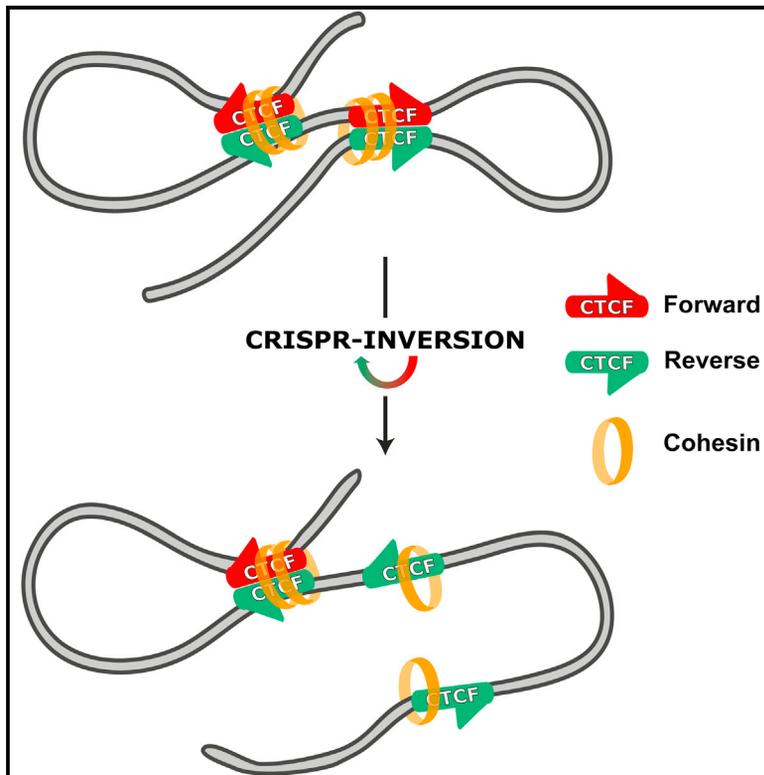


Molecular Cell

CTCF Binding Polarity Determines Chromatin Looping

Graphical Abstract



Authors

Elzo de Wit, Erica S.M. Vos,
Sjoerd J.B. Holwerda, ...,
Patrick J. Wijchers, Peter H.L. Krijger,
Wouter de Laat

Correspondence

w.delaat@hubrecht.eu

In Brief

CCCTC-binding factor (CTCF) shapes the three-dimensional genome. Here, de Wit et al. provide direct evidence for CTCF binding polarity playing an underlying role in chromatin looping. Cohesin association persists, but inverted CTCF sites fail to loop, which can sometimes lead to long-range changes in gene expression.

Highlights

- CTCF binding polarity determines chromatin looping
- Inverted or disengaged CTCF sites do not necessarily form new chromatin loops
- Cohesin recruitment to CTCF sites is independent of loop formation
- Phenocopied linear but altered 3D chromatin landscape can affect gene expression

Accession Numbers

GSE72720



CTCF Binding Polarity Determines Chromatin Looping

Elzo de Wit,^{1,2} Erica S.M. Vos,^{1,2} Sjoerd J.B. Holwerda,^{1,2} Christian Valdes-Quezada,^{1,2} Marjon J.A.M. Verstegen,¹ Hans Teunissen,¹ Erik Splinter,¹ Patrick J. Wijchers,¹ Peter H.L. Krijger,¹ and Wouter de Laat^{1,*}

¹Hubrecht Institute-KNAW and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, the Netherlands

²Co-first author

*Correspondence: w.delaat@hubrecht.eu

<http://dx.doi.org/10.1016/j.molcel.2015.09.023>

SUMMARY

CCCTC-binding factor (CTCF) is an architectural protein involved in the three-dimensional (3D) organization of chromatin. In this study, we assayed the 3D genomic contact profiles of a large number of CTCF binding sites with high-resolution 4C-seq. As recently reported, our data also suggest that chromatin loops preferentially form between CTCF binding sites oriented in a convergent manner. To directly test this, we used CRISPR/Cas9 genome editing to delete core CTCF binding sites in three loci, including the CTCF site in the *Sox2* super-enhancer. In all instances, CTCF and cohesin recruitment were lost, and chromatin loops with distal, convergent CTCF sites were disrupted or destabilized. Re-insertion of oppositely oriented CTCF recognition sequences restored CTCF and cohesin recruitment, but did not re-establish chromatin loops. We conclude that CTCF binding polarity plays a functional role in the formation of higher-order chromatin structure.

INTRODUCTION

Chromosome topology strongly influences genome function (Bickmore and van Steensel, 2013; de Laat and Duboule, 2013). Chromosomes are structurally subdivided into topologically associated domains (TADs), evolutionarily and developmentally stable regions within which DNA sequences preferentially contact each other (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). TADs and their developmentally more dynamic substructures, called sub-TADs (Phillips-Cremins et al., 2013; Rao et al., 2014) serve to functionally insulate intervening sequences. Within TADs, chromatin loops can be formed between enhancers and distant target genes to increase their transcription (Deng et al., 2012). Boundaries between TADs hamper enhancer action on genes in other TADs (Downen et al., 2014; Nora et al., 2012; Symmons et al., 2014).

CCCTC-binding factor (CTCF) is a developmentally essential protein (Heath et al., 2008) that plays a central role in the folding and segmentation of mammalian chromosomes (Ong and Corces, 2014). CTCF binds to tens of thousands of genomic sites, > 60% of which are bound in a tissue-invariant manner (Chen

