the p16INK4a tumor suppressor is associated with loss of CTCF binding and a chromatin boundary. *Mol. Cell* **34**(3), 271–284.

Xu, N., and Donohoe, M. E., *et al.*, (2007). Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. *Nat. Genet.* **39**(11), 1390–1396.

Xu, J., and Huo, D., *et al.*, (2010). CpG island methylation affects accessibility of the proximal BRCA1 promoter to transcription factors. *Breast Cancer Res. Treat.* **120**(3), 593–601.

Yao, H., Brick, K., *et al.*, (2010). Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.* 24 (22), 2543-2555.

Yoon, B., and Herman, H., *et al.*, (2005). Rasgrf1 imprinting is regulated by a CTCF-dependent methylation-sensitive enhancer blocker. *Mol. Cell. Biol.* **25**(24), 11184–11190.

Yu, W., and Ginjala, V., *et al.*, (2004). Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. *Nat. Genet.* **36**(10), 1105–1110.

Yusufzai, T. M., and Tagami, H., *et al*., (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol. Cell* **13**(2), 291–298.

Zhao, K., and Hart, C. M., *et al*., (1995). Visualization of chromosomal domains with boundary element-associated factor BEAF-32. *Cell* **81**(6), 879–889.

Zhao, Z., and Tavoosidana, G., *et al*., (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* **38**(11), 1341–1347.