CTCF/cohesin-mediated DNA looping is required for protocadherin α promoter choice

Ya Guo[a,b], Kevin Monahan[c,1], Haiyang Wu[a,b], Jason Gertz[d], Katherine E. Varley[d], Wei Li[a,b], Richard M. Myers[d], Tom Maniatis[c,2], and Qiang Wu[a,b,2]

[a]Key Laboratory of Systems Biomedicine (Ministry of Education), Center for Comparative Biomedicine, Institute of Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, China; [b]State Key Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200240, China; [c]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032; and [d]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806

Contributed by Tom Maniatis, November 6, 2012 (sent for review October 23, 2012)

The closely linked human protocadherin (***Pcdh***) **α**, **β**, and **γ** gene clusters encode 53 distinct protein isoforms, which are expressed in a combinatorial manner to generate enormous diversity on the surface of individual neurons. This diversity is a consequence of stochastic promoter choice and alternative pre-mRNA processing. Here, we show that ***Pcdhα*** promoter choice is achieved by DNA looping between two downstream transcriptional enhancers and individual promoters driving the ex[illegible]

cells, we characterized specific protein–DNA interactions within the α, β, and γ clusters by ChIP-seq analyses. First, we mapped the active (H3K4me3) and repressive (H3K27me3) histone methylation marks throughout the Pcdh gene clusters. We found that each of the active α promoters and the HS7 and HS5-1 enhancers are marked by H3K4me3 (Fig. 1C). By contrast, neither the active nor inactive promoters nor the enhancers of the Pcdh clusters were marked by the repressive H3K27me3 histone modification (Fig. 1 C and F and Fig. S2A). As a positive control for the antibodies and ChIP-seq protocol, we showed that the promoter of the nonclustered Pcdh1 gene is marked by both H3K27me3 and H3K4me3 (Fig. S2C).

ChIP-seq experiments using a specific antibody against RNA polymerase II (RNAPII) revealed that RNAPII is enriched at the promoter regions of αc2 and γc3 (Fig. 1 C and F), two isoforms that are expressed at high levels in the brain (1, 5). We presume that RNAPII is not detected on the alternate promoters because of the relative low levels of expression compared with the ubiquitously expressed promoters. RNAPII is also enriched at H3K4me3-marked enhancer sites of HS7 and HS5-1 of the α cluster. It is important to note that we find similar enhancer-like sites in the human γ cluster, designated as HS7-like (HS7L), HS5-1a–like (HS5-1aL), and HS5-1b–like (HS5-1bL) [corresponding to the mouse HS17 site (11)], which are enriched in RNAPII and/or H3K4me3 (Fig. 1F). Interestingly, we also identified spliced transcripts that are complementary to the HS7 (AK136271, AK042845), HS7L (DA549378, CF593989), HS5-1aL (DA087514, AI311853), and HS5-1bL (AK094264) sites (all accession numbers from Genbank), suggesting that these enhancers have promoter activities (13). Finally, we note that these α and γ enhancer sites are DNase I–hypersensitive in the Encyclopedia of DNA Elements (ENCODE) database, which is a characteristic of transcriptional enhancers (14).

CTCF and Rad21 Binding Correlates with Promoter Activity. To investigate the relationship between the level of gene expression of the human α, β, and γ clusters and the binding of CTCF and cohesin subunit Rad21, we examined their binding profiles by ChIP-seq and ChIP–quantitative PCR (ChIP-qPCR) experiments in SK-N-SH cells (Fig. 1 C and F and Fig. S2 A and B). Previous studies of the mouse Pcdhα cluster identified two CTCF-binding sites in the alternate α isoforms, one in the promoter and a second within the downstream exon (10). We observed the same binding pattern in the human α cluster. As shown in Fig. 1C, CTCF binds to both the promoter conserved sequence element (CSE) and the exonic CTCF-binding site (eCBS) in the α4, α8, and α12 genes. This binding profile directly correlates with the high levels of expression of α4, α8, and α12 in SK-N-SH cells (Fig. 1B).