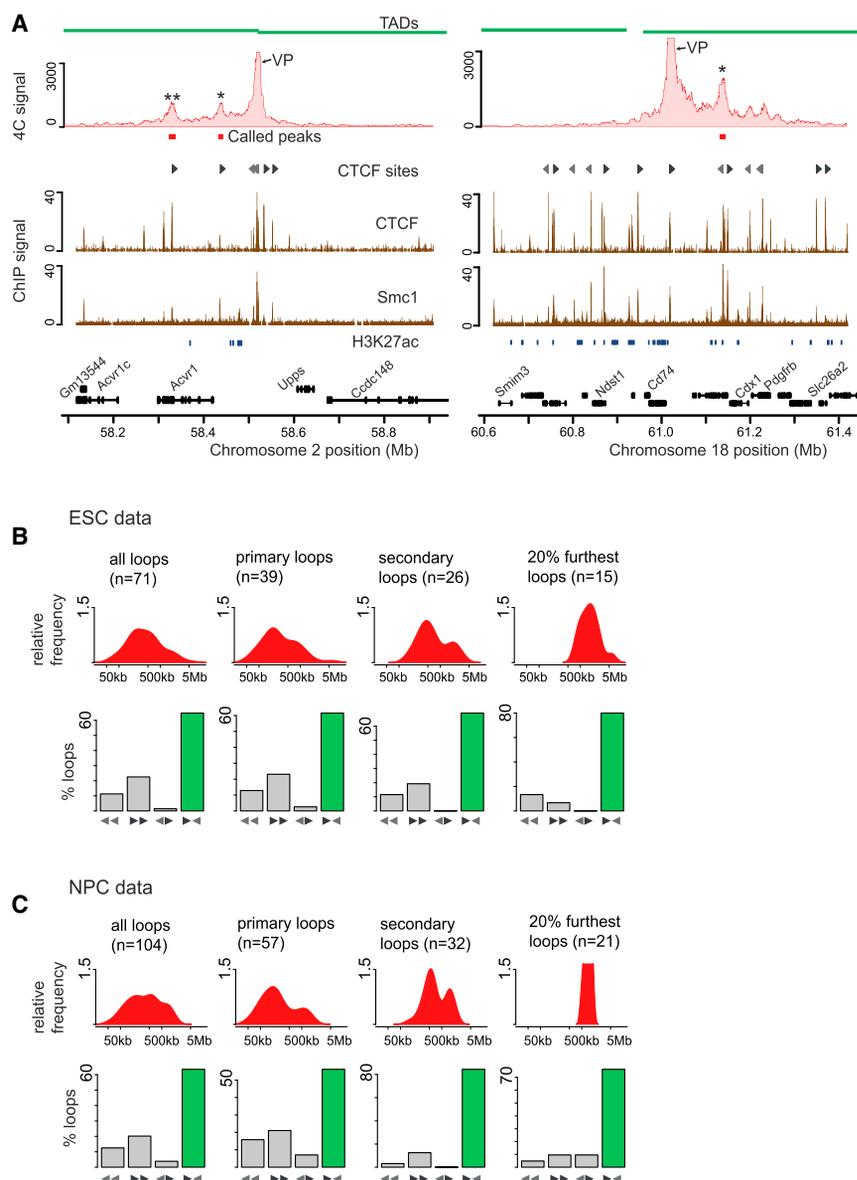
et al., 2012; Faure et al., 2012). It has been implicated in gene activation, repression, and insulation (Kim et al., 2007; Nakahashi et al., 2013). CTCF can form chromatin loops (Splinter et al., 2006) and is enriched at boundaries of topological domains (Dixon et al., 2012; Yaffe and Tanay, 2011). CTCF sites engaged in chromatin looping are often co-occupied by cohesin, a protein complex that also binds to enhancers, independent of CTCF. Cohesin also actively contributes to loop formation, possibly through its capacity to embrace and hold together two DNA molecules (Hadjur et al., 2009; Kagey et al., 2010; Parelho et al., 2008; Wendt et al., 2008). Based on depletion experiments, it is thought that CTCF is mainly involved in the segmentation of chromosomes, while cohesin may act more locally to organize structure within domains (Seitan et al., 2013; Sofueva et al., 2013; Zuin et al., 2014).

There are many more CTCF binding sites than TAD boundaries, and CTCF sites also far outnumber chromatin loops currently found per cell type (Handoko et al., 2011; Rao et al., 2014). This may be because genome-wide methods still lack the resolving power to detect all CTCF-bound loops. It can also be because only certain combinations of CTCF binding sites are capable of loop formation. It was recently shown that the great majority of CTCF chromatin loops involve pairs of sites with CTCF binding motifs oriented in a convergent manner (Gómez-Marín et al., 2015; Rao et al., 2014; Vietri Rudan et al., 2015). To further investigate the relationship between CTCF binding polarity and chromatin looping, we created high-resolution chromosome contact maps for a large number of CTCF sites. In addition, we studied the topological and functional consequences of deleting and inverting specific CTCF sites at multiple genomic locations.

RESULTS AND DISCUSSION

4C Peak Calling for Systematic Identification of Chromatin Loops

To analyze chromatin looping mediated by CTCF, we first defined the genome-wide occupancy of CTCF. We performed ChIP-seq in murine embryonic stem cells (ESCs) and neural progenitor cells (NPCs) and found 54,151 and 65,983 bound CTCF sites, respectively. The majority ($n = 37,298$) of CTCF sites were found to be conserved between these two cell types. Based on the ChIP-seq data, we selected 86 CTCF sites for subsequent 4C analysis. 4C-seq is a method that generates detailed contact profiles of selected genomic sites. For comparison, the highest-resolution Hi-C map to date assays on average ~750



independent contacts per individual restriction fragment (Rao et al., 2014), whereas our 4C-seq profiles interrogate a minimum of 4,314 and up to 87,462 independent contacts per target site. 4C-seq may thus better resolve DNA interactions of individual sites at lower sequencing depth.

In order to systematically identify chromatin loops in our compendium of 4C profiles, we developed a peak calling algorithm (see [Experimental Procedures](#) for details). Briefly, we define a peak as an increase in signal over (perceived) background. In contrast to ChIP-seq or related datasets, which have a close to uniform background distribution, 4C profiles have a strongly non-uniform data distribution (Figure S1A). Importantly, this distribution can differ between 4C experiments at different locations in the genome. The most important contributor to this phenomenon is distance to a TAD border. A 4C experiment on one side of a TAD border can have a completely different background distribu-

Figure 1. 4C-Seq Shows that CTCF Motif Orientation Is Associated with Chromatin Looping

(A) Example 4C-seq experiments showing looping. Top panels show smoothed (running mean) 4C profiles. Below the 4C signal, peaks identified by our peak calling algorithm are shown by red rectangles ("Called peaks"). "VP" indicates 4C viewpoint, (*) indicate primary loops, and (**) indicate secondary loops. Gray triangles show the forward and reverse orientation of shared CTCF-cohesin binding sites. Bottom panels show ChIP-seq signal for CTCF and Smc1 (cohesin). TADs (green rectangles, top) and H3K27ac sites (bottom) are indicated. See also [Figure S1](#).

