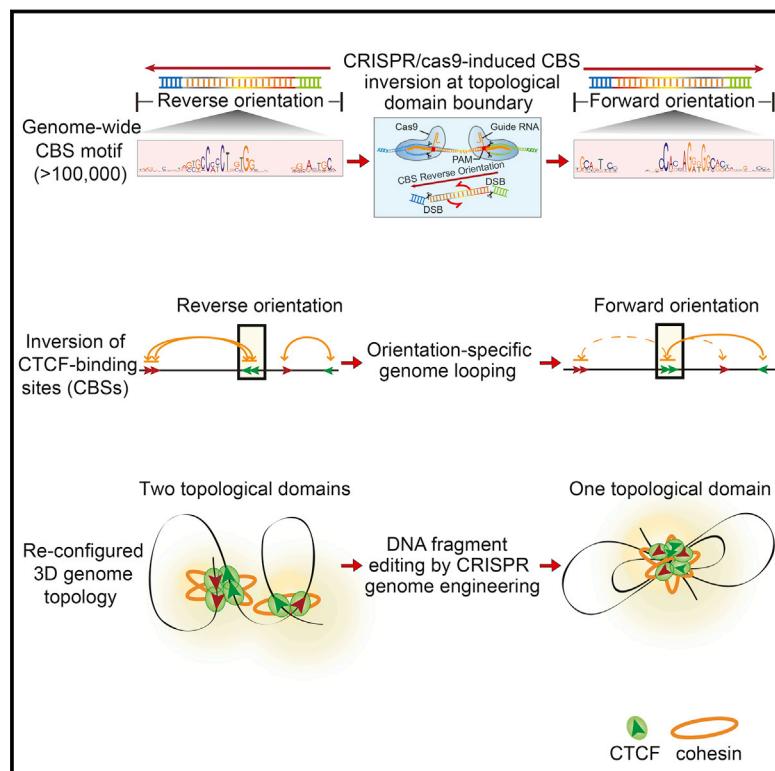


CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function

Graphical Abstract



Highlights

- The orientation of *Pcdh* CBSs determines the direction of topological DNA looping
- Directional CTCF binding to CBSs is crucial for loop topology and gene expression
- The CTCF binding orientation functions similarly in β -globin and the whole genome
- CTCF/cohesin-mediated directional DNA-looping determines chromosome architecture

Authors

Ya Guo, Quan Xu, Daniele Canzio, ..., Adrian R. Krainer, Tom Maniatis, Qiang Wu

Correspondence

tm2472@cumc.columbia.edu (T.M.), qwu123@gmail.com (Q.W.)

In Brief

The relative orientations of CTCF binding sites in enhancers and promoters determine the directionality of DNA looping and regulation of gene expression. The findings reveal how chromosome architecture is encoded by genome sequence and provoke thinking about how sequence orientation is functionally translated into a 3D genome.

Accession Numbers

GSE71275

GSE71072

CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function

Ya Guo,^{1,2,3,9} Quan Xu,^{1,2,3,9} Daniele Canzio,⁴ Jia Shou,^{1,2,3} Jinhuan Li,^{1,2,3} David U. Gorkin,⁵ Inkyung Jung,⁵ Haiyang Wu,^{1,2,3} Yanan Zhai,^{1,2,3} Yuanxiao Tang,^{1,2,3} Yichao Lu,^{1,2,3} Yonghu Wu,^{1,2,3} Zhiliang Jia,^{1,2,3} Wei Li,^{1,2,3} Michael Q. Zhang,^{6,7} Bing Ren,⁵ Adrian R. Krainer,⁸ Tom Maniatis,^{4,*} and Qiang Wu^{1,2,3,*}

¹Center for Comparative Biomedicine, MOE Key Laboratory of Systems Biomedicine, Institute of Systems Biomedicine, Collaborative Innovation Center of Systems Biomedicine, Shanghai Jiao Tong University (SJTU), Shanghai 200240, China

²State Key Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Renji Hospital, School of Medicine, SJTU, Shanghai 200240, China

³Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (MOE), Bio-X Center, School of Life Sciences and Biotechnology, SJTU, Shanghai 200240, China

⁴Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, 701 West 168th Street, New York, NY 10032, USA

⁵Ludwig Institute for Cancer Research and Department of Cellular and Molecular Medicine, University of California, San Diego School of Medicine, 9500 Gilman Drive, La Jolla, CA 92093, USA

⁶Department of Molecular and Cell Biology, Center for Systems Biology, University of Texas at Dallas, Richardson, TX 75080, USA

⁷MOE Key Laboratory of Bioinformatics and Bioinformatics Division, Center for Synthetic and System Biology, TNLIST/Department of Automation, Tsinghua University, Beijing 100084, China

⁸Cold Spring Harbor Laboratory, 1 Bungtown Rd, NY 11724, USA

*Co-first author

*Correspondence: tm2472@cumc.columbia.edu (T.M.), qwu123@gmail.com (Q.W.)

<http://dx.doi.org/10.1016/j.cell.2015.07.038>

SUMMARY

CTCF and the associated cohesin complex play a central role in insulator function and higher-order chromatin organization of mammalian genomes. Recent studies identified a correlation between the orientation of CTCF-binding sites (CBSs) and chromatin loops. To test the functional significance of this observation, we combined CRISPR/Cas9-based genomic-DNA-fragment editing with chromosome-conformation-capture experiments to show that the location and relative orientations of CBSs determine the specificity of long-range chromatin looping in mammalian genomes, using protocadherin (*Pcdh*) and β -globin as model genes. Inversion of CBS elements within the *Pcdh* enhancer reconfigures the topology of chromatin loops between the distal enhancer and target promoters and alters gene-expression patterns. Thus, although enhancers can function in an orientation-independent manner in reporter assays, in the native chromosome context, the orientation of at least some enhancers carrying CBSs can determine both the architecture of topological chromatin domains and enhancer/promoter specificity. These findings reveal how 3D chromosome architecture can be encoded by linear genome sequences.

INTRODUCTION

Interphase chromosomes fold into highly compartmentalized, hierarchical structures, and the topology of chromosome folding is

thought to play an important role in critical nuclear processes, including the regulation of gene expression (de Laat and Du-boule, 2013; Gibcus and Dekker, 2013; Levine et al., 2014). Individual chromosomes occupy a distinct space in the nucleus, referred to as a “chromosome territory” (Cremer and Cremer, 2001), and within this region are relatively stable chromatin domains containing specific DNA-looping interactions between proximal promoters and distal regulatory DNA elements, such as transcriptional enhancers and silencers, insulators, and locus control regions (LCR) (Dixon et al., 2012; Gibcus and Dekker, 2013; Levine et al., 2014; Lieberman-Aiden et al., 2009). Genome-wide studies of mammalian genomes have shown that there are far more enhancers than promoters and that spatiotemporal gene expression is regulated through one or more promoters and multiple enhancers (Bulger and Groudine, 2011; ENCODE Project Consortium, 2012; Zhang et al., 2004). Insulator elements play pivotal roles in orchestrating proper long-range DNA-looping interactions between remote enhancers and their cognate promoters via mechanisms that are poorly understood (Dowen et al., 2014; Jia et al., 2014; Narendra et al., 2015; Ong and Corces, 2014).

The mammalian CCCTC-binding factor (CTCF), a zinc-finger DNA-binding protein, is the best characterized insulator-binding protein, which also plays a key role in genome looping (Lobanenkov et al., 1990; Ong and Corces, 2014). In addition, the insulator activity of CTCF-binding sites (CBSs) requires the cohesin complex that is recruited by CTCF. Previous studies have implicated CTCF and cohesin complexes in genome-wide chromatin-looping interactions (Handoko et al., 2011; Zuin et al., 2014). Over 100,000 diverse CBSs have been identified in mammalian genomes (Kim et al., 2007; Shen et al., 2012; Xie et al., 2007), and the genome-wide pattern of CTCF occupancy is cell-type specific (Kim et al., 2007; Shen et al., 2012; Wang

et al., 2012); however, CBSs are enriched at constitutive boundaries of topologically associated domains (TADs) (Dixon et al., 2012; Gibcus and Dekker, 2013; Zuin et al., 2014). More recently, it was shown that CBSs at the anchors of chromatin loops are arranged in the forward-reverse orientations, suggesting that the relative positions and orientations of CBSs could be important for chromosome architecture (Alt et al., 2013; Guo et al., 2012; Monahan et al., 2012; Rao et al., 2014; Vietri Rudan et al., 2015). However, the underlying molecular mechanisms through which CTCF-mediated DNA-looping interactions lead to CTCF's many cellular functions remain obscure.

The mammalian protocadherin (*Pcdh*) α , β , and γ gene clusters provide a unique model system to investigate the role of CTCF/cohesin-mediated enhancer-promoter interactions in cell-specific gene expression (Guo et al., 2012; Hirayama et al., 2012; Monahan et al., 2012; Wu and Maniatis, 1999). In the α and γ (but not the β) clusters, the *Pcdh* "variable regions" contain more than a dozen large and highly similar "alternately expressed" variable exons followed by two or three "ubiquitously expressed" C-type variable exons in the α and γ clusters, respectively (Figure 1A). By contrast, the downstream "constant regions" of the α and γ clusters are organized into three small exons that encode the intracellular domain of all of the protein isoforms in each cluster (Figure 1A) (Wu and Maniatis, 1999). Previous studies revealed that each "variable" exon (except $\alpha c2$, $\beta 1$, $\gamma c4$, and $\gamma c5$) is preceded by a promoter containing a highly conserved sequence element (CSE) (Figure 1A) (Tasic et al., 2002; Wu et al., 2001). Subsequently, CTCF was shown to bind to the CSE and to a second CBS within the downstream exon (eCBS) of transcriptionally active α genes, and this binding is required for transcription (Guo et al., 2012; Hirayama et al., 2012; Monahan et al., 2012; Wu et al., 2001).

A key observation, relevant to the present study, was that the CBSs in the *Pcdh* HS5-1 enhancer downstream of the α cluster and in each of the promoters and downstream exons are configured in opposite orientations (Guo et al., 2012). Chromosome conformation capture (3C) studies revealed that CTCF/cohesin-mediated DNA looping occurs exclusively between paired CBSs within the enhancers and the active promoters (Guo et al., 2012). This organization of CBS sites within the *Pcdh* clusters was recently shown to reflect the genome-wide organization of CBS pairs at anchors of DNA contact loops (Alt et al., 2013; Rao et al., 2014; Vietri Rudan et al., 2015). This striking organization of oriented CBS sites in *Pcdh* clusters and the availability of powerful CRISPR genome editing methods provide the opportunity to address the functional significance of the genome-wide CBS organization.

Here, we provide direct functional evidence that the location and relative orientations of CBSs play a critical role in the establishment of chromosome architecture and proper enhancer-promoter interactions. We developed a CRISPR/Cas9-based DNA-fragment *in situ* inversion technology (Li et al., 2015)—in conjunction with 3C, as well as related 4C (circularized 3C) and Hi-C methods (Dekker et al., 2002)—to study the chromatin organization in the *Pcdh* clusters. We find that directional CTCF binding to the paired CBSs with a specific combination of forward-reverse orientations determines the formation of specific DNA-looping interactions between enhancers and promoters in

mammalian cells. The generality of this observation is demonstrated by showing that the same mechanism operates with CBSs in the β -globin gene cluster and throughout mammalian genomes. This mechanism of CTCF-determined looping directions has important implications regarding chromosomal architecture and the insulator functions of genome-wide CBSs in genome folding and gene regulation.

RESULTS

Two CTCF/Cohesin-Mediated Chromatin Domains in the *Pcdh* Locus

We used 3C, 4C, and Hi-C to study CTCF/cohesin-mediated DNA looping and chromatin organization in the *Pcdh* α , β , and γ clusters (Figures 1A–1G and S1). Specifically, we performed 4C using the HS5-1 enhancer, the α promoters, or the region immediately upstream of the α cluster as anchors and showed that promoters within the α cluster interact with HS5-1 in human SK-N-SH cells (Figure 1B) and mouse neuro2A (N2A) cells (Figure 1C) and brain tissues (Figures 1D, S1A, and S1B). By contrast, the downstream promoters of the β cluster display virtually no interactions with HS5-1 (Figures 1B–1D).

We also performed 4C using promoters of the γ cluster as anchors and identified a downstream regulatory region (Figures S1C–S1E). This region contains a cluster of CBSs (CBS sites a–h) located within several DNaseI HS sites (see Figures 1A and S1F) and is enriched with molecular marks typical of enhancers (Figure S1G) (ENCODE Project Consortium, 2012). Similar to the α cluster, this downstream regulatory region interacts with promoters of the γ cluster in human SK-N-SH cells (Figure 1E), as well as in mouse N2A cells (Figure 1F) and brain tissues (Figure 1G). Interestingly, promoters of the β cluster also interacted with this remote enhancer when the β promoters were used as 4C anchors (Figures S1H and S1I), which is consistent with the previous observation that this region is required for maximum levels of β gene expression and regulates the γ cluster (Yokota et al., 2011). Finally, we performed Hi-C experiments on SK-N-SH and analyzed the results along with previously published Hi-C data from H1-hESC and NPC cells (Dixon et al., 2012, 2015) (Figure S1J). We observed two TAD-like chromatin domains covering α and $\beta\gamma$ clusters, respectively, by calculating a directionality index with a sliding window of 300 kb (Figure S1J). Taken together, these data show that the HS5-1 enhancer forms a CTCF/cohesin-mediated chromatin domain (CCD) within the α cluster, and the $\beta\gamma$ -regulatory region forms a CCD that includes both the β and γ clusters (Figure 1A).