In the case of the human β and γ clusters, the binding of CTCF to CSE correlates with the high expression levels of respective isoform promoters (Fig. 1F and Fig. S2A). We do not observe a corresponding eCBS in the members of the β or γ clusters. However, we observed a second site in the ubiquitous γc3 gene that binds to high levels of CTCF (Fig. 1F). The binding profiles of Rad21 in the α, β, and γ clusters are similar to that of CTCF, i.e., Rad21 binds to two sites in active alternate promoters of the α cluster but only one site in the promoters of the β and γ clusters (Fig. 1 C and F and Fig. S2A).

Binding of CTCF and Rad21 to c-Type Promoters. As was shown in mouse cells, CTCF binds to the CSE of αc1 (Fig. 1C

CTCF binds to two sites within the HS5-1 enhancer (Fig. 1C). Reminiscent of the two HS5-1 CTCF sites downstream of the α cluster, we observed high levels of CTCF binding to HS5-1aL and HS5-1bL, downstream of the γ cluster (Fig. 1F). ChIP-seq experiments with SK-N-SH cells revealed that Rad21 is enriched at the ubiquitously expressed αc1 promoter as well as at the enhancer sites HS5-1a and HS5-1b, possibly recruited by the bound CTCF (Fig. 1C). In addition, Rad21 is enriched less so at the ubiquitously expressed αc2 promoter and the enhancer HS7 (Fig. 1C). It is important to note that CTCF does not bind to these two sites, so Rad21 binding at these sites is CTCF-independent.

We also observed Rad21 enrichments in the Pcdhγ cluster at γc3, HS5-1aL, and HS5-1bL in SK-N-SH cells (Fig. 1F), possibly also recruited by the bound CTCF. In addition, Rad21 is enriched to a lesser extent at the promoter regions of γc4 and γc5, as well as HS7L. Interestingly, the location of HS7L in the γ constant regions corresponds to the position of HS7 in the α constant region (Fig. 1F). These CTCF and Rad21-binding patterns strongly suggest that regulatory mechanisms of cell-specific Pcdh gene expression are similar for the α and γ clusters.

In Vitro Binding of Recombinant CTCF to *Pcdh* Promoters and Enhancers. To confirm the ChIP-seq data and examine the sequence requirements for CTCF binding, we carried out in vitro DNA-binding experiments with recombinant CTCF. We performed EMSA experiments using recombinant full-length human CTCF proteins and a series of probes each containing the CSE of human α1, α4, α5, α8, α12, αc1, αc2, β1, β3, β9, β15, γa2, γb1, γb5, γa10, γb7, γc3, αc4, and γc5 genes (Fig. S3A). In contrast to a previous report (15), we observed a specific binding of full-length CTCF to every probe except that of αc2, β1, γc4, and γc5 (Fig. S3A), which is consistent with motif predictions (2). The specificity of direct CTCF binding to each probe was confirmed by the detection of a supershifted complex using a CTCF antibody (Fig. S3A). In addition, mutations in either the CGCTG core sequences within the CSE or within the immediate upstream 5 nt abolished the retarded gel-shift complex (Fig. S3B). These data demonstrate that both the CGCTG core sequences and the immediate upstream 5 nt, which correspond to the conserved CCCTC motif within the CTCF consensus (16), are essential for specific binding of CTCF to the CSE.

As shown in Fig. S4A, exonic CTCF-binding sites can be identified in the human α2 to α13 but not α1 genes. EMSA experiments with the recombinant CTCF revealed a shifted complex and specific supershifted band for the eCBS of α4, α5, α8, and α12 but not α1 (Fig. S3A). Moreover, mutation of the eCBS CTCF site abolished the gel-shift complex (Fig. S4B). Thus, CTCF binds directly to two sites in each of human alternate promoters α2 to α13 but only to one site of CSE in the human α1 promoter, consistent with recent findings of two CTCF sites at the 5′ end of each mouse alternate α isoform by ChIP-seq studies (10).