(B) Quantification of all CTCF motif orientations that we identified engaged in chromatin looping in ESCs. Top panels show a density plot of the \log_{10} size distribution of the loops in each class.

(C) Same as (B) but for NPCs.

tion compared to a viewpoint on the other side. To account for this, our algorithm explicitly models the background separately on each side of the viewpoint using monotonic regression. In a monotonic regression analysis, the signal is forced to decrease with increased distance, which is a fair assumption for an unstructured chromatin fiber (Rippe et al., 1995). The result is a regression line that closely follows the background distribution in a 4C experiment (Figure S1B). Using a re-sampling-based statistical analysis (see [Experimental Procedures](#)), we subsequently identify regions that show a robust increase over the background (Figure S1C). This method stringently detects chromatin loops. Compared to FourCSeq, a recently published analysis pipeline for 4C-seq data based on similar principles (Klein et al., 2015), our analysis pipeline identifies fewer contacts. The number of

peaks we find are more similar to the number chromatin loops identified in high-resolution Hi-C data (Rao et al., 2014).

Looping Preferentially Occurs between Convergent CTCF Motifs

We investigated whether the correlation between CTCF binding site polarity and chromatin loops is also apparent in our 4C data. We identified the summits of called peaks (i.e., the highest value within a contacted region) and with them associated the single nearest CTCF site, regardless of whether or not it also had cohesin associated. This enabled us to compare a CTCF site's orientation to that of the 4C target CTCF site. Indeed, as was found by genome-wide high-resolution Hi-C contact maps (Rao et al., 2014; Vietri Rudan et al., 2015), loop formation preferentially occurred between convergent CTCF sites (see [Figure 1A](#) for examples). This was true not only for primary loops formed with the

nearest contacting CTCF site (65% convergent, 1% divergent, and 34% in the same orientation), but also for secondary loops formed with sequences beyond the first contacted CTCF site (Figure 1B). Even the 20% most distal contacts scored, which formed loops over 0.9–5.8 Mb, showed this preference for convergence (Figure 1B).

We found that ~10%–20% of chromatin loops occurred between identically oriented CTCF sites, which is relatively high compared to previous Hi-C results (Rao et al., 2014). We attribute this finding to the frequent close linear juxtaposition of divergent CTCF sites (Pugacheva et al., 2015). 4C applied to one of these closely juxtaposed sites frequently results in the detection of loops both up- and downstream of the viewpoint, of which the equivalently oriented site is then necessarily misattributed. The all-versus-all nature of Hi-C better enables one to discern which of the juxtaposed CTCF sites is responsible for each oppositely oriented loop.

We also generated 81 4C profiles from CTCF sites in NPCs. Here, we find the same preference for convergent CTCF sites to form loops, with an almost equal preference for convergent sites in primary and secondary loops (Figure 1C), despite the fact that some CTCF sites, most notably around tissue-specific genes, clearly showed tissue-specific contact profiles (Figure S1D). These results confirm the importance of CTCF motif orientation in chromatin loop formation.

Genome Editing for the Deletion and Inversion of Endogenous CTCF Sites

To test the functional importance of CTCF binding polarity, we deleted and inverted a number of CTCF binding sites in ESCs. We selected target sites based on CTCF binding site motif orientation, efficient recruitment of CTCF and cohesin (both judged by ChIP-seq scores), and the formation of a convergent loop (as judged by 4C-seq). Based on these criteria, three individual CTCF binding sites were selected (Figure S2A), located near the genes *Malt1*, *Sox2*, and *Fbn2*. We used the CRISPR/Cas9 system (Ran et al., 2013) with oligonucleotide repair templates to site-specifically modify the selected CTCF sites in murine ESCs. For CTCF binding site deletions, we removed 9, 13, and 16 bp of the core CTCF binding motif at the *Malt1*, *Fbn2*, and *Sox2* loci, respectively, and replaced them by an EcoRI recognition site (GAATTC) to facilitate screening (Figure S2B). We obtained multiple homozygous deletion (del/del) clones for each targeted site.

To invert the orientation of the CTCF binding sites, the same 9, 13, and 16 consensus base pairs were removed, but were replaced by an inverted CTCF binding motif. To prevent the repair template from being targeted by the same guide RNA and modified by CRISPR/Cas9, the replacement site had a different (oppositely oriented) CTCF binding motif. Since chromatin loops are mediated by associated proteins, we reasoned that this would well serve to test the importance of CTCF binding polarity as long as (1) our deletion clones showed complete absence of CTCF binding and (2) the reintroduced, oppositely oriented binding motif fully restored CTCF binding. The result would be a phenocopied association of CTCF protein to the exact same genomic location, but now in an opposite orientation. For this purpose, we selected a binding site that efficiently recruits

CTCF (based on ChIP-seq) and which we confirmed participates in convergent looping by 4C-seq (Figure S3). This site was inserted at all targeted locations as a 50-bp sequence that contained both CTCF “core” and “downstream” binding modules (“M1” and “M2” [Schmidt et al., 2012]). We obtained one clone with an inverted and an untargeted CTCF site (inv/wt) at the *Malt1* locus, seven homozygous inversion clones (inv/inv) for the CTCF site near *Sox2*, one homozygous inversion (inv/inv) clone for the CTCF site at *Fbn2* (with on both alleles only the “core” M1 binding module, not the M2 motif), and one clone with an inverted and a deleted CTCF site at the two *Fbn2* alleles (inv/del). All deletions and inversions were confirmed by Sanger sequencing (Figures S2B and S3C).

CTCF Binding Is Necessary for Loop Formation

To evaluate the relevance of CTCF binding for chromatin looping, we first revisited an available transgenic mouse line carrying four mutated nucleotides in one of the CTCF binding sites (3'HS1) of the β -globin locus. We previously found that this mutant 3'HS1 fails to bind CTCF and no longer participates in loop formation with other CTCF sites, judged at the time by 3C technology (Splinter et al., 2006). 4C-seq can provide a much more comprehensive overview of chromosome contact changes. It confirmed that mutated 3'HS1 had strongly reduced contact frequencies with downstream CTCF sites at the other end of the β -globin locus, which indeed were in the convergent orientation (Figure S2C).