Non-random CBS Orientations in the Two *Pcdh* CCDs

A CBS motif is located within all of the *Pcdh* α , β , and γ promoters, except $\alpha c2$, $\beta 1$, $\gamma c4$, and $\gamma c5$ (Figure 1A) (Guo et al., 2012; Kim et al., 2007; Monahan et al., 2012; Nakahashi et al., 2013; Rhee and Pugh, 2011; Schmidt et al., 2012; Wu et al., 2001). We defined the CBS from modules 1 to 4 as being in the forward orientation (Figure 1H). Interestingly, all of the α CSEs and eCBSs are in the forward orientation; by contrast, both HS5-1 CBSs (HS5-1a and b) are in the reverse orientation within the α CCD (Figure 1A). Similarly, all of the $\beta\gamma$ CSEs are in the forward orientation, whereas the first five CBSs (a–e) in the $\beta\gamma$

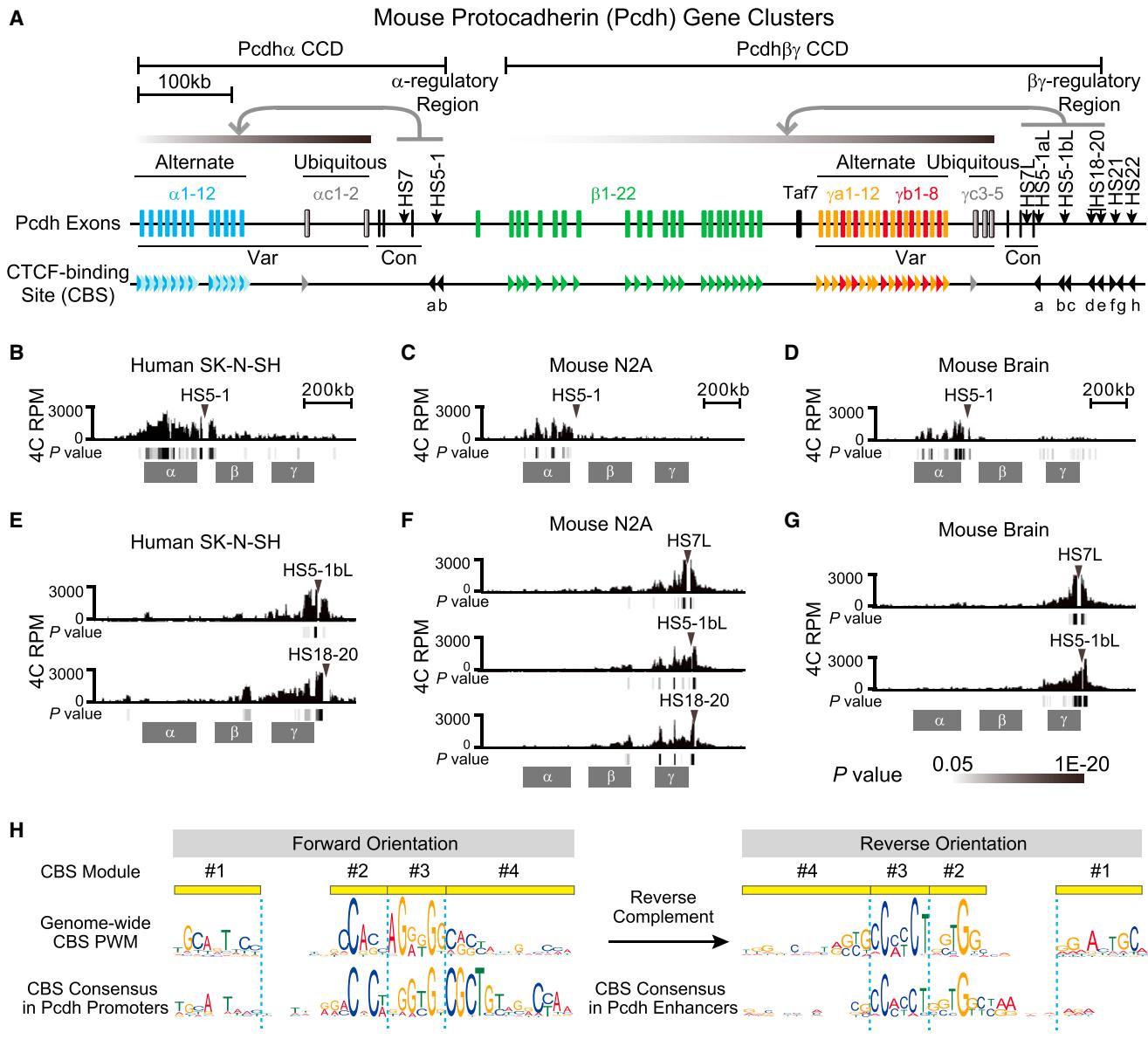


Figure 1. Two Distinct CTCF/Cohesin-Mediated Chromatin Domains in the Pcdh Locus

(A) Diagram showing the *Pcdh* α and $\beta\gamma$ CCDs in the three mouse *Pcdh* gene clusters. The variable (Var) and constant (Con) exons are also indicated. The CBSs and their orientations are indicated as arrowheads. Different types of *Pcdh* CBSs are represented by differently colored arrowheads. The dark and light blue CBSs represent the CSE and eCBS, respectively, for each of the 12 “alternate promoters” ($\alpha 1-12$) of the α cluster. The 21 tandem green arrowheads represent the CSE for each member of the β cluster (except $\beta 1$). The yellow and red arrowheads represent CSEs for γa and γb , respectively. The two gray arrowheads represent the C-type CSEs ($\alpha c 1$ and $\gamma c 3$). The two CBS sites (a and b) downstream of the α cluster and the eight CBS sites (a-h) downstream of the γ cluster are indicated in black arrowheads. The DNasel hypersensitive sites (HS) in the α and $\beta\gamma$ regulatory regions are also shown.

(B-G) Relative distributions of the 4C reads per million (RPM) obtained in human SK-N-SH cells (B), mouse N2A cells (C), and brain tissues (D) using the HS5-1 enhancer as an anchor. 4C interaction profiles in human SK-N-SH cells (E), mouse N2A cells (F), and brain tissues (G) with the regulatory region downstream of the γ cluster as an anchor are also shown. The significance of interactions (p value) is shown under the reads density for each panel.

(H) Showing the forward orientation of CBS sites in *Pcdh* promoters and reverse orientation of CBS sites in *Pcdh* enhancers.

See also Figure S1.

enhancer complex are in the reverse orientation within the $\beta\gamma$ CCD (Figures 1A and S1F). The last three CBSs (f-h) downstream of the $\beta\gamma$ -regulatory region are in different orientations, and they do not interact with the $\beta\gamma$ promoters (see Figures 1A and S1C-S1F). Thus, the *Pcdh* chromatin-looping interactions occur

between CBS pairs in the forward-reverse orientations in the promoters and enhancers, respectively (Figure 1A). Previously reported weak DNA-looping interactions between two CBSs in the same orientation in the α promoter region may be the consequence of their interactions with common CBSs within the HS5-1

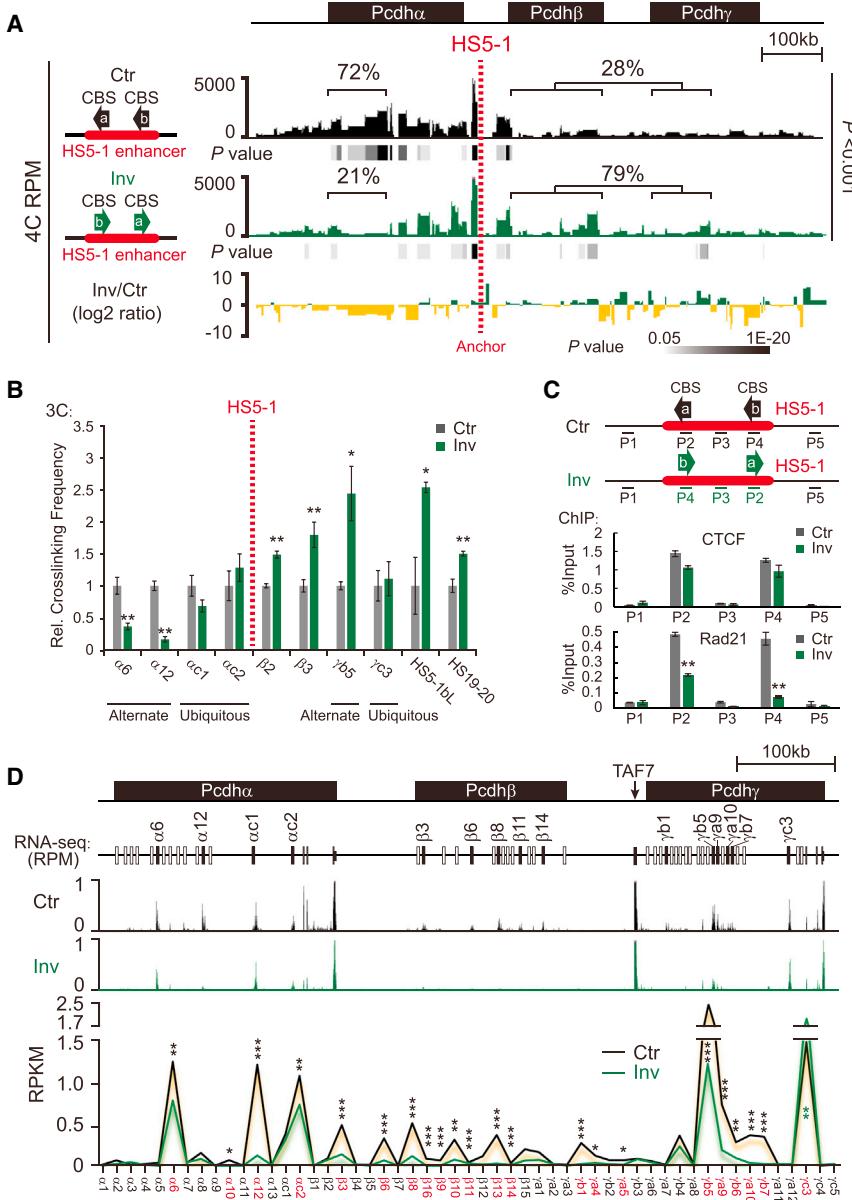


Figure 2. Inversion of the *Pcdh* HS5-1 Enhancer with CBSs Switches DNA Looping Direction and Alters Gene Expression

(A) Long-range chromatin-looping interaction profiles of the HS5-1 anchor in wild-type control (Ctr) or in a HS5-1 inversion (Inv) cell line generated from subcloned HEC-1B cells by CRISPR engineering. The log2 ratio between inversion and control is also shown.

(B) The relative crosslinking frequency measured by quantitative 3C assays in the control or inversion cell lines with HS5-1 as an anchor (HS5-1 is within the same 3C restriction fragment in the genomes of both Ctr and Inv cell lines). Data are means \pm SEM ($n = 4$). * $p < 0.05$ and ** $p < 0.01$.

(C) Control experiments showing functional CTCF/cohesin binding after inversion. Data are means \pm SEM ($n = 3$); ** $p < 0.01$.

(D) RNA-seq experiments showing expression reduction of the α , β , and γ clusters (except $\gamma c3$) after inversion. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

See also Figure S2.

enhancer in the opposite orientation (Guo et al., 2012). Overall, these observations strongly suggest that the relative orientations of CBSs determine the topology of CTCF/cohesin-mediated DNA looping (Alt et al., 2013; Guo et al., 2012; Rao et al., 2014; Vietri Rudan et al., 2015).

In Situ Inversion of the Boundary CBS Element Alters DNA Looping and Gene Expression

To directly determine whether CBS orientation is important for enhancer-promoter interactions and DNA looping, we used an efficient in situ CRISPR inversion of DNA fragment editing method we recently developed (Li et al., 2015) to invert the core HS5-1 element in its endogenous chromosomal location. We screened for CRISPR inversion cell clones derived from HEC-1B cells, which have three alleles at the *Pcdh* locus (Li

et al., 2015) and express a subset of the α (Tasic et al., 2002) and γ clusters (Figure S1F, also see Figure 2D, below). Out of 32 clones that were genotyped, we identified a cell clone (V28) in which the orientation of HS5-1 was inverted for two alleles and deleted for one allele (Figure S2A). We then performed 4C using HS5-1 as an anchor. Strikingly, we observed a significant increase in DNA-looping interactions between HS5-1 and promoters in the $\beta\gamma$ clusters (from 28% to 79%) and a corresponding decrease in DNA-looping interactions with the promoters driving the expression of the alternate *Pcdh* α isoforms (from 72% to 21%) (Figure 2A). We confirmed these changes in DNA looping by quantitative 3C assays (Figure 2B). Chromatin immunoprecipitation (ChIP)-qPCR studies showed that

CTCF binds to the inverted HS5-1 element; however, a significant decrease in the binding of the cohesin subunit Rad21 to this sequence was observed (Figure 2C). We conclude that inversion of the oriented CBSs in the HS5-1 enhancer profoundly alters enhancer-promoter interactions in the *Pcdh* clusters.