We next carried out EMSA experiments with the HS5-1a and HS5-1b sites, confirming that CTCF binds to both sites within the HS5-1 enhancer (Fig. S3A). Furthermore, mutations of the core consensus for each of these two sites abolish the CTCF binding (Fig. S4 C and D). In addition, we observed a HS5-1a–like CTCF-binding site (HS5-1aL) in the 3′ UTR of the γ constant region (Fig. 1F). EMSA experiments showed the gel-shifted and supershifted complex with the HS5-1aL probe, demonstrating the specific binding of CTCF to HS5-1aL in vitro (Fig. S3A). In addition, mutation of the core CTCF consensus within HS5-1aL also abolishes the CTCF binding in vitro (Fig. S4 E and F). Finally, mutations of the CGCTG or CCCTC core sequences, which disrupt CTCF binding to CSE, result in a significant decrease of the promoter activity for α1, α4, α8, α12, αc1, β3, γb1, γb5, γa10, and γb7 in a luciferase reporter assay (Fig. S5). In summary, we show that CTCF binds directly to a repertoire of regulatory sequences within the Pcdh locus in vitro and in vivo and that this binding correlates with Pcdh gene expression.

Methylation of CSE DNA Blocks CTCF Binding in Vitro. DNA methylation in the Pcdh promoter regions was shown previously to correlate with their activities (3, 17, 18). The core sequences within CSE contain a CpG dinucleotide (2). Recent studies suggested that binding of CTCF to its target site is insensitive to methylation (15); however, CpG methylation has been shown to play an essential role in CTCF-dependent allele-specific expression regulation of imprinted genes (19, 20).

To investigate whether methylation of CpG within the CSE influences CTCF binding, we prepared probes with methylated CpG of both strands for α8, β3, γa2, and γb1, each representing the α, β, γa, and γb groups, respectively. EMSA experiments show that CpG methylation results in a dramatic decrease or complete absence of the CTCF binding (Fig. 2 A and B). In addition, we detected CpG hypomethylation of expressed isoforms and CpG hypermethylation of silenced isoforms in the promoter proximal regions between the two CTCF-binding sites. However, there is constitutive hypermethylation in the promoter distal regions downstream of the eCBS sites in both expressed and silenced isoforms (Fig. 2 C and D). These data, in conjunction with the correlation between Pcdh promoter activity and its hypomethylation in cultured cells, strongly suggest that CpG methylation within the promoter regions plays an important role in the regulation of CTCF-binding and promoter activity.

CpG methylation may also regulate Pcdh enhancer function. CTCF binds to HS5-1b in every cell line in the ENCODE dataset; however, CTCF binds HS5-1a in only a subset of these cell lines (14). Interestingly, HS5-1a contains a CpG dinucleotide at the position corresponding to the methylation-sensitive CTCF site in CSE, whereas HS5-1b does not contain a CpG site and, thus, cannot be methylated (Fig. S4C). This observation suggests that CTCF binds to HS5-1b constitutively but to HS5-1a in a methylation-sensitive manner. Interestingly, the first site of HS5-1aL contains a CpG dinucleotide at the same position corresponding to that in HS5-1a and CSE (Fig. S4E). Moreover, the CTCF binding to HS5-1aL appears to be regulated, probably in a methylation-sensitive manner, whereas CTCF binding to HS5-1bL appears to be constitutive (14).