We then analyzed the del/del ESC clones we had obtained by genome editing. Importantly, ChIP confirmed that CTCF binding was abolished in all del/del clones (Figure 2A). Consistently, this was accompanied by loss of cohesin association with the target sites (Figure 2B). 4C-seq uncovered that spatial contacts with the deleted sites were disrupted in the CTCF del/del clones to varying degrees. For example, upon deletion of the reverse oriented CTCF binding site centered at chr18:65,649,680 in the *Malt1* locus (indicated by scissors, Figure 2C), visual inspection of spatial contact profiles revealed the specific disruption of a chromatin loop with an upstream (convergently oriented) CTCF site nearly 200 kb away (indicated by an asterisk, Figure 2C). This loss of contact was confirmed by the peak caller. The deleted site is the only one of multiple juxtaposed CTCF sites, all within 10 kb, that is in the reverse orientation. All others, of which the closest is less than 500 bp from the target site, are in the forward orientation. They are likely responsible for the downstream loops that we also score using this viewpoint. The primary downstream loop also appeared affected by the deletion, but to a lesser extent. These results may suggest that of the four clustered CTCF sites, the reversely oriented site is specifically responsible for the formation of the upstream loop.

In the *Sox2* locus, we deleted the sole CTCF site that is contained within the 13 kb super-enhancer recently described to loop toward the *Sox2* gene and responsible for 90% of its expression in ESCs (Li et al., 2014; Zhou et al., 2014). This CTCF site is in a convergent orientation with the CTCF site at the promoter of *Sox2* (Figure 2D) and is surrounded by sequences within ~5 kb that efficiently recruit cohesin apparently in a CTCF-independent manner (Figure S2A) (Whyte et al.,

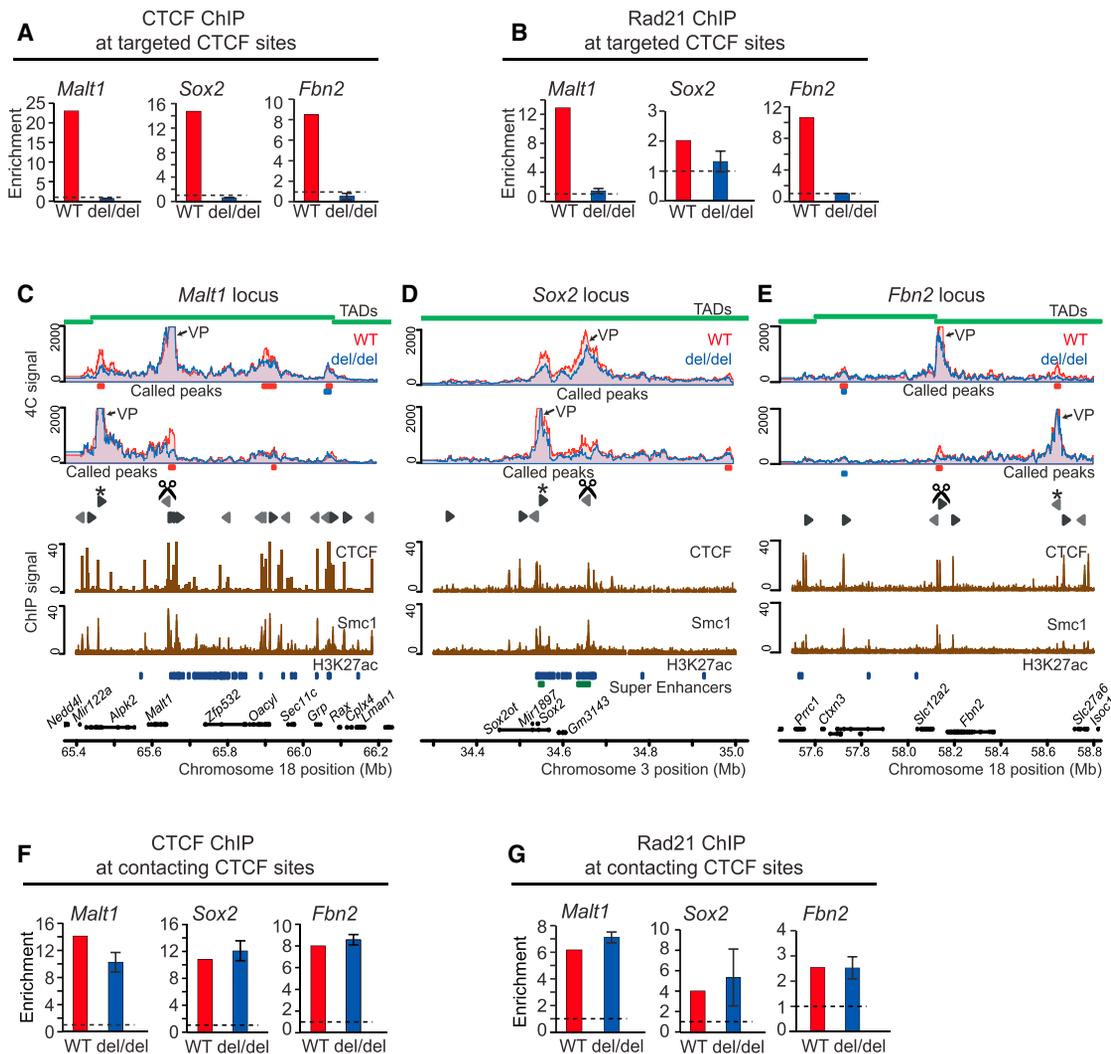


Figure 2. Chromatin Looping Requires CTCF Association

(A and B) ChIP-qPCR results showing that disruption of the CTCF consensus sequence via genome editing abolishes CTCF (A) and Rad21/cohesin (B) binding. Shown are the average (\pm SEM) of two independent del/del clones (blue) for each locus (*Malt1*, *Sox2*, *Fbn2*), versus wild-type (red), expressed as ChIP enrichment over an unbound site (*Actin* promoter). Value of 1 (dashed line) indicates no enrichment.

(C–E) 4C-seq contact profiles (averages of at least two biological replicates) showing overlays of contacts formed by wild-type (red) and deleted (del/del, blue) CTCF sites at the *Malt1* (C), *Sox2* (D), and *Fbn2* (E) loci. The orientation of shared CTCF/cohesin sites is shown by gray triangles. Targeted CTCF sites (scissors), the convergent contact partners (asterisks), and 4C viewpoints (VP) are indicated. Top plots show 4C results from targeted site, bottom from contacted site. Called contacts in wild-type (red) and deleted (blue) clones are indicated. Underneath 4C plots, CTCF and Smc1/cohesin ChIP-seq profiles, genes, H3K27ac peaks (dark blue), and super-enhancers (dark green) (Whyte et al., 2013) are indicated. See also Figure S2.