To assess the effects of these alterations, we carried out an RNA sequencing (RNA-seq) analysis on the HEC-1B cells in which the HS5-1 enhancer is inverted. As shown in Figure 2D, the decrease of DNA looping between HS5-1 and α promoters resulted in a significant reduction in α transcription. However, a corresponding enhancement in β transcription was not observed in spite of the observed increase of interactions between the inverted HS5-1 enhancer and the β cluster. Similarly, a reduction of γ transcription (except an increase of $\gamma c3$) was also observed

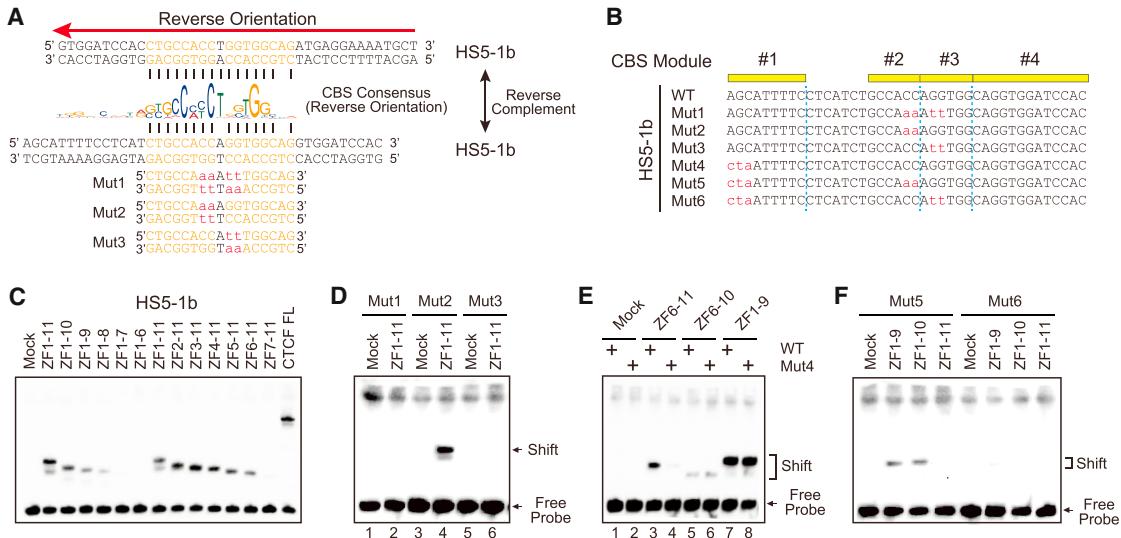


Figure 3. CTCF Recognition of the HS5-1b Site in Only One Direction

(A) Showing the HS5-1b CBS sequence (double-stranded) of the reverse orientation (indicated above by a red arrow) with the palindromic core highlighted. The double-stranded reverse complement HS5-1b CBS sequences (along with three probes with core sequences mutated) are also shown below the CBS consensus. The nucleotides that match to the CBS consensus are indicated by vertical lines. Note that mut2 and mut3 are exactly the same for the palindrome core sequence.

(B) The wild-type (WT) or mutant (Mut) sequences of *HS5-1b* probes (shown in the reverse complement).

(C) Gel-shift assays of the wild-type HS5-1b probe using a set of recombinant CTCF proteins with sequentially deleted zinc-finger domains.

(D-F) Gel-shift assays using recombinant CTCF proteins with probes of Mut1-3 (D), Mut4 (E), Mut5 and Mut6 (F).

See also Figure S3.

(Figure 2D). Thus, the inappropriate engagement of the HS5-1 enhancer with the downstream β and γ clusters appears to disrupt rather than enhance transcription.

The function of enhancers tested in mammalian cell transfection experiments with reporter genes is independent of the relative orientations of the enhancer and promoter (Banerji et al., 1981). However, the data of Figures 1 and 2 clearly show that the activity and specificity of enhancers in their normal chromosomal context are highly orientation specific, likely as a consequence of differences in the altered organization of CCDs caused by the DNA sequence inversion.

Directional CTCF Binding to *Pcdh* CBS Sequences

A large number of palindromic CBSs have been identified in the human genome (Xie et al., 2007), and yet, intriguingly, CTCF binds to CBSs in a preferred orientation (Nakahashi et al., 2013; Renda et al., 2007; Schmidt et al., 2012). How CTCF binds directionally to large numbers of diverse and seemingly palindromic CBSs therefore remains a mystery. Careful examination of the 17 bp core sequences of the *HS5-1b* CBS revealed that they are perfectly palindromic (Figure 3A). Considering that the reverse-complement sequences also conform to the CTCF-binding consensus, one would expect that CTCF recognizes *HS5-1b* in both directions, thus eliminating the apparent asymmetry of CBS pairs in the α promoters and the *HS5-1* enhancer. To investigate whether CTCF binding to the *HS5-1b* CBS is directional, we generated three DNA probes bearing combined 2 bp mutations designed to distinguish between the two putative CTCF-binding directions (Figures 3A and 3B). We

also generated a series of 17 CTCF expression constructs encoding two sets of truncated CTCFs in which each zinc finger (ZF) domain was sequentially deleted from either the C or N terminus (**Figures 3C** and **S3A**).

Remarkably, electrophoretic mobility shift assay (EMSA) experiments revealed that CTCF recognizes palindromic *HS5-1b* in only one direction relative to its sequences because mutation of “GG” to “tt” (mut1 and mut3, **Figures 3A** and **3B**) abolished CTCF binding (lanes 2 and 6, **Figure 3D**) whereas mutation of “CC” to “aa” (mut2, **Figures 3A** and **3B**) did not abolish CTCF binding (lane 4, **Figure 3D**). To further investigate the directional CTCF recognition, we generated combinations of these mutations with 3 bp mutations in the *HS5-1b* module 1 (mut5 and mut6 with mut4 as the control, **Figure 3B**). We found that the first three nucleotides of module 1 are recognized by the C-terminal ZF11, and this recognition determines the direction of CTCF binding to the CBS with palindromic core sequences (**Figures 3E** and **3F**). In particular, introduction of mutations into the first tri-nucleotide from “AGC” to “cta” did not alter the binding to ZF6-10 (compare lanes 6 and 5 in **Figure 3E**) but did reduce the binding of ZF6-11 to levels similar to those of ZF6-10 (compare lanes 4 and 3 with lanes 5 and 6 in **Figure 3E**). Thus, the C-terminal ZF11 of CTCF determines its directional binding to the *HS5-1b* CBS with palindromic core sequences, suggesting that module 1 is the key directional element in CBSs with palindromic core sequences.

To further determine the directionality of CTCF binding at the *Pcdh* CBS repertoire and the recognition profile of the 11 ZF domains of CTCF, we mutated distinct sets of 3 bp sequences in

modules 1, 2, or 4 of a large set of *Pcdh* CBSs (Figure S3B). We found that the C-terminal ZF domains of CTCF recognize module 1 of the CBS and that the N-terminal ZF domains recognize module 4 (Figures S3C–S3W). For example, CTCF ZF3 and ZF2 recognize the CGC and TGT tri-nucleotides of the α 8 CBS, respectively, because mutations of these tri-nucleotides reduced CTCF binding only when ZF3 and ZF2 were present (Figures S3B–S3F). In addition, module 2 of the β 3 CBS appears to be bound by CTCF ZF6/7 (Figures S3B and S3G–S3I). Moreover, CTCF ZF2-11 and ZF4-11 are essential for binding the CSE of γ a10 (Figures S3B, S3J, S3K, and S3N) and γ b7 (Figures S3B, S3L, and S3N), respectively. In particular, ZF11 of CTCF is absolutely required for CTCF binding of the CSE of γ a10 and γ b7, as deletion of ZF11 abolished CTCF binding to these two CBSs (Figures S3J and S3L). Furthermore, CTCF ZF11 recognizes the first tri-nucleotide TGC in module 1 of the $\beta\gamma$ -*b* CBS (Figures S3B, S3M, and S3N). Finally, we show that different types of *Pcdh* CBSs are recognized by distinct combinations of the CTCF ZF domains (Figures S3O–S3W).

Taken together, these observations clearly show that CTCF recognizes CBSs in only one direction relative to its target sequences and that distinct combinations of CTCF ZF domains recognize different types of *Pcdh* CBSs. Thus, the configuration of directional CTCF binding determines the topology of CTCF/cohesin-mediated DNA looping in the *Pcdh* clusters. Although the nature of the interactions between CTCF/cohesin complexes on the active *Pcdh* alternate promoters and the *HS5-1* enhancer is not known, these observations suggest that functional interactions require directional binding of CTCF/cohesin in the forward-reverse orientations to the *Pcdh* CBS pairs.

Directional CTCF Binding in Genome-wide DNA Looping Specificity

Recent whole-genome Hi-C experiments revealed that the vast majority of DNA loops correlate with the presence of CBS pairs arranged in a convergent orientation (Rao et al., 2014; Vietri Rudan et al., 2015). However, because chromatin contacts detected by Hi-C are unbiased and do not specifically relate to CTCF/cohesin binding, these loops may or may not be established by CTCF and the associated cohesin complex. Based on our observations that directional CTCF binding to forward-reverse CBS pairs determines topological looping domains in the *Pcdh* clusters, we investigated genome-wide CBS orientation and CTCF/cohesin-mediated DNA-looping topology by analyzing published datasets of chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) and ChIP-seq with specific CTCF/cohesin antibodies (ENCODE Project Consortium, 2012; Handoko et al., 2011). We first determined the orientations of 88,332 CBSs and their CTCF occupancies in K562 cells (Table S1) using position weight matrices (PWM) (Schmidt et al., 2012). We then screened for ChIA-PET interactions (from a total of 24,887) in which both tethered DNA fragments contain CBSs and identified 19,532 such interactions (Figure 4A and Table S2). We found that 76.4% of the CTCF-mediated interactions (14,928) are in the forward-reverse orientations; by contrast, only 2.3% (443) are in the reverse-forward orientations. In addition, 11.0% of the interactions (2,155) are in the forward-forward orientations and 10.3% (2,006) are in the reverse-reverse orienta-

tions. Finally, we measured the chromatin-looping strength by counting the number of overlapped looping PETs of the ChIA-PET datasets. Interestingly, the percentages of CBS pairs in the forward-reverse orientations dramatically increased with enhanced chromatin-looping strength (Figure 4B and Table S3). Similar results were observed in data collected from mouse E14 embryonic stem cells (Table S2) (Handoko et al., 2011) and human MCF-7 breast cancer cells (Tables S1 and S2). These observations clearly show that the majority of genome-wide chromatin-looping interactions correlate with directional CTCF binding to CBS pairs in the forward-reverse orientations.

We previously demonstrated that CTCF and the cohesin complex colocalize to promoters and enhancers in the *Pcdh* clusters (Monahan et al., 2012; Guo et al., 2012). In order to investigate the relationship between the binding of CTCF and cohesin, CBS orientation, and DNA looping, we identified chromatin-looping interactions containing CTCF/cohesin co-occupied CBSs in K562 cells. We found 16,610 such interactions, in which the majority (78.7%) occur between CBS pairs in the forward-reverse orientations (Table S4). In addition, we found that the tethered CBSs have a higher occupancy of CTCF and cohesin than the non-tethered CBSs (Figure S4A), suggesting that high levels of CTCF/cohesin co-occupancy at CBSs are required for establishing these long-range chromatin-looping interactions. Thus, in addition to the location and orientation of CBSs, levels of their CTCF/cohesin occupancy are also an important determinant for directional chromatin looping.

Interestingly, in the K562 cell genome, 46% of the p300 enhancer marks (Heintzman et al., 2007) have at least one CBS located within 2 kb (Figure S4B). On the other hand, 54% of the marks of the silencer factor REST/NRSF (Johnson et al., 2007) have at least one CBS located within 2 kb (Figure S4C). These observations suggest that CTCF/cohesin-mediated DNA-looping interaction may enhance or inhibit gene expression, depending on the proximity of the CBS to p300 or REST/NRSF. This possibility is consistent with the observation that a REST/NRSF binding site in *HS5-1* is required for repression of the α cluster in non-neuronal cells (Kehayova et al., 2011).

We next identified genome-wide overlapping CTCF/cohesin-mediated chromatin-looping interactions and merged clusters of the overlapping interactions as single CCDs. The two CCDs in the *HoxD* locus are shown as examples in Figure 4C. The cumulative features of CBSs in the looping PETs of all human CCDs demonstrate that most CBSs are located near the boundaries (Figure 4D and Tables S4 and S5). By analyzing the orientations of the boundary CBS pairs between neighboring CCDs, we found that the vast majority (90.0%) of the boundary CBS pairs between neighboring CCDs in K562 cells are in the reverse-forward orientations (1,626) (Figure 4E and Tables S4 and S6). Similar results were obtained for MCF-7 cells (Tables S4, S5, and S6). Taken together, these genome-wide data suggest that directional CTCF binding to CBS pairs in the reverse-forward orientations at the boundary between neighboring CCDs is important for establishing distinct topological domains.