DNA Looping Between Enhancers and Active Promoters. To investigate whether there are long-range DNA-looping interactions between enhancers and promoters of the α cluster, we performed quantitative chromosome conformation capture (3C) assays using SK-N-SH cells. We used two nonneuronal human cell lines K562 and 293T, which do not express Pcdhα isoforms, as controls in the 3C analyses (Fig. S6 A and B). When using an anchor primer

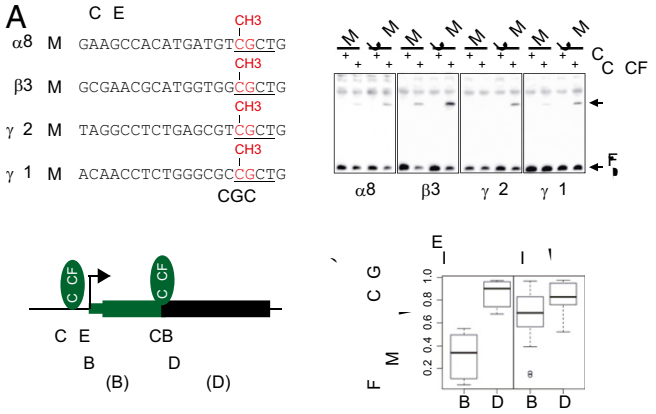

Fig. 2. CpG methylation regulates *Pcdh* gene expression through alteration of CTCF binding. (*A*) The methylated CSE sequences as representative probes. (*B*) EMSA experiments with methylated (Me) and unmethylated control (UMe) CSE probes. (*C*) Schematic of CTCF binding at each alternate promoter region of the *Pcdhα* cluster. (*D*) Inverse relationships between *Pcdh* gene expression and extent of CpG methylation (β score).

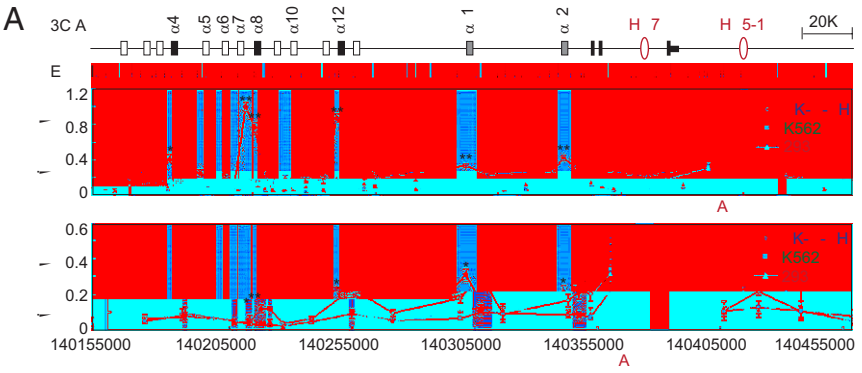

Fig. 3. The HS5-1 and HS7 enhancers engage in multiple long-range DNA-looping interactions with alternate and ubiquitous promoters. Relative cross-linking frequencies between enhancers HS5-1 (*A*) or HS7 (*B*) and promoters of the α cluster were measured by the 3C assay in SK-N-SH (blue), K562 (green), and 293T (red) cells. Data are presented as means ± SEM ($n = 3$). $^{*}P < 0.05$; $^{**}P < 0.01$. Only significance of comparison with K562 is shown. The *P* value of comparison with 293T is similar to that of K562 for all sites except *α4* ($P = 0.056$).

matching HS5-1, strong interactions were detected with the active promoters *α*8 and *α*12, whereas weak interactions were detected with the active promoters of *α*4, *α*c1, and *α*c2 (Fig. 3A). Importantly, interactions with the inactive promoters of *α*5, *α*6, *α*7, and *α*10 were not detected (Fig. 3A).

We note that there is an EcoRI site between the *α*8 CSE and eCBS, thus making it possible to separate the promoter region into two DNA fragments. Importantly, both fragments form strong DNA-looping interactions with HS5-1 (Fig. 3A). Interestingly, sequencing of the looped-DNA fragments between *α*8 and HS5-1 revealed that both alleles of the *α*8 isoform interact with HS5-1 in SK-N-SH cells (Fig. S6 C and D). In addition, sequencing of the *α*8 mRNA demonstrates that both alleles are expressed in SK-N-SH cells (Fig. S6G). We note that HS5-1 interacts more weakly with the ubiquitously expressed promoters of *α*c1 and *α*c2, consistent with the observation that only a minor decrease in the expression of these c-type genes was detected in HS5-1–deletion mice (9, 11).