(F and G) ChIP-qPCR results plotted as above, showing that CTCF (F) and Rad21/cohesin (G) binding at the detached looping partners (asterisks) are not affected.

2013). Abrogating CTCF binding at the enhancer clearly reduced, but did not abolish, enhancer-promoter contacts, suggesting that CTCF is only partially responsible for the promoter-enhancer loop in the *Sox2* locus (Figure 2D). Note that our peak caller misses the clear chromatin loop formed between the super enhancer and the *Sox2* gene. This is because the presence of a broad peak relatively close to the viewpoint leads to a signal that does not reach our strict threshold for peak calling.

At the *Fbn2* locus, deletion of the forward-oriented CTCF binding site centered at chr18:58,136,460 abolished the wild-type

loop formed with a convergent CTCF site more than 400 kb downstream of the target site (Figure 2E). An oppositely oriented loop with a forward oriented CTCF site \sim 350 kb upstream of the target site was mildly destabilized. Presumably this loop relied more on the reversely oriented CTCF site near the target site (Figures 2E and S2A).

We validated the results by generating 4C-seq profiles at the originally contacted CTCF sites (asterisks, Figures 2C–2E). In wild-type cells, these showed chromatin looping with the target sites, as expected. These loops were abolished in the *Malt1* del/del and *Fbn2* del/del clones and destabilized in the *Sox2* del/del

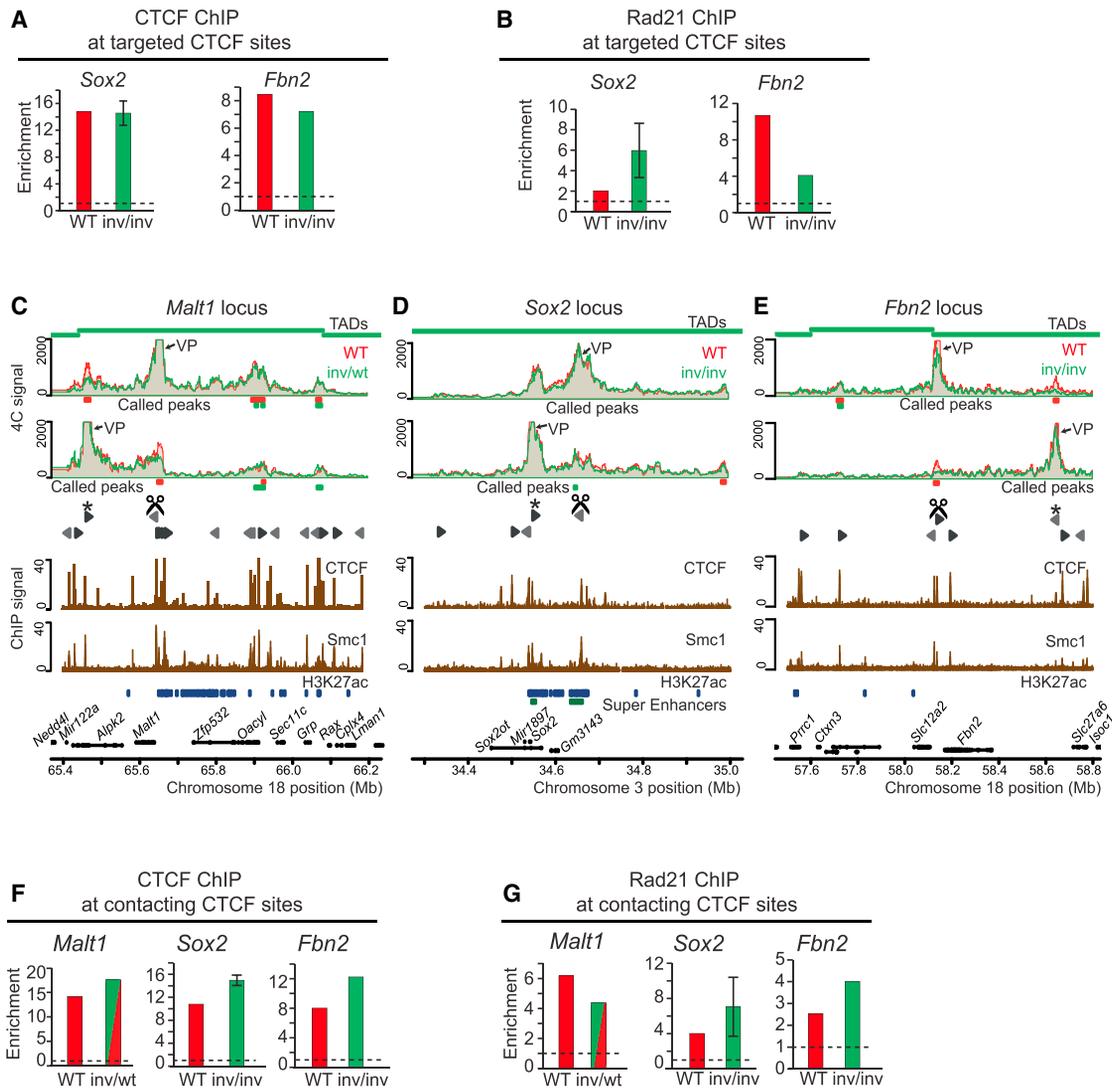


Figure 3. Inverted CTCF Sites Restore CTCF Binding but Not Chromatin Looping

(A and B) ChIP-qPCR results showing that inversion of the CTCF consensus sequence via genome editing enables CTCF (A) and Rad21/cohesin (B) recruitment to the targeted site. Shown is the average (\pm SEM) of two independent inv/inv clones for *Sox2*, or the ChIP enrichment for the one inv/inv of *Fbn2* (both in green), versus wild-type (red).

(C–E) 4C-seq contact profiles (averages of at least two biological or two technical [*Malt1* inv/wt, *Fbn2* inv/inv] replicates) showing overlays of contacts formed by wild-type (red) and inverted (green) CTCF sites at *Malt1* (C), *Sox2* (D), and *Fbn2* (E) loci. Called contacts in wild-type (red) and inverted (green) clones are indicated. See also Figure S3.