Finally, because CBSs are enriched at the boundaries of TADs (Dixon et al., 2012; Shen et al., 2012), we analyzed the orientations of CBSs of the TAD boundaries identified in H1-hESC and IMR90 cells. We found that CBS pairs in the reverse-forward

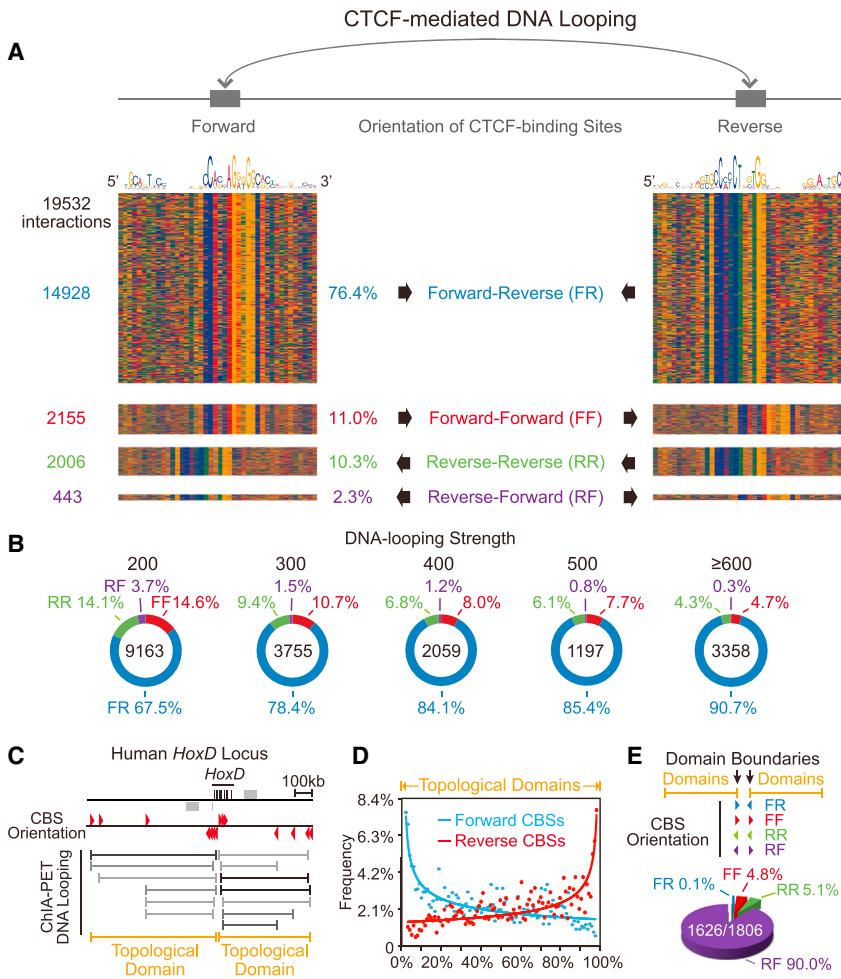


Figure 4. The Role of CBS Location and Orientation in CTCF-Mediated Genome-wide DNA Looping

(A) Diagram of CTCF-mediated long-range chromatin-looping interactions between CBS pairs in the forward-reverse orientations. The color charts represent 19,532 interactions of CBS pairs in K562 cells. The number and percentage of CBS pairs in the forward-reverse (FR), forward-forward (FF), reverse-reverse (RR), and reverse-forward (RF) orientations are shown.

(B) The percentage of CBS pairs in the forward-reverse orientations increases from 67.5% to 90.7% as the chromatin-looping strength is enhanced.

(C) Schematic of the two topological domains in the *HoxD* locus. The orientations of CBSs are indicated by arrowheads. CTCF/cohesin-mediated looping interactions and the two resulting topological domains (CCDs) are also shown.

(D) Cumulative patterns of CBS orientations of topological domains in the human genome.

(E) Distribution of genome-wide orientation configurations of CBS pairs located in the boundaries between two neighboring domains in the human K562 genome. Note that the vast majority (90.0%) of boundary CBS pairs between two neighboring domains are in the reverse-forward orientation.

See also Figure S4 and Tables S1, S2, S3, S4, S5, and S6.

orientation exist in >60% neighboring TAD boundaries (Figure S4D), suggesting that the boundary reverse-forward CBS pairs play an important role in the formation of most of TADs. For example, there is a CBS pair in the reverse-forward orientation in a Chr12 genomic region of H1-hESC cells, located at or very close to each of the six TAD boundaries (boundaries 1–6), except for boundary 5, which has only one closely located CBS in the forward orientation (Figure S4E). These data, taken together, strongly suggest that directional binding of CTCF to boundary CBS pairs in the reverse-forward orientations causes opposite topological looping and thus appears to function as insulators.

The Human β -globin Locus Provides an Additional Example of CBS Orientation-Dependent Topological Chromatin Looping

Based on the location and orientation of CBSs, as well as their CTCF/cohesin occupancy, we identified four CCDs (domains 1–4) in the well-characterized β -globin cluster (Figure 5A). The β -globin gene cluster is located between CBS3 (5'HS5) and CBS4 (3'HS1) in domain1 (Figure 5A) (Hou et al., 2010; Splinter et al., 2006). We generated a series of CBS4/5 mutant K562 cell lines using CRISPR/Cas9 with one or two sgRNAs (Li et al.,

2015) (Figures S2B and S2C). In the CRISPR cell lines D3, D7, and D19 (out of 38 clones screened) in which the internal CBS4 (3'HS1) was deleted (Figure S2B), chromatin-looping interactions between CBS3 (5'HS5) in the forward orientation and the boundary CBS5 in the reverse orientation in domain1 persisted, although its interaction with the CBS4 (3'HS1) region was abolished (Figures S5A and S5B). As expected, the interactions between CBS6/7 and CBS8/9 in domain2 were unchanged (Figure S5C). Strikingly, however, in the CBS4 (3'HS1) and CBS5 double-knockout CRISPR cell lines C2, C4, and C14 (out of 49 clones screened) (Figure S2C), novel chromatin-looping interactions between CBS3 (5'HS5) in the forward orientation of domain1 and CBS8/9 in the reverse orientation of the neighboring domain2 were observed, suggesting that these two domains merge as a single domain in CRISPR cell lines with CBS4/5 double knockout (Figure S5B). Similarly, when CBS8 was used as an anchor, this reverse-oriented CBS in domain2 establishes new long-range chromatin-looping interactions with CBS1–3 in the forward orientation of domain1 in the CBS4/5 double-deletion CRISPR cell lines (Figure S5C). We conclude that cross-domain interactions can be established after deletion of CBSs up to the boundary of topological domains, but not after deletion of the internal CBS in the β -globin locus.

To further test the functional significance of this organization of CBSs, we again performed CRISPR/cas9-mediated DNA-fragment editing in the HEK293T cells and screened 198 CRISPR

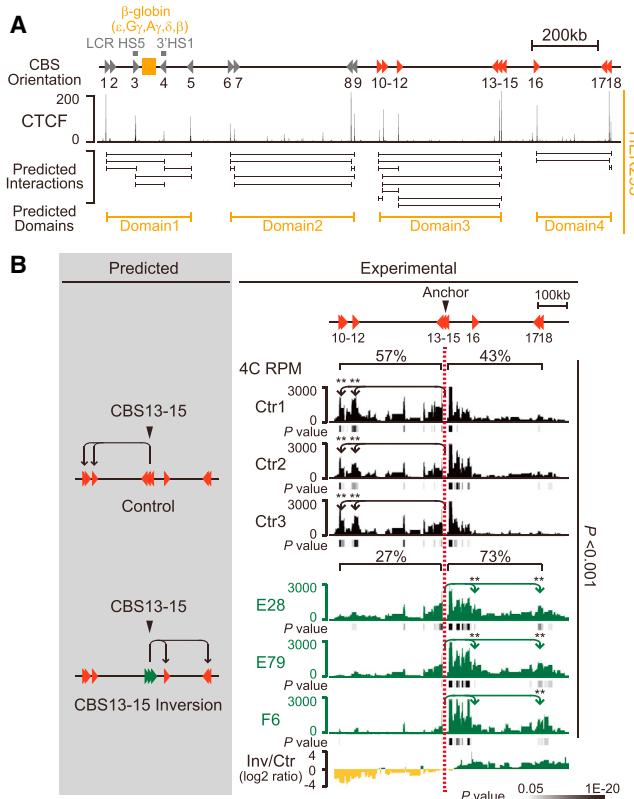


Figure 5. CRISPR Inversion of CBS13–15 in the Human β -globin Cluster Confirms the CTCF/Cohesin-Mediated Directional Looping Mechanism

(A) Diagram of the human β -globin region. Predicted looping interactions and topological domains are shown, based on CTCF occupancy in HEK293 cells. (B) The predicted interactions (left) and the altered looping directions (right) in the three subcloned CRISPR cell lines with inversion of CBS13–15 (E28, E79, and F6) are confirmed by 4C with CBS13–15 as an anchor. The looping interactions of three mock controls are also shown. The average log₂ ratios of interactions between inversions and controls are also indicated. **p < 0.01. See also Figures S2 and S5.

cell clones for inversions of CBS13–15, which is located at the boundary of domain3 (Figure 5A), and obtained three CRISPR inversion cell clones (E28, E79, and F6) (Figure S2D). We then performed 4C using CBS13–15 as an anchor. Strikingly and similar to the inversion of the *Pcdh* domain boundary, we observed a significant increase of chromatin-looping interactions with the downstream domain containing CBS16–18 (from 43% to 73%) and a corresponding decrease of chromatin-looping interactions with the upstream domain containing CBS10–15 (from 57% to 27%) (Figure 5B). These observations, taken together, clearly show that the orientations of CBSs determine the directionality of topological DNA looping.

In summary, we find that changing the relative orientations of CBS elements at domain boundaries by CRISPR/Cas9 alters the direction of CTCF/cohesin-mediated topological chromatin looping, which consequently leads to the establishment of new chromatin-looping interactions with CBS targets located in neighboring topological domains.

DISCUSSION

The diverse neuronal cell-surface PCDH repertoires, encoded by more than 50 clustered mammalian *Pcdh* genes, provide individual neurons with “identity tags” that engage in highly specific combinatorial homophilic interactions (Chen and Maniatis, 2013; Hirayama et al., 2012; Schreiner and Weiner, 2010; Thu et al., 2014; Wu, 2005; Wu and Maniatis, 1999). The functional significance of these interactions, based on direct evidence and by analogy to the *Dscam* system of invertebrates, is that they are required for the normal assembly of neural circuits during brain development (Chen et al., 2012; Garrett et al., 2012; Lefebvre et al., 2012; Suo et al., 2012; Thu et al., 2014; Wu and Maniatis, 1999). Therefore, understanding how PCDH diversity is generated in individual neurons is of fundamental importance.

The architecture rule of *Pcdh* CBSs provides interesting insights into their insulator functions. Rather than, or in addition to, blocking the cross-domain activities of enhancers as generally thought, the location and relative orientation of CBSs in enhancers determine the direction of looping and therefore indirectly “insulate” one expression domain from another. This perspective may explain seemingly contradictory data previously obtained from reporter gene assays or transgenic mice experiments that addressed whether insulators function in an orientation-dependent manner. The enhancer activity of HS5-1 was demonstrated both with reporter genes in transgenic mice (Ribich et al., 2006) and by targeted deletion (Kehayova et al., 2011). The presence of both oriented CBSs (Guo et al., 2012) and a functional REST/NRSF binding site (NRSE) in the HS5-1 enhancer regulates the directional looping and neuronal cell-specific activity of the enhancer (Guo et al., 2012; Kehayova et al., 2011). Analysis of HS5-1 reporter constructs revealed that the NRSE functions as a silencer in transfection experiments, and deletion of the HS5-1 enhancer in mice resulted in an increase in *Pcdh* gene expression in the kidney (Kehayova et al., 2011). The computational analyses presented here revealed that 46% of the potential enhancers genome wide have a nearby CBS (Figure S4B), and 54% of genome-wide REST/NRSF sites (Figure S4C) have a close CBS, suggesting that CTCF functions as an activator or a silencer of transcription by controlling directional looping in different genomic contexts or specific cell types.

Insulators function to ensure proper interactions between remote enhancers and cognate promoters *in vivo* by blocking enhancers from targeting non-cognate promoters (Ong and Corces, 2014). Considering that CTCF and cohesin play a pivotal role in the enhancer-blocking activity of insulators, in conjunction with the striking switching of looping directions with *in situ* CRISPR inversion, as well as biophysical and computational analyses, we propose that directional CTCF/cohesin recognition of CBS pairs in the forward-reverse orientations establishes topological domains in mammalian genomes, resulting in a boundary element with CBS pairs in the reverse-forward orientations between adjacent domains (Figure 6). In particular, the directional CTCF binding to forward-reverse CBS sites and the asymmetric recruitment of cohesin through the CTCF C-terminal domain (Xiao et al., 2011) determine

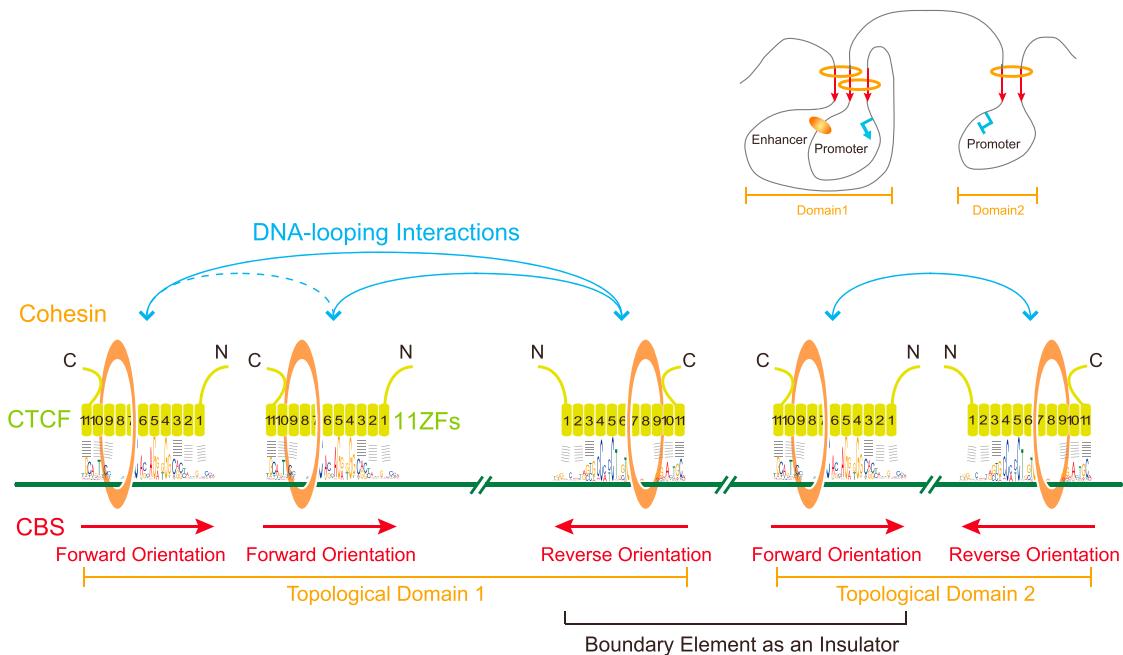


Figure 6. A Model of CTCF/Cohesin-Mediated Topological 3D Genome Folding and Gene Regulation

