SUPPLEMENTAL FIGURES

Figure S1. Additional 3C, 4C, 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 γ*a3*, γ*a6*, γ*b6*, or γ*c4* in mouse brain tissues (C), of γ*a3* or γ*a12* in mouse N2A cells (D), or of γ*a7*, γ*b7*, or γ*c3* 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 ± 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 DNaseI hypersensitive sites in this region are indicated by vertical arrows.

(H and I) Relative distribution of 4C reads obtained using *β4*, *β14*, or *TAF7* as anchors in human SK-*N*-SH cells (H) or *β2* and *β20* as anchors in mouse brain (I). The *α*- and *βγ*-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* α*8*, β*3*, γ*a10*, γ*b7*, and $\beta\gamma$*-b*.

(C-E) Gel-shift assays using recombinant CTCF proteins with the CBS probes of the *Pcdh*α*8* CSE.

(F) Recognition of the *Pcdh*α*8* 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 *Pcdh*β*3* CSE.

(I) Recognition of the *Pcdh*β*3* 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* γ*a10* (J and K), γ*b7* (L), or $\beta\gamma$*-b* (M).

(N) Recognition of the *Pcdh* γ*a10*, γ*b7*, 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.

(P-W) EMSA reveals that different combinations of the CTCF 11 ZF domains have distinct CBS binding patterns.

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×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 –log10 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-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.

(C) The predicted interactions in the CBS4-knockout or CBS4/5 double-knockout cell lines are also confirmed by 4C with CBS8 as an anchor. **P* <0.05 and ***P* <0.01.