(F and G) ChIP-qPCR results plotted as above, showing that CTCF (F) and Rad21/cohesin (G) binding at the detached looping partners (asterisks) is not affected.

clones, consistent with our other 4C observations. Notably, none of these three contacting CTCF sites showed specific new loops to other CTCF sites. Judging by their CTCF and cohesin occupancy (Figures 2F and 2G), this was not because they had intrinsically lost looping capacity. Productive loop formation therefore also seems to rely on genomic context.

Inverted CTCF Sites Restore CTCF Binding but Not Chromatin Looping

Next, we analyzed the impact of re-inserting an oppositely oriented CTCF binding motif at the same three sites that no longer

bound CTCF or cohesin in the deletion clones. At the *Sox2* and *Fbn2* sites, for which we had homozygous inv/inv clones, this fully restored CTCF recruitment (Figure 3A). Cohesin was also recruited, albeit with varying efficiency (Figure 3B). At the *Malt1* locus, for which we only obtained a heterozygous inv/wt clone, ChIP followed by semiquantitative PCR analysis confirmed that the inverted site bound CTCF and cohesin at least as efficiently as its wild-type counterpart on the other allele (Figure S3D). The inverted sites are therefore indiscernible from the wild-type sites with respect to CTCF binding as assayed by ChIP. Their chromosomal contact profiles, however, were different from wild-type

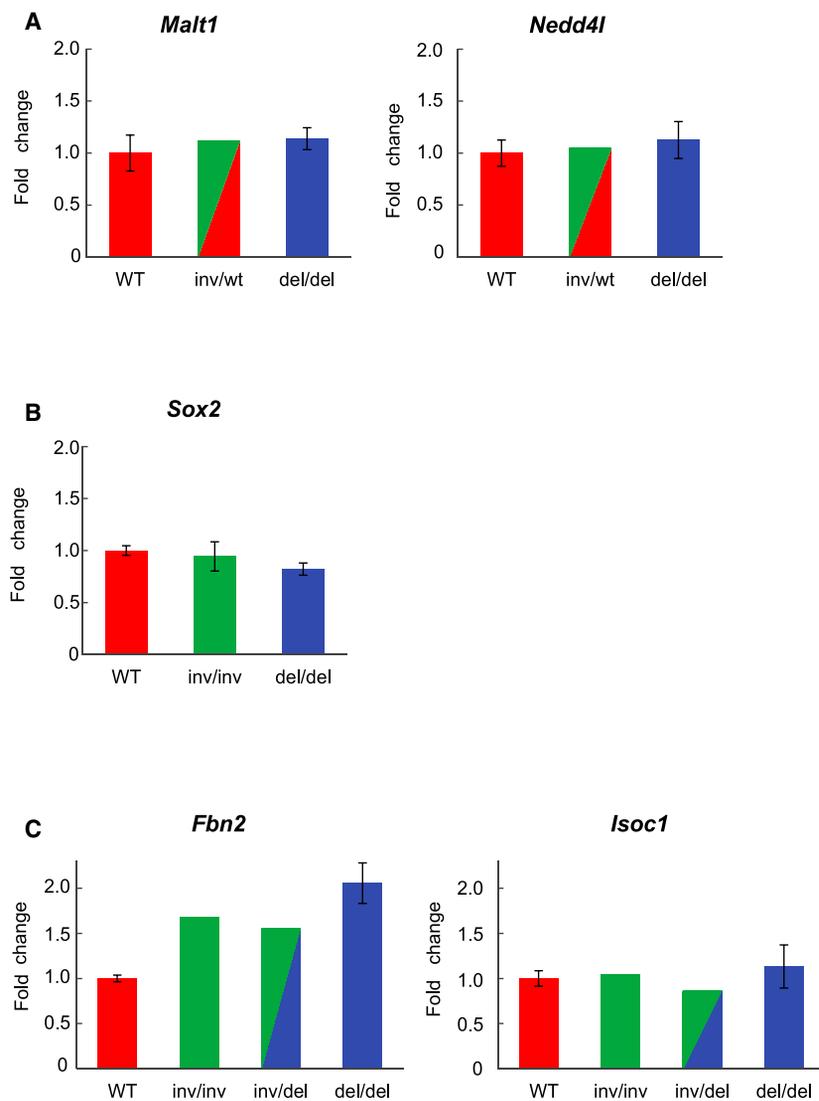


Figure 4. Chromatin Loop Disruption Can Dysregulate Gene Expression

(A–C) qPCR expression analysis of genes at the (A) *Malt1* locus, (B) *Sox2* locus, and (C) *Fbn2* locus. Shown is the average (\pm SEM) of multiple clones, each analyzed in triplicate. Expression without error bars indicates that a single clone was analyzed in triplicate. See also Figure S4.

sites in new replacement loops, despite unaltered cohesin and CTCF recruitment to these sites (Figures 3F and 3G).

Chromatin Loop Disruption May Alter Gene Expression

Finally, we addressed whether changes in chromatin architecture would affect the expression of nearby genes. We previously found no difference in β -globin gene expression in fetal livers carrying a CTCF binding site mutation in 3'HS1 (Splinter et al., 2006). Deletion (del/del) or inversion (inv/wt) of the CTCF site at the *Malt1* locus did not alter the expression level of the gene contained within the disrupted chromatin loop (*Malt1*; the other gene, *Alpk2*, is not active), nor the levels of *Nedd4l*, the gene flanking the original loop (Figure 4A). At the *Sox2* locus, contact frequencies between the promoter and the CTCF site in the super-enhancer were reduced, yet this led to no (inv/inv) or only a subtle (del/del) reduction in *Sox2* expression (Figure 4B). At the *Fbn2* locus, however, we found that the *Fbn2* gene, contained within the wild-type chromatin loop, was upregulated 1.5- to 2.5-fold. This was consistently found with two independent qPCR primer sets for five independent del/del clones as

and resembled those seen for the deletion clones (Figures 2C and 3C).