When we used primers against HS7 as anchors, we observed weak interactions with the active promoters of *α*8, *α*12, *α*c1, and *α*c2, but not with *α*4, compared with the control nonneuronal cells of K562 or 293T (Fig. 3B), which do not express the α cluster (Fig. S6 A and B). Similar to the situation with the HS5-1 enhancer, sequencing of the looped-DNA fragments between *α*8 and HS7 revealed that both alleles of the *α*8 interact with HS7 in SK-N-SH cells (Fig. S6 C and E). It is important to note that HS7 does not interact with inactive promoters (Fig. 3B). Thus, the pattern of Pcdh*α* isoform expression observed in SK-N-SH cells strongly correlates with specific enhancer/promoter interactions. We conclude that long-distance (over 250 kb in some cases) DNA looping plays an important role in stochastic promoter choice.

CTCF and Cohesin Are Required for DNA Looping. To determine whether CTCF and cohesin are required for enhancer/promoter interactions, we infected SK-N-SH cells with a lentivirus expressing shRNAs directed against GFP (control), CTCF, or Rad21. The shRNA knockdowns led to a decrease in the levels of both CTCF and Rad21 mRNAs and proteins (Fig. S7 A–C). As expected, knockdown of CTCF by shRNA results in a significant decrease of CTCF binding to the CSE and eCBS of *α*4, *α*8, and *α*12; the CSE of *α*c1, *β*3, *γ*b1, *γ*b7, and *γ*c3; and the enhancer sites of HS5-1a and HS5-1b, as well as HS5-1aL in SK-N-SH cells (Fig. S8A).

ChIP-qPCR experiments demonstrated that knockdown of CTCF leads to a significant decrease in the levels of the cohesin subunits Rad21 and SMC3, which bind to the CSEs of *α*4, *α*8, *α*12, and *α*c1, as well as to HS5-1a and HS5-1b (Fig. S7 D and E). Similarly, knockdown of Rad21 results in a significant decrease in the binding of Rad21 and SMC3 to these sites (Fig. S7 F and G). Most importantly, knockdown of CTCF or Rad21 results in a significant decrease in the long-range DNA-looping interactions between HS5-1 and the *α*8 or *α*12 promoter (Fig. 4). In addition, CTCF knockdown decreases Rad21 enrichment at the eCBS sites of *α*8 and *α*12, as well as CSEs of *β*3, *γ*b1, *γ*b7, *γ*c3, and HS5-1aL in the β and γ clusters, suggesting that cohesin is recruited by CTCF to these sites (Fig. S8 B and C) (21). We conclude that the binding of CTCF/cohesin to active promoters and enhancers is required for Pcdh*α* enhancer/promoter interactions through DNA looping.

Interaction of Paired CTCF/Cohesin-Binding Sites. We investigated DNA-looping interactions between the pair of CTCF-binding sites in the Pcdh*α*8 promoter region (CSE and eCBS) and the pair of binding sites in the HS5-1 enhancer region (HS5-1a, HS5-1b) in SK-N-SH cells. Based on the PstI-digestion patterns, we designed two specific primers, P1 and P2, corresponding to the two fragments containing the CSE and eCBS of *α*8, respectively. In addition, we designed two primers, P3 and P4, corresponding to the two fragments containing HS5-1a and HS5-1b, respectively. We also designed two primers, P5 and P6, complementary to sequences outside of HS5-1 as negative controls (Fig. 5A).