In mammalian genomes, CTCF directionally recognizes CBSs by distinct combinations of its 11 ZF domains and asymmetrically recruits the cohesin complex to CBS sites through its C-terminal domain (Xiao et al., 2011). CTCF, together with the cohesin complex, establishes specific long-range chromatin-looping interactions between CBS pairs in the forward-reverse orientations to form distinct topological domains (domains 1 and 2, see the upper right inset). The weak interactions between the two CBSs in the same forward orientation in topological domain 1 may be the consequence of their looping interactions with a common CBS in the reverse orientation (Guo et al., 2012). The two CBSs in the reverse-forward orientations form a boundary insulator element between the two neighboring domains 1 and 2, blocking remote enhancers located within one domain from aberrantly activating promoters located in the neighboring domain and thus “indirectly” ensuring proper activation of cognate promoters by distal enhancers within the same topological domain (see inset).

the looping direction with adjacent CBS sites (Figure 6). The reverse-forward boundary element between neighboring topological domains functions as an insulator to ensure the proper targeting of cognate promoters by distal enhancers (inset, Figure 6). This model provides a molecular explanation for the pivotal role of CTCF in organizing chromatin during higher-order chromosome folding and defines a unifying mechanism for the multivalent and seemingly conflicting functions of CTCF in the regulation of gene expression. Interestingly, computational simulation suggests that chromatin loops can “facilitate” or “insulate” enhancer-promoter interactions, depending on their locations outside or inside of the loops (Doyle et al., 2014). We note, however, that additional levels of control over directional DNA looping must exist, as the orientations of CBSs alone could not explain the specificity of DNA looping at such long distances since the chromatin fiber is likely to be sufficiently flexible to allow the DNA to be positioned to bring enhancers into proper orientation to interact with promoters.

Computational analyses reveal that the vast majority of genome-wide chromatin loops occurs between forward-reverse CBSs and a minority of loops occurs between forward-forward or reverse-reverse clusters of CBSs. Together with CRISPR and conformation capture evidence, the genome-wide architectural mechanism of CTCF/cohesin-mediated chromosome topology (Figure 6) has important implications not only for

long-distance chromatin-looping contacts ranging from several kb to several Mb but also for the enhancer insulation functions of insulators to ensure proper promoter activation by distal enhancers. We suggest that genome-wide topological chromatin looping can be predicted based on CTCF/cohesin directional binding and its controlling elements can be engineered by CRISPR genome editing. Thus, our findings reveal how nonlinear 3D genome topology could be encoded by linear 1D genomic sequences.

EXPERIMENTAL PROCEDURES

CRISPR/Cas9 System

The templates for producing target sgRNAs were constructed in pLKO.1 or pGL3-U6-sgRNA-PGK-Puro plasmids (Li et al., 2015). All constructs were confirmed by sequencing. To screen for inversion cell clones, cells cultured to about 80% confluence were transfected with Lipofectamine 2000 (Invitrogen) in a 6-well plate with 6 µg of plasmid DNA, including 2 µg of pcDNA3.1-Cas9 and 4 µg of sgRNA constructs (2 µg each). One day after transfection, puromycin was added to a final concentration of 2 µg/ml. Ten to twelve days later, the cells were serially diluted and plated in 96-well plates to isolate clonal CRISPR cell lines. The primer sets used are shown in Table S7.

Circularized Chromosome Conformation Capture

The circularized chromosome conformation capture (4C)-seq libraries were constructed as described (Guo et al., 2012; Jia et al., 2014). A series of

4C-seq libraries were generated by inverse PCR using a high-fidelity DNA polymerase. High-throughput sequencing was performed using 49 bp single-end reads on the Illumina HiSeq 2000 platform. The sequenced reads were mapped to reference genomes using the Bowtie program (version 1.0.0). The r3Cseq program in the R/Bioconductor package was used to detect statistical significance. All 4C-seq experiments were performed with at least two biological replicates.

Chromatin Immunoprecipitation

ChIP was performed as previously described (Guo et al., 2012; Jia et al., 2014). Briefly, HEK-1B cells were cross-linked with 1% formaldehyde for 10 min at 37°C. The lysate was immunoprecipitated with antibodies against CTCF (07-729; Millipore) or RAD21 (ab992; Abcam). The DNA was purified for real-time PCR. Statistical analysis was performed using Student's t test.

Electrophoretic Mobility Shift Assay

EMSA was performed using LightShift Chemiluminescent EMSA reagents as described (Guo et al., 2012). The probes were incubated with in-vitro-synthesized proteins in binding buffer containing 10 mM Tris, 50 mM KCl, 5 mM MgCl₂, 0.1 mM ZnSO₄, 1 mM dithiothreitol, 0.1% Nonidet P-40 (NP-40), 50 ng/μl poly (dI-dC), and 2.5% (v/v) glycerol.

ACCESSION NUMBERS

The GEO accession numbers for the RNA-seq and Hi-C datasets are GEO: GSE71275 and GEO: GSE71072, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and seven tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.07.038>.

AUTHOR CONTRIBUTIONS

Q.W. conceived the project. Y.G. performed 3C, 4C, ChIP-qPCR, and RNA-seq. Q.X. performed computational analyses. D.C., D.U.G., and I.J. preformed Hi-C. J.S. and J.L. performed CRISPR. H.W. and Y.Z. performed EMSA. Y.T., Y.L., Y.W., Z.J., and W.L. performed cell culture and mouse work. Y.G., M.Q.Z., B.R., A.R.K., T.M., and Q.W. prepared the manuscript.

ACKNOWLEDGMENTS

We thank M. Capecci, Z. Chen, S. Lomvardas, K. Monahan, S. O'Keeffe, L. Peng, and Y. Shi for critical reading of the manuscript, and we thank X. Huang and J. Xi for providing CRISPR/Cas9 plasmids. This study was supported by grants from MOST (2009CB918700), NSFC (31171015 and 31470820), Shanghai Municipality (13XD1402000 and 14JC1403600), and State Key Laboratory of Medical Genomics to Q.W.; from NIH (NS043915) to T.M.; from a collaborative grant between NSFC (81261120390) and NCI (3P01-CA013106-41S1) to Q.W. and A.R.K.; and from NIH (HG001696), MOST (2012CB316503), and NSFC (91019016) to M.Q.Z. D.C. is a Helen Hay Whitney postdoctoral fellow. Q.W. is a Shanghai Subject Chief Scientist.

Received: March 17, 2015

Revised: May 30, 2015

Accepted: July 22, 2015

Published: August 13, 2015

REFERENCES

- Alt, F.W., Zhang, Y., Meng, F.L., Guo, C., and Schwer, B. (2013). Mechanisms of programmed DNA lesions and genomic instability in the immune system. *Cell* 152, 417–429.
- Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27, 299–308.
- Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. *Cell* 144, 327–339.
- Chen, W.V., and Maniatis, T. (2013). Clustered protocadherins. *Development* 140, 3297–3302.
- Chen, W.V., Alvarez, F.J., Lefebvre, J.L., Friedman, B., Nwakweze, C., Geiman, E., Smith, C., Thu, C.A., Tapia, J.C., Tasic, B., et al. (2012). Functional significance of isoform diversification in the protocadherin gamma gene cluster. *Neuron* 75, 402–409.
- Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* 2, 292–301.
- de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 502, 499–506.
- Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306–1311.
- 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.
- Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature* 518, 331–336.
- Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schijvers, 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.
- Doyle, B., Fudenberg, G., Imakaev, M., and Mirny, L.A. (2014). Chromatin loops as allosteric modulators of enhancer-promoter interactions. *PLoS Comput. Biol.* 10, e1003867.
- ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.
- Garrett, A.M., Schreiner, D., Lobas, M.A., and Weiner, J.A. (2012). γ-protocadherins control cortical dendrite arborization by regulating the activity of a FAK/PKC/MARCKS signaling pathway. *Neuron* 74, 269–276.
- Gibcus, J.H., and Dekker, J. (2013). The hierarchy of the 3D genome. *Mol. Cell* 49, 773–782.
- Guo, Y., Monahan, K., Wu, H., Gertz, J., Varley, K.E., Li, W., Myers, R.M., Maniatis, T., and Wu, Q. (2012). CTCF/cohesin-mediated DNA looping is required for protocadherin α promoter choice. *Proc. Natl. Acad. Sci. USA* 109, 21081–21086.
- 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.
- Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39, 311–318.
- Hirayama, T., Tarusawa, E., Yoshimura, Y., Galjart, N., and Yagi, T. (2012). CTCF is required for neural development and stochastic expression of clustered *Pcdh* genes in neurons. *Cell Rep.* 2, 345–357.
- Hou, C., Dale, R., and Dean, A. (2010). Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl. Acad. Sci. USA* 107, 3651–3656.
- Jia, Z., Guo, Y., Tang, Y., Xu, Q., Li, B., and Wu, Q. (2014). Regulation of the protocadherin *Celsr3* gene and its role in *globus pallidus* development and connectivity. *Mol. Cell. Biol.* 34, 3895–3910.
- Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of *in vivo* protein-DNA interactions. *Science* 316, 1497–1502.
- Kehayova, P., Monahan, K., Chen, W., and Maniatis, T. (2011). Regulatory elements required for the activation and repression of the protocadherin-alpha gene cluster. *Proc. Natl. Acad. Sci. USA* 108, 17195–17200.
- Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., and Ren, B. (2007). Analysis of the

- vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 128, 1231–1245.
- Lefebvre, J.L., Kostadinov, D., Chen, W.V., Maniatis, T., and Sanes, J.R. (2012). Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. *Nature* 488, 517–521.
- Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. *Cell* 157, 13–25.
- Li, J., Shou, J., Guo, Y., Tang, Y., Wu, Y., Jia, Z., Zhai, Y., Chen, Z., Xu, Q., and Wu, Q. (2015). Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. *J. Mol. Cell Biol.* 7, 284–298.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293.
- Loabanenkov, V.V., Nicolas, R.H., Adler, V.V., Paterson, H., Klenova, E.M., Polotskaja, A.V., and Goodwin, G.H. (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene* 5, 1743–1753.
- Monahan, K., Rudnick, N.D., Kehayova, P.D., Pauli, F., Newberry, K.M., Myers, R.M., and Maniatis, T. (2012). Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of protocadherin- α gene expression. *Proc. Natl. Acad. Sci. USA* 109, 9125–9130.
- Nakahashi, H., Kwon, K.R., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruitt, 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.
- Ong, C.T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* 15, 234–246.
- 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.
- Renda, M., Baglivo, I., Burgess-Beusse, B., Esposito, S., Fattorusso, R., Felsenfeld, G., and Pedone, P.V. (2007). Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. *J. Biol. Chem.* 282, 33336–33345.
- Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* 147, 1408–1419.
- Ribich, S., Tasic, B., and Maniatis, T. (2006). Identification of long-range regulatory elements in the protocadherin-alpha gene cluster. *Proc. Natl. Acad. Sci. USA* 103, 19719–19724.
- Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Gonçalves, A., Kutter, C., Brown, G.D., Marshall, A., Flícek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* 148, 335–348.
- Schreiner, D., and Weiner, J.A. (2010). Combinatorial homophilic interaction between gamma-protocadherin multimers greatly expands the molecular diversity of cell adhesion. *Proc. Natl. Acad. Sci. USA* 107, 14893–14898.
- Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Loabanenkov, V.V., and Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature* 488, 116–120.
- 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.
- Suo, L., Lu, H., Ying, G., Capecchi, M.R., and Wu, Q. (2012). Protocadherin clusters and cell adhesion kinase regulate dendrite complexity through Rho GTPase. *J. Mol. Cell Biol.* 4, 362–376.
- Tasic, B., Nabholz, C.E., Baldwin, K.K., Kim, Y., Rueckert, E.H., Ribich, S.A., Cramer, P., Wu, Q., Axel, R., and Maniatis, T. (2002). Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing. *Mol. Cell* 10, 21–33.
- Thu, C.A., Chen, W.V., Rubinstein, R., Cheevee, M., Wolcott, H.N., Felsovalyi, K.O., Tapia, J.C., Shapiro, L., Honig, B., and Maniatis, T. (2014). Single-cell identity generated by combinatorial homophilic interactions between α , β , and γ protocadherins. *Cell* 158, 1045–1059.
- 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.
- Wang, H., Maurano, M.T., Qu, H., Varley, K.E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., et al. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* 22, 1680–1688.
- Wu, Q. (2005). Comparative genomics and diversifying selection of the clustered vertebrate protocadherin genes. *Genetics* 169, 2179–2188.
- Wu, Q., and Maniatis, T. (1999). A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* 97, 779–790.
- Wu, Q., Zhang, T., Cheng, J.F., Kim, Y., Grimwood, J., Schmutz, J., Dickson, M., Noonan, J.P., Zhang, M.Q., Myers, R.M., and Maniatis, T. (2001). Comparative DNA sequence analysis of mouse and human protocadherin gene clusters. *Genome Res.* 11, 389–404.
- Xiao, T., Wallace, J., and Felsenfeld, G. (2011). Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol. Cell. Biol.* 31, 2174–2183.
- Xie, X., Mikkelsen, T.S., Gnirke, A., Lindblad-Toh, K., Kellis, M., and Lander, E.S. (2007). Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *Proc. Natl. Acad. Sci. USA* 104, 7145–7150.
- Yokota, S., Hirayama, T., Hirano, K., Kaneko, R., Toyoda, S., Kawamura, Y., Hirabayashi, M., Hirabayashi, T., and Yagi, T. (2011). Identification of the cluster control region for the protocadherin-beta genes located beyond the protocadherin-gamma cluster. *J. Biol. Chem.* 286, 31885–31895.
- Zhang, T., Haws, P., and Wu, Q. (2004). Multiple variable first exons: a mechanism for cell- and tissue-specific gene regulation. *Genome Res.* 14, 79–89.
- Zuin, J., Dixon, J.R., van der Reijden, M.I., Ye, Z., Kolovos, P., Brouwer, R.W., van de Corput, M.P., van de Werken, H.J., Knoch, T.A., van IJcken, W.F., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. USA* 111, 996–1001.