The heterozygous inv/wt site at the *Malt1* locus selectively lost contacts with the upstream CTCF site at 200 kb (asterisk, Figure 3C), as also was seen upon deleting the site. The inverted CTCF site in the *Sox2* super-enhancer showed a reduction in contact frequencies with the *Sox2* promoter (Figure 3D), albeit less obvious than seen after deleting the site. Finally, the inverted CTCF binding motif at the *Fbn2* locus caused disruption of the same > 400 kb downstream chromatin loop that was also lost upon deletion of this same CTCF binding site (Figure 3E). In none of the three instances did we find evidence for the inverted CTCF site engaging in new, oppositely oriented chromatin loops (Figures 3C–3E). Reciprocal 4C-seq experiments that looked from the originally contacted CTCF sites confirmed disruption of the chromatin loops at the *Malt1* and *Fbn2* loci and also confirmed a slightly reduced contact frequency between the *Sox2* promoter and enhancer (Figures 3C–3E). These profiles likewise revealed no engagement of the originally contacted

well as the inv/inv clone, as compared to six independent clones carrying a wild-type *Fbn2* locus (Figures 4C and S4). This implies that disruption of the encompassing CTCF loop leads to *Fbn2* upregulation. Expression levels of the more downstream gene *Isoc1* varied more than 2-fold between all independent clones (Figure S4), but this appeared unrelated to the targeting of the CTCF site. The example of *Fbn2* demonstrates that in a genomic context indistinguishable from wild-type by ChIP, a gene can be expressed at different levels because of an oppositely oriented chromatin-associated CTCF molecule \sim 230 kb downstream of its transcriptional start site.

Collectively, we have shown that deletion of a few nucleotides that abolish CTCF binding is sufficient to specifically disrupt chromatin loops between convergent CTCF sites. In all cases, re-introduction of an oppositely oriented 50 bp CTCF binding motif resulted in a CTCF site that is indistinguishable by ChIP from the original wild-type site. Yet the oppositely oriented CTCF molecules were unable to re-establish the original chromatin loops. We therefore conclude that CTCF binding polarity

plays an underlying role in chromatin looping. Our data are in agreement with the recent demonstration that inversion of a kilobase-sized region carrying multiple CTCF/cohesin sites affects chromosome topology and expression (Guo et al., 2015). However, rather than inverting a genomic region, our genome editing experiments only specifically disrupt single CTCF binding sites.

The inverted sites behaved like CTCF-depleted sites: they not only lost the ability to form chromatin loops in one direction, but they also did not engage in specific contacts with CTCF sites in the other direction. The same was true for the disengaged CTCF sites (i.e., the original looping partners), which also did not form new long-range contacts with other CTCF sites. The observation that cohesin association at these sites persists when disengaged suggests that cohesin recruitment to CTCF sites is independent of stable chromatin loop formation. Loop formation by CTCF therefore also seems dependent on chromosomal context, most likely requiring the availability of compatible binding sites in close enough linear proximity and a local 3D configuration that can facilitate loop formation. A super-enhancer, with CTCF-independent associated cohesin, seems capable of efficient looping to a gene promoter even without the help of paired CTCF sites, as seen here for the *Sox2* locus.

Although disruption of CTCF-anchored loops does not always cause dysregulation of genes, it can lead to gene expression changes. This was recently seen in rostral motor neurons, where disruption of a CTCF site caused ectopic *HoxA7* expression (Narendra et al., 2015), as well as here for the *Fbn2* gene, which was upregulated by both the deletion and inversion of a CTCF site ~230 kb away from its transcription start site.

EXPERIMENTAL PROCEDURES

CTCF ChIP-Seq, ChIP-qPCR, and 4C

ChIP was performed according to standard protocols using 6 μ l anti-CTCF (Millipore #07-729) and 7 μ l anti-Rad21 (Abcam #ab992) antibody per ChIP, respectively. ChIP-seq was performed with anti-CTCF antibody (Millipore #07-729) according to manufacturer protocol, with slight modifications (see Supplemental Information). 4C templates were prepared as previously described (Splinter et al., 2012).

4C Peak Calling

In contrast to, for example, ChIP-seq, the background in a 4C experiment is distributed non-uniformly. To perform peak calling, we explicitly model the background of up- and downstream genomic regions independently. We assume that in an unstructured chromatin fiber the contact profile is the result of random encounters between genomic loci. Therefore, the contact probability monotonically decreases as a function of distance to the viewpoint. By performing monotonic regression of the 4C signal as a function of distance to the viewpoint, we can model this decay. This is done using the Pool Adjacent-Violators Algorithm (PAVA) from the R package isotone, which is an implementation of monotonic regression (de Leeuw et al., 2009). Preferential contacts are defined as windows ($n = 21$) that show an increase over the background model. To identify regions that show a robust increase over the background, our algorithm contains the following steps: (1) to prevent spurious peak calling, we subsample 80% of the 4C data and recalculate the 4C scores to mitigate the effect of outliers (this process is repeated 1,000 times); (2) we only select windows where the average subsampled 4C score is 1.2-fold over the background in more than 99.5% of the iterations; and (3) we select only windows where the average 4C score is at least 50 reads/million over the background distribution. Peaks should be at least 5 kb in size and more than 40 kb from

the viewpoint. Within the peak region, we identify the window with the highest coverage, which we call the summit of the peak (see also Supplemental Information).

Genome Editing

Cells were transfected using Lipofectamine 2000 (Life Technologies) per manufacturer protocol; each transfection was performed using 4 μ g pSpCas9(BB)-2A-Puro (PX459, a gift from Feng Zhang [Addgene plasmid #48139]) and 4 μ g ssODN repair template (IDT). After 12–14 hr, cells were replated at low density. Twenty-four hours later, cells were placed under puromycin selection (1.8 μ g/ml) for 48 hr. Individual colonies were picked and seeded in 96-well plates. For details, gRNAs, and repair templates, see Supplemental Information.

Expression Analysis

Each qPCR was performed in triplicate per PCR plate, and all experiments were repeated two more times such that the average of nine measurements represents the expression value of a given clone. For expression analysis at the *Malt1* locus, expression values from two wild-type clones, two del/del clones, and the one inv/wt clone were compared. For *Sox2* expression analysis, expression values from five wild-type clones, five del/del clones, and five inv/inv clones were compared. For expression analysis at the *Fbn2* locus, expression values from six wild-type clones, five del/del clones, one inv/inv clone, and one inv/del clone were compared.

ACCESSION NUMBERS

4C and ChIP-seq data have been deposited under accession number GEO: GSE72720.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and four figures and can be found with this article online at <http://dx.doi.org/10.1016/j.molcel.2015.09.023>.