The 3C experiments using the primers P1 and P3 or P4 demonstrated significant DNA-looping interactions between the CSE of *α*8 to both HS5-1a and HS5-1b (Fig. 5B). Similarly, significant DNA-looping interactions were observed between the eCBS of *α*8 and both HS5-1a and HS5-1b (Fig. 5C), as demonstrated with primers P2 and P3 or P4. Control PCR experiments with BAC DNA preparations showed that these primers amplify predicted products with similar efficiency (Fig. 5 B and C). ChIP-qPCR experiments with four pairs of specific primers matching the CSE, eCBS, HS5-1a, and HS5-1b sites and six pairs of intervening and flanking primers confirmed the binding specificity of CTCF and Rad21 to these four sites in SK-N-SH (Fig. S9 A and B). The conservation of the position and spacing of the CSE/eCBS sites on the one hand (about 650 bp) and the HS5-1a,b on the other (1,009 bp) and the DNA-looping data of Fig. 5 suggest that DNA looping between the HS5-1 enhancer and active alternate promoters involves a "double clamp" in which the four CTCF/cohesin sites interact simultaneously (Fig. 5D).

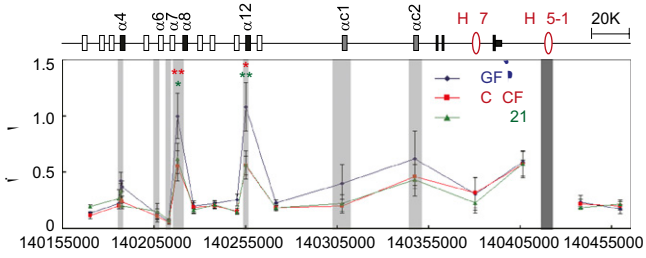

Fig. 4. CTCF or cohesin knockdown results in significant decreases of long-range DNA interactions between HS5-1 and promoters. SK-N-SH cells were infected with shGFP (control), shCTCF, or shRad21 lentivirus, and crosslinking frequencies were measured by 3C. Data are presented as means ± SEM ($n = 4$). $^{*}P < 0.05$; $^{**}P < 0.01$.

in a similar manner to that observed for the HS7, HS5-1a, and HS5-1b and variable promoters of the α cluster. The variable region of the human γ cluster contains 19 alternate promoters (12 a-type: γa1 to γa12; and 7 b-type: γb1 to γb7) and 3 c-type (γc3 to γc5) ubiquitous promoters. Similar to the α cluster, each of the 19 alternate promoters and the first ubiquitous promoter (γc3) of the γ cluster contains a CSE (2) that binds to CTCF (Fig. 1F and Fig. S3). The last two c-type ubiquitous promoters (γc4 and γc5), similar to the last ubiquitous α promoter (αc2), do not contain a CSE and cannot bind to CTCF (Fig. 1C and Fig. S3A). Similar to HS7 of the α cluster, we found an HS7-like (HS7L)

Discussion

The remarkable genomic organization of the
cadherin genes (1, 2, 4, 5) and the generation of
through promoter choice and alternative pre
provide enormous single-cell diversity (3, 6
appears to play a fundamental role in neural circ
25). Understanding the mechanism by which
generated is, therefore, fundamental to under
velopment and function of the nervous system.
have identified Pcdhα isoform promoters and e
(2–5, 8) and have shown that they bind to CTC
10). However, neither the function of CTCF/
possible role in enhancer/promoter interaction
Here, we made use of the diploid human ne
line SK-N-SH to address these questions. We
and cohesin are required for the long range
enhancers and active isoform promoters, an
methylation of CpG dinucleotides in the CS
binding of CTCF to the CSE in vitro. We find th
the HS5-1 enhancer form long-range DNA-loo
with two sites in the active promoters, α8 and α1
interactions with promoters of α4, αc1, and αc2
SH cells. These interactions correlate with th
HS5-1 for alternate promoter activity in mice (b
and qualitatively), except for αc2, which does n
for maximal activity (9, 11). Interestingly, the m
tous promoter in the cluster, αc2, does not conta
is not recognized by CTCF (Fig. S3A). The HS7 e
located between constant exons 2 and 3 (8), is a
CTCF (10) but is enriched in RNAPII (Fig. 1C
αc2 and HS7 are bound to Rad21. We found
hancer forms long-range DNA-looping interact
α12, αc1, and αc2 promoters in human SK-N
again, is consistent with requirement of HS7
ubiquitous promoter activity in mouse genetic
there appears to be CTCF- and cohesin-depend
(α1 to α13 and αc1) and CTCF-independent
pendent DNA looping (αc2). Taken together,
vide conclusive evidence that DNA looping occ
characterized enhancers and multiple active Pcdh
α cluster, and that this looping requires specific
cohesin to the transcriptionally active promoter