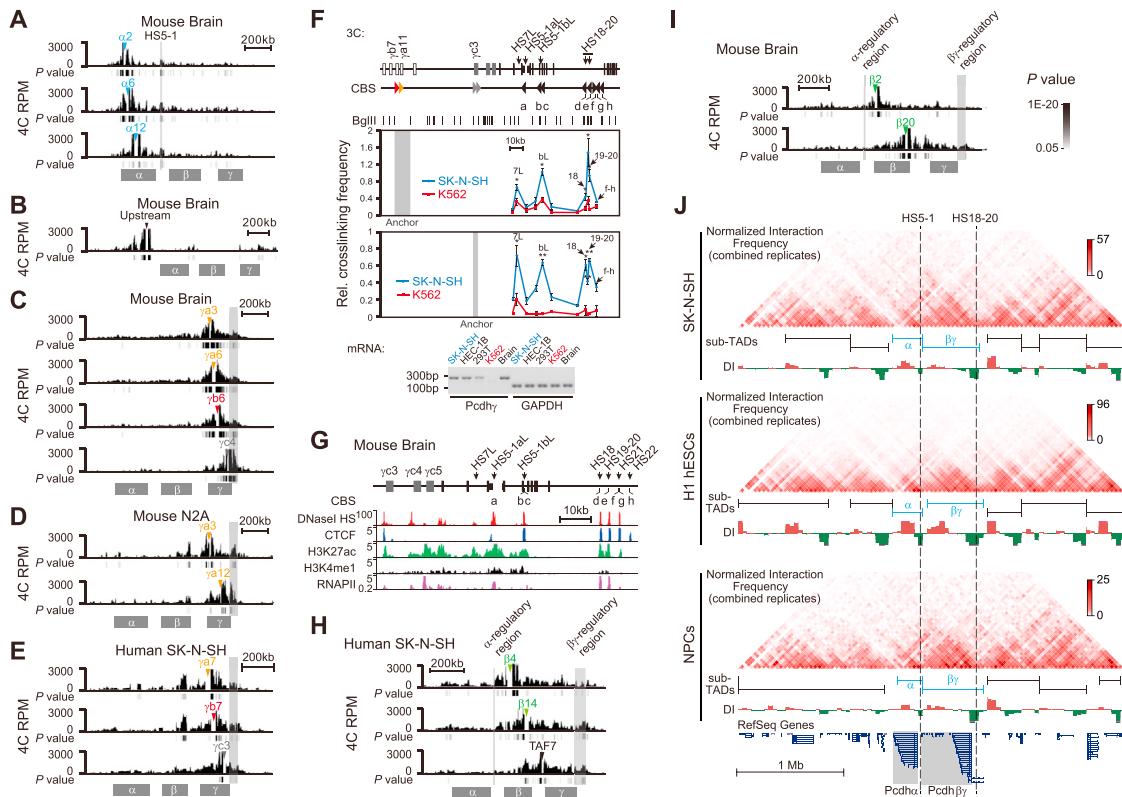


Figure S1. Additional 3C, 4C, and Hi-C Data Characterizing Clusters of *Pcdh* CBS Sites and Supporting the Two CTCF/Cohesin-Mediated Topological Domains in the Three *Pcdh* Gene Clusters, Related to Figure 1

(A–E) The 4C interaction profiles of members of the *Pcdh* α and γ gene clusters. Relative distribution of 4C reads using the α_2 , α_6 , or α_{12} promoter region (A) and a CBS upstream of the α cluster (B) as anchors in mouse brain tissues. HS5-1 is highlighted. 4C interaction profiles of the $\gamma\alpha_3$, $\gamma\alpha_6$, $\gamma\beta_6$, or $\gamma\gamma_4$ in mouse brain tissues (C), of $\gamma\alpha_3$ or $\gamma\alpha_{12}$ in mouse N2A cells (D), or of $\gamma\alpha_7$, $\gamma\beta_7$, or $\gamma\gamma_3$ in human SK-N-SH cells (E). The downstream regulatory region that interacts with the anchoring variable promoters is highlighted. 4C-seq reads were plotted as reads per million (4C RPM).

(F) The long-range chromatin-looping interactions were confirmed by quantitative 3C in SK-N-SH cells, with K562 cells as a negative control. RT-PCR analysis of mRNA expression reveals no expression of *Pcdh* γ in K562 cells and confirms the expression of *Pcdh* γ in SK-N-SH cells. Data are means \pm SEM ($n = 3$). * $p < 0.05$ and ** $p < 0.01$.

(G) Molecular marks in the $\beta\gamma$ regulatory region. Shown are the signal profiles of ChIP-seq in the regulatory region downstream of the *Pcdh* γ cluster in mouse brain tissues (ENCODE Project Consortium, 2012). The locations of DNasel hypersensitive sites in this region are indicated by vertical arrows.

(H and I) Relative distribution of 4C reads obtained using β_4 , β_{14} , or *TAF7* as anchors in human SK-N-SH cells (H) or β_2 and β_{20} as anchors in mouse brain (I). The α - and $\beta\gamma$ -regulatory regions are indicated as gray boxes.

(J) Hi-C interaction frequency and TAD calls at the *Pcdh* gene clusters. Hi-C data from 3 cell types are shown for a 3.6 MB region centered on the *Pcdh* gene clusters (Chr5:138720000-142320000; hg19). From top to bottom, data are from SK-N-SH cells, H1 human Embryonic Stem Cells (H1-hESC), and H1-derived Neural Precursor Cells (NPCs). Normalized interaction frequency, topological associating domains ("sub-TADs") (Dixon et al., 2015), and Directionality Index (DI), a measure of directional chromatin interactions at a particular region, with a sliding window of 300 kb are shown for each cell type, together with RefSeq genes at the bottom. DI measures whether interactions from a given anchor point are skewed toward upstream regions (green) or downstream regions (red). Vertical dotted (dashed) lines indicate the locations of 4C anchor points HS5-1 and HS18-20.

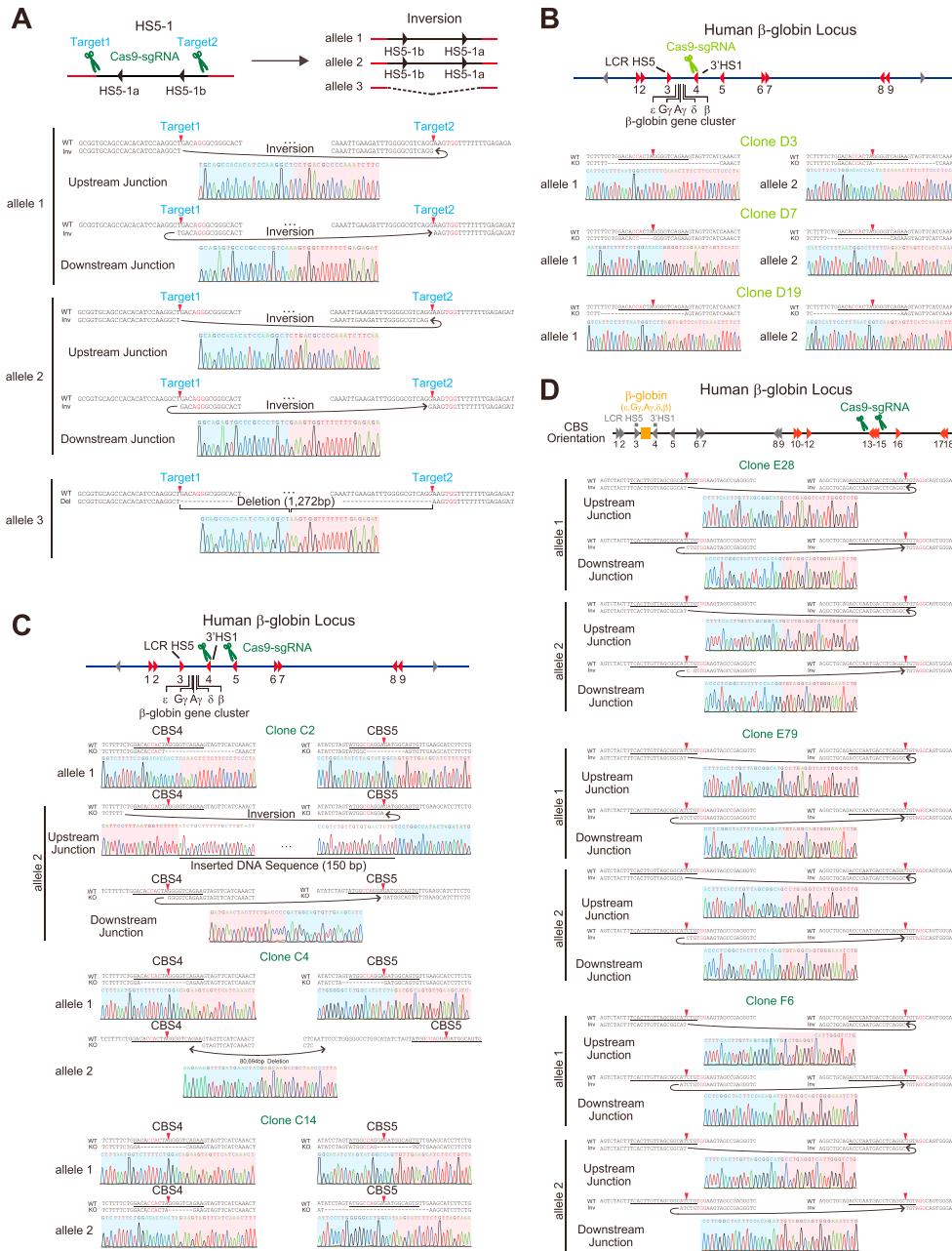


Figure S2. DNA Sequencing of CRISPR Cell Clones Generated with Two sgRNAs, Related to Figures 2 and 5

(A) Generation of subcloned cell line of the *HS5-1* inversion using the CRISPR/Cas9 System, related to Figure 2. Upper panel, schematic diagram showing the dual-sgRNA-mediated DNA-fragment inversion. CBS *HS5-1a* and *HS5-1b* are indicated. Lower panel, confirmation of the three CRISPR alleles of the *HS5-1* targeting cell line by DNA sequencing: two alleles with *HS5-1* inversion and one with *HS5-1* deletion. Note that the inversion junctions for the two alleles are different, so these two inversion alleles can easily be distinguished. The DNA sequences of the junctions of inversion or deletion for the three alleles are shown.

(B) Generation of subcloned cell lines with CBS4(3'HS1) knockout by CRISPR/Cas9. Diagram showing the targeting of 3'HS1 by the CRISPR/Cas9 system. Sequencing confirmation of the two targeted alleles from subcloned D3, D7, and D19 knockout cell lines. The PAM sequence is labeled in red. The 20-nt core sequence of the CBS is underlined.

(C) Diagram showing the double knockout of both CBS4 and CBS5 by the CRISPR/Cas9 system. Sequencing confirmation of the two targeted alleles from subcloned C2, C4, and C14 double-knockout cell lines. The PAM sequences are labeled in red. The 20-nt core sequences of the CBSs are underlined.

(D) Generation of subcloned cell line with inversion of CBS13–15 by CRISPR/Cas9, related to Figure 5. Diagram showing the targeting of CBS13–15 by the CRISPR/Cas9 system. Sequencing confirmation of the two targeted alleles from subcloned E28, E79, and F6 inversion cell lines. The PAM sequences are labeled in red. The 20-nt target sequences are underlined.