AUTHOR CONTRIBUTIONS

E.d.W. computationally analyzed data, helped design experiments, and helped write the manuscript; E.S.M.V. designed and carried out genome editing experiments and related 4C and helped write the manuscript; S.J.B.H. performed ChIP-seq and helped design and perform the large-scale 4C experiments; C.V.-Q. helped design the experiments and performed 4C, ChIP, and expression experiments; M.J.A.M.V. helped with experiments; H.T. performed 4C experiments; E.S. carried out work on β -globin locus; P.J.W. and P.H.L.K. helped design experiments; and W.d.L. designed experiments, supervised the project, and wrote the manuscript.

ACKNOWLEDGMENTS

We thank Harmen van de Werken and Geert Geeven for help and Michal Mokry for mapping CTCF ChIP-seq data. This work was supported by an NWO (Netherlands Organisation for Scientific Research)/CW TOP grant (714.012.002), an NWO Vici grant (724.012.003), a NanoNextNL grant, a European Research Council Starting Grant to W.d.L. (209700, "4C"), and an NWO Veni grant to E.d.W. (700.10.402).

Received: July 10, 2015

Revised: August 25, 2015

Accepted: September 25, 2015

Published: October 29, 2015

REFERENCES

Bickmore, W.A., and van Steensel, B. (2013). Genome architecture: domain organization of interphase chromosomes. *Cell* 152, 1270–1284.

- Chen, H., Tian, Y., Shu, W., Bo, X., and Wang, S. (2012). Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PLoS ONE* 7, e41374.
- de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 502, 499–506.
- de Leeuw, J., Hornik, K., and Mair, P. (2009). Isotone Optimization in R: Pool-Adjacent-Violators Algorithm (PAVA) and Active Set Methods. *J. Stat. Softw.* 32, 1–24.
- Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 149, 1233–1244.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380.
- Downen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schuijers, J., Lee, T.I., Zhao, K., and Young, R.A. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* 159, 374–387.
- Faure, A.J., Schmidt, D., Watt, S., Schwalie, P.C., Wilson, M.D., Xu, H., Ramsay, R.G., Odom, D.T., and Flicek, P. (2012). Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. *Genome Res.* 22, 2163–2175.
- Gómez-Marín, C., Tena, J.J., Acemel, R.D., López-Mayorga, M., Naranjo, S., de la Calle-Mustienes, E., Maeso, I., Beccari, L., Aneas, I., Vielmas, E., et al. (2015). Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proc. Natl. Acad. Sci. USA* 112, 7542–7547.
- Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* 162, 900–910.
- Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.G., and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* 460, 410–413.
- Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W., Ye, C., Ping, J.L., Mulawadi, F., et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.* 43, 630–638.
- Heath, H., Ribeiro de Almeida, C., Sleutels, F., Dingjan, G., van de Nobelen, S., Jonkers, I., Ling, K.W., Gribnau, J., Renkawitz, R., Grosveld, F., et al. (2008). CTCF regulates cell cycle progression of alphabeta T cells in the thymus. *EMBO J.* 27, 2839–2850.
- Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430–435.
- Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenko, V.V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 128, 1231–1245.
- Klein, F.A., Pakozdi, T., Anders, S., Ghavi-Helm, Y., Furlong, E.E., and Huber, W. (2015). FourCSeq: analysis of 4C sequencing data. *Bioinformatics* 31, 3085–3091.
- Li, Y., Rivera, C.M., Ishii, H., Jin, F., Selvaraj, S., Lee, A.Y., Dixon, J.R., and Ren, B. (2014). CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells. *PLoS ONE* 9, e114485.
- Nakahashi, H., Kwon, K.R., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruetz, N., Nelson, S., Yamane, A., et al. (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep.* 3, 1678–1689.
- Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). Transcription. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* 347, 1017–1021.
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385.
- Ong, C.T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* 15, 234–246.
- Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132, 422–433.
- Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* 153, 1281–1295.
- Pugacheva, E.M., Rivero-Hinojosa, S., Espinoza, C.A., Méndez-Catalá, C.F., Kang, S., Suzuki, T., Kosaka-Suzuki, N., Robinson, S., Nagarajan, V., Ye, Z., et al. (2015). Comparative analyses of CTCF and BORIS occupancies uncover two distinct classes of CTCF binding genomic regions. *Genome Biol.* 16, 161.
- Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308.
- Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680.
- Rippe, K., von Hippel, P.H., and Langowski, J. (1995). Action at a distance: DNA-looping and initiation of transcription. *Trends Biochem. Sci.* 20, 500–506.
- Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Gonçalves, A., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* 148, 335–348.
- Seitan, V.C., Faure, A.J., Zhan, Y., McCord, R.P., Lajoie, B.R., Ing-Simmons, E., Lenhard, B., Giorgetti, L., Heard, E., Fisher, A.G., et al. (2013). Cohesin-based chromatin interactions enable regulated gene expression within pre-existing architectural compartments. *Genome Res.* 23, 2066–2077.
- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* 148, 458–472.
- Sofueva, S., Yaffe, E., Chan, W.C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S.M., Schroth, G.P., Tanay, A., and Hadjur, S. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* 32, 3119–3129.
- Splinter, E., Heath, H., Kooren, J., Palstra, R.J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* 20, 2349–2354.
- Splinter, E., de Wit, E., van de Werken, H.J., Klous, P., and de Laat, W. (2012). Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods* 58, 221–230.
- Symmons, O., Uslu, V.V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., and Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* 24, 390–400.
- Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* 10, 1297–1309.
- Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirhuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 457, 796–801.
- Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and

mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307–319.

Yaffe, E., and Tanay, A. (2011). Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat. Genet.* 43, 1059–1065.

Zhou, H.Y., Katsman, Y., Dhaliwal, N.K., Davidson, S., Macpherson, N.N., Sakthidevi, M., Collura, F., and Mitchell, J.A. (2014). A Sox2 distal enhancer

cluster regulates embryonic stem cell differentiation potential. *Genes Dev.* 28, 2699–2711.

Zuin, J., Dixon, J.R., van der Reijden, M.I.J.A., Ye, Z., Kolovos, P., Brouwer, R.W.W., van de Corput, M.P.C., van de Werken, H.J.G., Knoch, T.A., van IJcken, W.F.J., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. USA* 111, 996–1001.