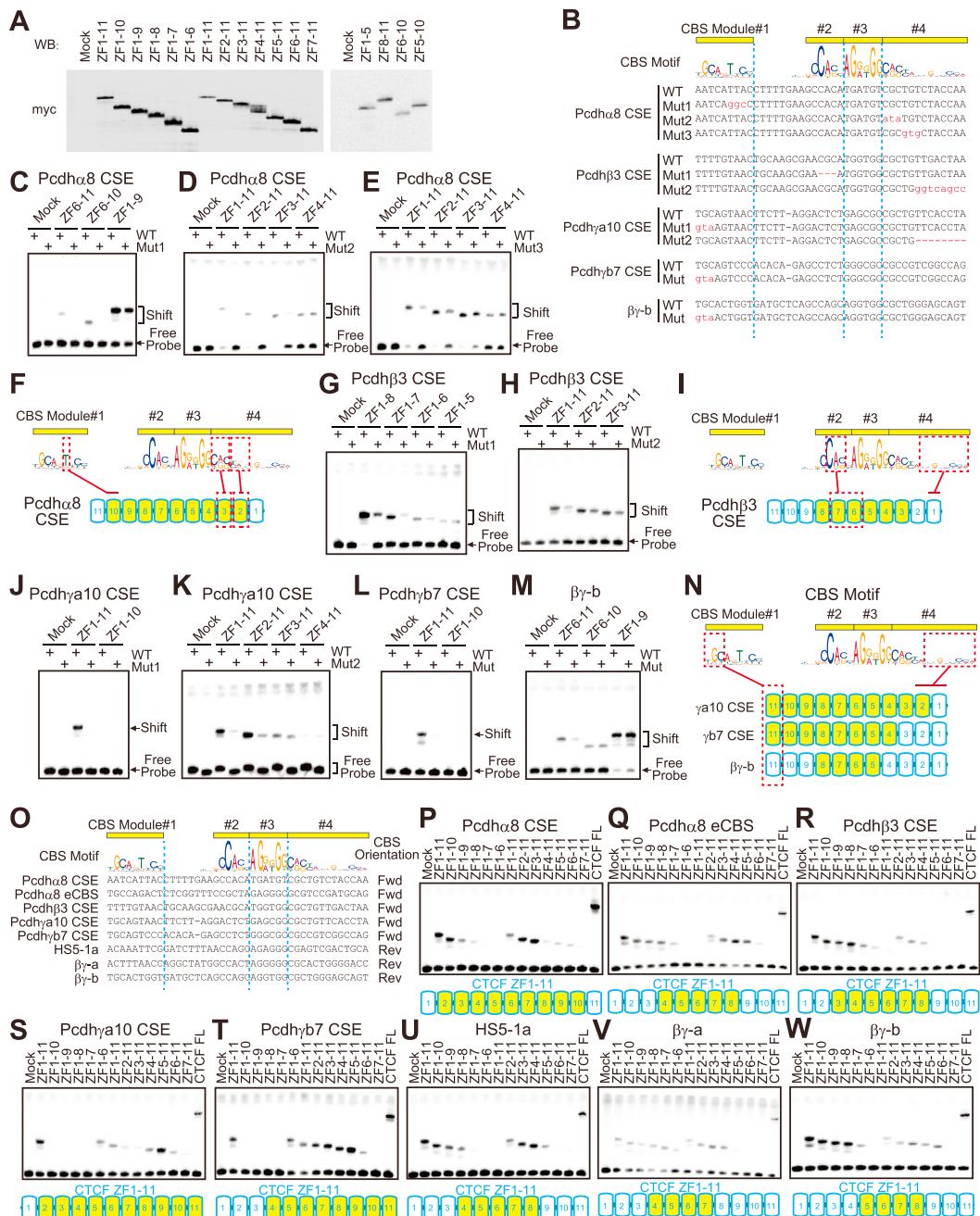


Figure S3. Directional Binding of CTCF to a Repertoire of the *Pcdh* CBSs, Related to Figure 3

(A) Western blot confirmation of a series of recombinant CTCF ZF domains with sequential deletions of ZFs from either N terminus or C terminus.

(B) The wild-type (WT) or mutant (Mut) CBS sequences of *Pcdh α8, β3, γa10, γb7*, and *βγ-b*

(C-E) Gel-shift assays using recombinant CTCF proteins with the CBS probes of the *Pcdha8* CSE.

(F) Recognition of the *Pcdha8* CBS modules 1 and 4 by the CTCF ZF domains.

(G and H) Gel-shift assays using recombinant CTCF proteins with the CBS probes of the *Pcdhb3* CSE.

(I) Recognition of the *Pcdhβ3* CBS modules 2 and 4 by the CTCF ZF domains.

(J–M) Gel-shift assays using recombinant CTCF proteins with the CBS probes of *Pcdh* γa10 (J and K) and *WIF1* (L and M).

(N) Recognition of the *Pcdh* $\gamma\alpha10$, $\gamma\beta7$, or $\beta\gamma-b$ CBS modules 1 and 4 by the CTCF ZF domains.

(O) The CBS sequences in the *Pcdh* promoter region or regulatory region and their orientations are shown.

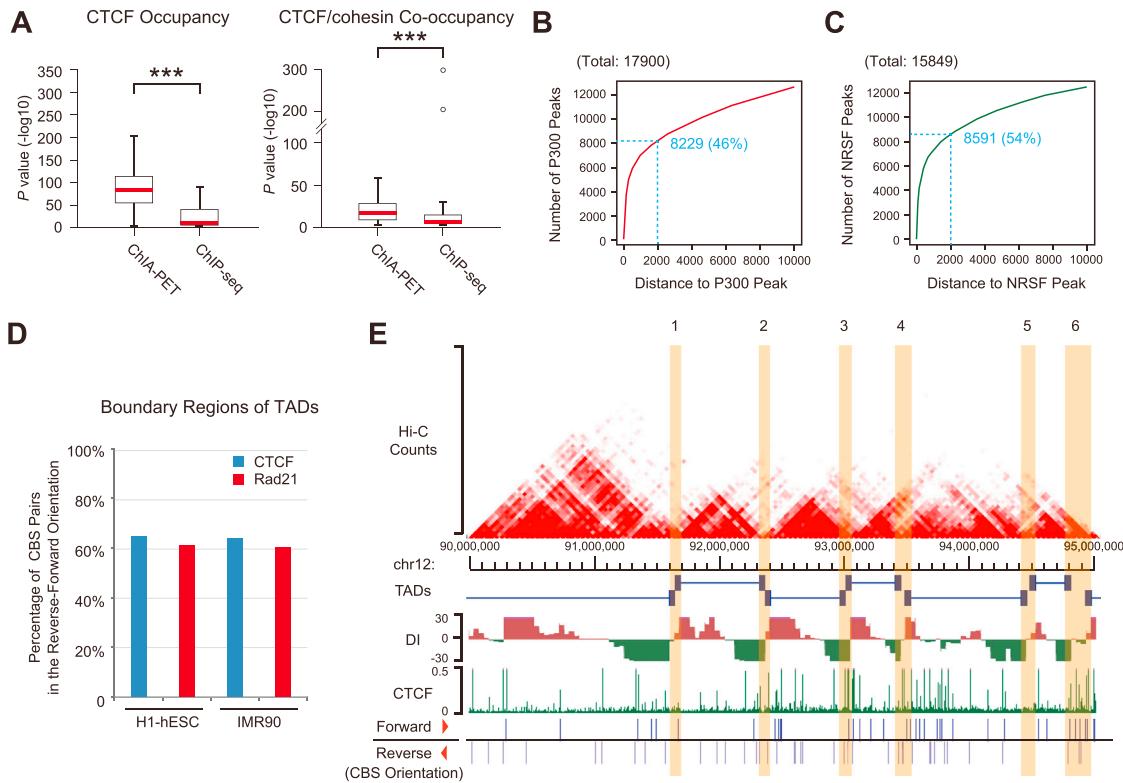


Figure S4. Additional Computational Analyses Characterizing Genome-wide Directional CTCF/Cohesin Recognition of CBS Sites, Related to Figure 4

(A) Higher occupancy of CTCF/cohesin complex in ChIA-PET than ChIP-seq in the K562 genome. Boxplots show the interquartile range (IQR) between first and third quartiles and the red line inside marks the median. The whiskers indicate the lowest and highest values within $1.5 \times$ IQR from the first and third quartiles. Outliers beyond the whiskers are indicated with circles. *** $p < 0.001$ by the Mann-Whitney test.

(B and C) The relationship between the CTCF and p300 binding sites (B) or between the CTCF and REST/NRSF binding sites (C) in the human genome. The ChIP-seq dataset (ENCODE Project Consortium, 2012) was filtered for a $-\log_{10}$ score (P value) of > 20 to retain 32,271 peaks of CTCF occupancy in K562 cells.

(D and E) Cumulative features of CBS orientations in the TAD boundaries in the human genome. (D) The percentage of CBS pairs in the reverse-forward orientation, identified within the CTCF and cohesin (Rad21) ChIP-seq peaks (Dixon et al., 2012; ENCODE Project Consortium, 2012) and located in the boundary regions of TADs in human H1-hESC and IMR90 cells. (E) The normalized Hi-C counts, TADs, directionality index (DI), CTCF occupancy (Dixon et al., 2012; ENCODE Project Consortium, 2012), and the orientation of CBSs of a Chr12 genomic region are shown as an example. Note that all boundary regions of neighboring TADs (number 1-6) have CBS pairs in the reverse-forward orientation, except boundary number 5, which only has one close CBS in the forward orientation.

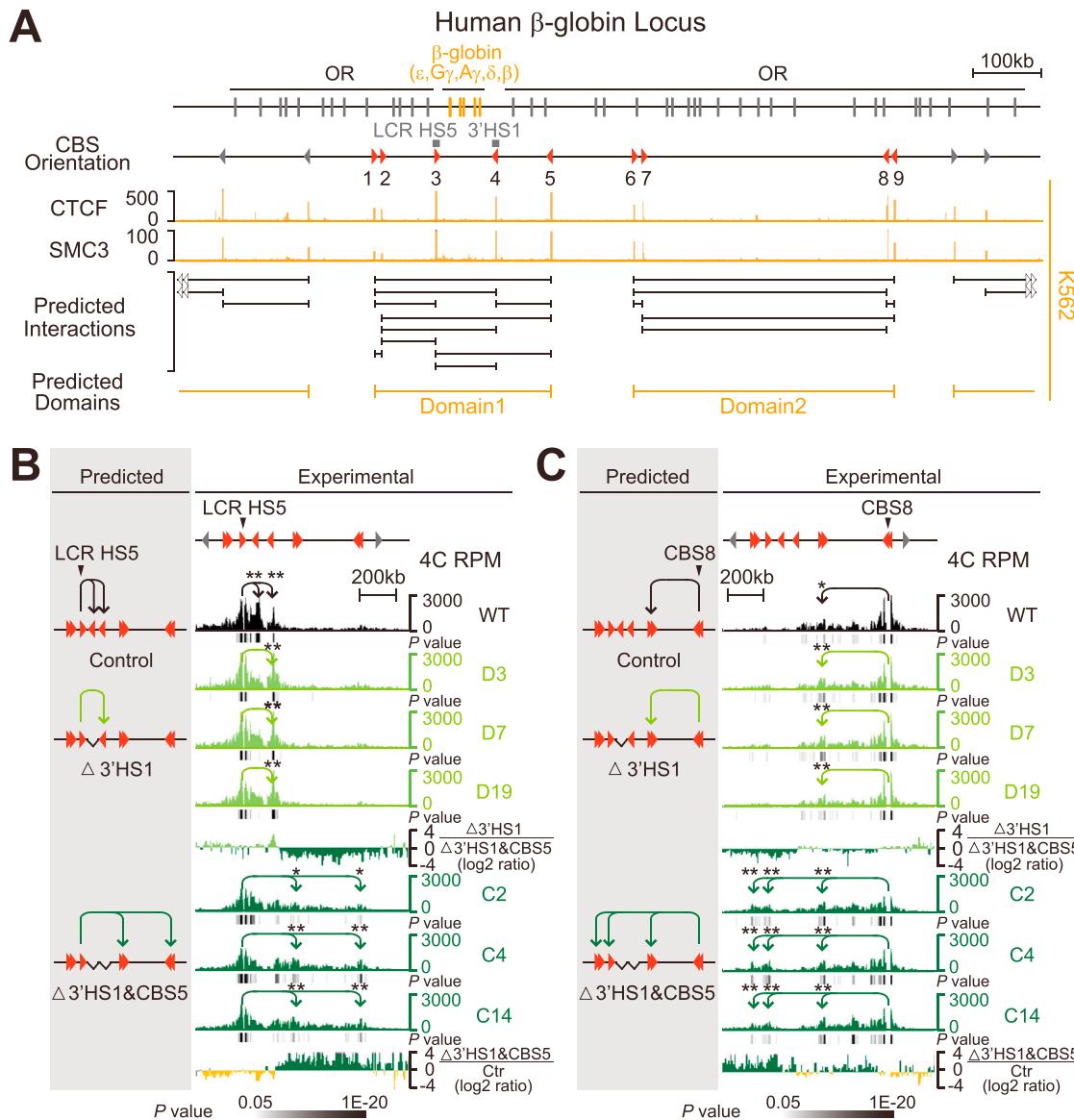


Figure S5. CTCF/Cohesin-Mediated Directional DNA Looping in the Human β -globin Gene Cluster, Related to Figure 5

(A) Diagram of the human β -globin cluster (yellow) and the flanking olfactory receptor (OR) clusters (gray). The HS5 within LCR and 3'HS1 are indicated. Predicted looping interactions and topological domains, based on location and orientation of CBSs as well as both CTCF and cohesin (SMC3) co-occupancy in K562 cells, are also shown.

(B) The predicted interactions, in the control, CBS4(3'HS1)-knockout (D3, D7, and D19), or CBS4/5 double-knockout (C2, C4, and C14) subcloned CRISPR cell lines, are confirmed by 4C with CBS3(5'HS5) as an anchor. The average log2 ratios of interactions are indicated between the profiles. *p < 0.05 and **p < 0.01.

Cell

Supplemental Information

CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function

**Ya Guo, Quan Xu, Daniele Canzio, Jia Shou, Jinhuan Li, David U. Gorkin, Inkyung Jung,
Haiyang Wu, Yanan Zhai, Yuanxiao Tang, Yichao Lu, Yonghu Wu, Zhilian Jia, Wei Li,
Michael Q. Zhang, Bing Ren, Adrian R. Krainer, Tom Maniatis, and Qiang Wu**

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Cell Culture

Human neuroblastoma SK-N-SH and endometrial adenocarcinoma HEC-1B cells, and mouse neuro2A (N2A) cells were cultured in MEM (Gibco), supplemented with 10% (v/v) FBS (Gibco), 2 mM GlutaMAX (Gibco), 1 mM sodium pyruvate (Sigma), and 1% penicillin-streptomycin (Gibco). Human K562 and HEK293T cells were cultured in DMEM (Hyclone) supplemented with 10% (v/v) FBS and 1% penicillin-streptomycin. Cells were maintained at 37 °C in a humidified incubator containing 5% CO₂.

Recombinant CTCF Protein Production

The recombinant full-length CTCF proteins were prepared by a pTNT-CTCF plasmid through *in vitro* translation in the rabbit reticulocyte lysate as previously described (Guo et al., 2012; Jia et., 2014). A series of truncated CTCF proteins with sequential deletion of each zinc finger domain from either N- or C- terminus

were prepared similarly from a repertoire of 17 sequencing-confirmed plasmids constructed by PCR and subcloning. The primer sets used are listed in Table S7.

Western Blotting

The *in-vitro*-synthesized proteins were diluted with RIPA lysis buffer containing 1 mM PMSF. Proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes. The membranes were then incubated with mouse anti-myc antibody (Millipore). Finally, the membranes were incubated with goat anti-mouse secondary antibody and scanned by using the Odyssey System (LI-COR Biosciences).

Electrophoretic Mobility Shift Assay (EMSA)

The sequences containing various CBS sites were cloned into the pGEM-T Easy plasmid (Promega). The mutations of each CBS site were constructed by PCR on the wild-type templates. Probes were amplified by PCR with high-fidelity DNA polymerase using 5' biotin-labeled primers and were gel-purified. The DNA concentration of the probes was measured with a NanoDrop (Thermo). Each binding reaction contained equimolar amounts of the biotin-labeled probes. The primers used are listed in Table S7. EMSA was performed using LightShift Chemiluminescent EMSA reagents (Thermo) as described in

the manufacturer's manual. Briefly, the probes were incubated with *in-vitro*-synthesized proteins in binding buffer containing 10 mM Tris, 50 mM KCl, 5 mM MgCl₂, 0.1 mM ZnSO₄, 1 mM dithiothreitol, 0.1% (v/v) Nonidet P-40 (NP-40), 50 ng/μl poly (dI-dC), and 2.5% (v/v) glycerol at room temperature for 20 min. The binding complex was electrophoresed on 5% nondenaturing polyacrylamide gels in ice-cold 0.5×TBE buffer (45 mM Tris-borate, 1 mM EDTA, pH8.0). The gel was electrotransferred to a nylon membrane in ice-cold 0.5×TBE buffer. After crosslinking using UV-light for 10 min, the membrane was incubated with stabilized streptavidin-horseradish peroxidase conjugate and rinsed with the washing buffer. The biotin-labeled DNA was then detected by chemiluminescence using the ChemiDoc XRS+ system (Bio-Rad).

Chromosome Conformation Capture (3C)

Quantitative 3C was performed according to described procedures (Guo et al., 2012). Briefly, cells were cross-linked with 1% (v/v) formaldehyde for 10 min at 37 °C. After nuclear extraction, cross-linked DNA was digested overnight with 400 U of BgIII or EcoRI at 37 °C while shaking at 900 rpm. After self-ligation, the DNA was purified and quantified using PicoGreen reagents (Invitrogen). The final quantitative PCR reactions were performed in triplicates by using SYBR Green (Roche) with 100 ng DNA as templates.

The 3C control experiments were performed according to the published method (Guo et al., 2012). BAC clones (CTD-3042N4, CTD-2506I6, CTD-2538C12, CTD-2527B17 and CTD-2371G16 from Invitrogen) were isolated and purified using a large-construct DNA isolation kit (Qiagen). The molar amount of these BAC clones was detected by qPCR titration using primers matching the BAC backbone sequences. Equimolar amounts of BAC clones were digested with 400 U BgIII or EcoRI at 37 °C overnight. After purification, the DNA was then ligated with T4 DNA ligase at 16 °C overnight. The ligated BAC DNA was used to establish a standard PCR amplification curve by serial dilutions.

To compare long-range DNA interaction frequencies in different cell lines, the PCR reactions were normalized to the ligation frequency of six restriction fragments of the tubulin, phosphoglycerate kinase 1, and 14-3-3 loci. The following six primer pairs were used for normalization between SK-N-SH and K562 cells: 3 pairs of tubulin (BgIII-PGK-1 and BgIII-PGK-3, BgIII-PGK-2 and BgIII-PGK-4, BgIII-PGK-3 and BgIII-PGK-5), and 3 pairs of 14-3-3 ζ/δ (BgIII-YWHAZ-1 and BgIII-YWHAZ-3, BgIII-YWHAZ-2 and BgIII-YWHAZ-3, BgIII-YWHAZ-2 and BgIII-YWHAZ-4) (Table S7). The following six primer pairs were used for normalization between inversion and wild-type HEC-1B cell lines: 3 pairs of tubulin (TUBB-1 and TUBB-2, TUBB-1 and TUBB-4, TUBB-2 and TUBB-4), and 3 pairs of 14-3-3 ζ/δ (YWHAZ-1 and YWHAZ-2, YWHAZ-1 and YWHAZ-5, YWHAZ-2 and YWHAZ-6). These 3C experiments were performed

at least three times. Data are means ± SEM. The significance of the differences was evaluated by the Student's t-test.

Circularized Chromosome Conformation Capture (4C)

The 4C-seq libraries were constructed as described (Guo et al., 2012; Jia et al., 2014; Simonis et al., 2006; Splinter et al., 2012) with some modifications. Briefly, mouse brain tissues were dispersed by collagenase (1.25 mg/ml, Sigma) treatment in DMEM supplemented with 10% (v/v) FBS for 45 min at 37 °C while rotating at 700 rpm. Cells were then filtered through a 40-μm cell strainer (BD Biosciences) to make a single-cell suspension. A total of 10⁷ cells were cross-linked and then lysed to prepare cell nuclei. The cross-linked DNA in the nuclear preparations was digested with HindIII or EcoRI overnight while rotating at 900 rpm and then ligated with T4 DNA ligase. After purification, the DNA was digested with a second enzyme, DpnII or NlaIII, and was ligated again. The religated DNA was then purified using a High-Pure PCR Product Purification system (Roche). A series of 4C-seq libraries were generated by inverse PCR using a high-fidelity DNA polymerase with primer pairs containing Illumina adapter sequences (Table S7). High-throughput sequencing was performed using 49-bp single-end reads on an Illumina HiSeq 2000 platform. The sequenced reads were mapped to the mouse (NCBI37/mm9) or human (GRCh37/hg19) reference genomes using Bowtie (version 1.0.0) (Langmead et al., 2009). The r3Cseq program in the R/Bioconductor package (Thongjuea

et al., 2013) was used to detect statistically significant long-range chromatin-looping interactions. The sequencing data were visualized in the UCSC genome browser (Kent et al., 2002). All 4C-seq experiments were performed with at least two biological replicates.

Hi-C Data Generation and Analysis

Hi-C for SK-N-SH cells was performed as previously described (Dixon et al., 2012; 2015). We performed two biological replicates, each with roughly 2.5×10^8 cells. We obtained a total of more 200 million read pairs per replicate. We constructed normalized Hi-C contact matrices at 40-kb resolution after removing intrinsic biases in Hi-C data by using HiCNorm (Hu et al., 2012). Normalized contact matrices for the two replicates were highly correlated (Pearson correlation coefficient > 0.89). Topologically associated domains (“TADs” or “sub-TADs”) were identified based on Directionality Index (“DI”) as described (Dixon et al., 2012; 2015) with one exception: DI was calculated using a sliding window of 300 kb upstream/downstream of the anchor point. A smaller DI window will yield smaller TADs, while a larger window will yield larger TADs. Hi-C data in H1 human Embryonic Stem Cells and H1-derived Neural Precursor Cells were previously generated (Dixon et al., 2015).

CRISPR/Cas9 System

The DNA fragment inversion and deletion by CRISPR/Cas9 were performed as previously described (Li et al. 2015). The templates for producing target sgRNAs were constructed by PCR using pLKO.1 or pGL3-U6-sgRNA-PGK-Puro plasmid (Chang et al., 2013; Cong et al., 2013; Mali et al., 2013; Shen et al., 2014; Li et al., 2015) with appropriate primers (Table S7). All plasmids were confirmed by sequencing. To screen for inversion cell clones, HEC-1B or HEK293T cells at about 80% confluence were transfected with Lipofectamine 2000 reagents (Invitrogen) in a 6-well plate with plasmid DNA including pcDNA3.1-Cas9 and sgRNA constructs (2 µg each). The primers used for genotyping are listed in Table S7.

Reverse-transcriptase PCR

Total RNA was extracted from brain tissues or cultured cells using the Qiagen RNeasy system. Reverse-transcription was performed using reagents from Promega with 1 µg of total RNA. PCR was then performed as follows: 94 °C for 4 min; 35 cycles of 94 °C for 30 sec, 60 °C for 30 sec, 72 °C for 30-60 sec; and 72 °C for 5 min.

Genome-wide Computational Analyses

To identify putative CBSs and their orientations, we scanned ChIP-seq peak regions in human K562, MCF-7, H1-hESC, and IMR90 as well as mouse E14 pluripotent cells using the STORM program (CREAD-0.84) and the CTCF position weight matrices (PWM) (Schones et al., 2007; Schmidt el al., 2012). We defined CBS sequences with the highest PWM score on the forward (forward orientation) or on the reverse (reverse orientation) strands as a CTCF-occupied CBS by using the STORM program. To study the correlation between CBS orientation and CTCF-mediated chromatin looping, we first filtered CTCF ChIA-PET interactions for tethered DNA fragments in which both fragments contain CBSs (ENCODE Project Consortium, 2012; Handoko et al., 2011). ChIA-PET measures interactions of DNA fragments of paired-end tags (PETs) in a form of “tag-linker-tag”. Inter-ligation PETs refer to the reads from different DNA fragments. PET sequences that overlap at both ends form PET clusters. PET clusters of multiple PETs reflect the strength of chromatin interactions. Thus, inter-ligation PETs predict the chromatin interactions by clustering (Li et al., 2012). We then screened for CTCF/cohesin-mediated interactions in different combinations of orientation configuration of CBS pairs (i.e., forward-reverse, forward-forward, reverse-reverse, and reverse-forward) using a Python script. Clusters of overlapping chromatin-looping interactions with looping strength >300 (Li et al. 2010; Handoko et al., 2011) were merged to form a CTCF/cohesin-mediated chromatin domain (CCD). The orientation

configuration of CBS pairs between neighboring domains was quantified for each chromosome and combined to give the total number of domains in the whole human genome. The sources of the public data (ENCODE Project Consortium, 2012; Handoko et al., 2011) used: Gene Expression Omnibus (GEO) accession numbers: GSM822297, GSM935379, GSM935404, GSM935624, GSM935407, GSM935310, GSM970216, GSM1022658, GSM1010791, and GSM970215.

Animals

Animal experiments were approved by the Institutional Animal Care and Use Committee (IACUC) of Shanghai Jiao Tong University.

RNA sequencing (RNA-Seq)

RNA-Seq experiments were performed as previously described (Mortazavi et al., 2008; Shen et al., 2012). Briefly, total RNA was extracted from cultured cells (inversion and WT control in replicates) using an RNeasy plus mini kit (Qiagen) according to the manufacturer's protocol. Messenger RNA was isolated from the total RNA by oligo (dT) magnetic beads, and fragmented under heating condition. After the first and second cDNA strand synthesis, as well as ends repairing, the 3' ends of cDNA fragments were added with a single 'A' nucleotide to facilitate adapter ligation. The sequencing libraries were then generated by PCR amplification of the cDNA products, and validated by an

Agilent 2100 Bioanalyzer before sequencing on the Illumina sequencing platform of HiSeq 2000. The resulting sequencing reads (49 bp, single-read) were mapped onto the human genome build GRCh37 using TopHat-v2.0.14 (Trapnell et al., 2009) with the setting of “-N 0 –g 1 –x 1”. The output data were then averaged among biological replicates and normalized to reads per million (RPM) by the genomeCoverageBed program (Quinlan and Hall, 2010). The images were then generated using the University of California Santa Cruz (UCSC) Genome Browser (Kent et al., 2002). In order to identify genes that were changed in expression between two groups of data, the TopHat mapped reads were analyzed by the DEGseq program with the setting of “MARS” (Wang et al., 2010). A revised genome annotation file containing the clustered *Pcdh* genes (removing the three constant exons of *Pcdh α* and γ gene clusters which can affect the comparison of gene expression, as well as adding the annotation of *Pcdh $\gamma/b5$* gene which is missing in the public genome annotation file) was used in the analysis.

SUPPLEMENTAL REFERENCES

Chang, N., Sun, C., Gao, L., Zhu, D., Xu, X., Zhu, X., Xiong, J.W., and Xi, J.J. (2013). Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. Cell Res. 23, 465-472.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819-823.

Hu, M., Deng, K., Selvaraj, S., Qin, Z., Ren, B., and Liu, J.S. (2012). HiCNorm: removing biases in Hi-C data via Poisson regression. *Bioinformatics* 28, 3131-3133.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res.* 12, 996-1006.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.

Li, G., Fullwood, M.J., Xu, H., Mulawadi, F.H., Velkov, S., Vega, V., Ariyaratne, P.N., Mohamed, Y.B., Ooi, H.S., Tennakoon, C., et al. (2010). ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol.* 11, R22.

Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148, 84-98.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823-826.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621-628.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-842.

Schones, D.E., Smith, A.D., and Zhang, M.Q. (2007). Statistical significance of cis-regulatory modules. *BMC Bioinformatics* 8, 19.

Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., Wang, L., Hodgkins, A., Iyer, V., Huang, X., et al. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods* 11, 399-402.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemse, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* 38, 1348-1354.

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.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105-1111.

Thongjuea, S., Stadhouders, R., Grosveld, F.G., Soler, E., and Lenhard, B. (2013). r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. *Nucleic Acids Res.* 41, e132.

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26, 136-138.