We propose that CTCF/cohesin-dependent
looping interactions among the putative enhanc
1aL, HS5-1bL) of the γ cluster and its variable

(22–25). Strikingly, a repertoire of heterozygous mutations in chromatid-segregating cohesin (SMC1A, SMC3, RAD21) and cohesin regulators (NIPBL, ESCO2) have been found to cause a class of developmental disorders known as Cornelia de Lange syndrome (CdLS) and Robert syndrome in humans (31). One of the most intriguing unknown aspects of these syndromes is the molecular etiology of CdLS in neurodevelopmental delay and mental retardation. The important role of cohesin in the regulation of clustered Pcdh genes (this study and refs. 10, 32, and 33) suggests a neuropathologic basis for these syndromes caused by human heterozygous cohesin mutations.

The observation that both the HS5-1 and HS7 enhancers interact with and are required for maximal expression of multiple members of the α cluster and that CTCF/cohesin is required for this interaction suggests a complex mechanism of promoter choice. We propose that the DNA-looping interactions recruit the promoters bound to CTCF to enhancers in an active "transcriptional hub" (Fig. 6). This model is based on extensive and highly specific functional and physical interactions between promoters and enhancers, and the fact that the formaldehyde cross-linking used in the ChIP-seq and 3C studies would be expected to crosslink a large complex containing multiple enhancers and promoters. In this model, the DNA-looping interactions between HS5-1 and the promoters of α8 and a12 are formed by a double-clamping mechanism between the HS5-1a/HS5-1b sites of the enhancer and the CSE/eCBS sites of alternate promoters (Fig. 5D). At the same time, HS5-1 must interact with αc1, because HS5-1 is required for its expression (9, 11). In addition, HS7 must directly interact with αc2, as well as with the active alternate promoters and ubiquitous promoters, because this enhancer interacts with and is required for the maximum activation of these promoters (9). Finally, long-range DNA-looping interactions are formed between promoters of αc1 and α8 and both alleles of the α8 promoters form looping interactions with αc1 (Fig. S6 C and F). All of these observations are consistent with the model of Fig. 6, in which the apparent simultaneous interactions between the two enhancers and multiple promoters are the consequence of the formation of a large transcriptional hub. Presumably, promoter/enhancer interactions within this hub must be maintained during DNA replication, because the pattern of Pcdh expression in SK-N-SH cells has been maintained during many rounds of cell division. Moreover, the stability of this complex is required to maintain self-identity during the postmitotic life of individual neurons. Further studies of the mechanisms of Pcdh promoter choice during development and whether and how the choice is maintained should provide important insights into the role of Pcdh diversity in the assembly of neural circuits.

Materials and Methods

Recombinant Protein Production and EMSA. Full-length human CTCF cDNA was cloned from total RNA preparations of HEC-1-B by RT-PCR with a pair of specific primers and confirmed by sequencing. The recombinant human CTCF proteins were synthesized from pTNT-CTCF by using TNT T7 System (Promega). All of wild-type (WT) and mutation probes were confirmed by sequencing. See *SI Materials and Methods* for details.

Methylation Analysis. To determine CpG methylation state in SK-N-SH cells, genomic DNA was analyzed on Infinium HumanMethylation450 BeadChips (Illumina). CpGs located near each variable exon were identified and grouped by location.

ChIP-qPCR and ChIP-seq. ChIP was performed by using protein A agarose beads from the Millipore and followed by standard qPCR experiments. The ChIP-seq experiments were performed similarly, except that the precipitated compleW6.7335.iGo84e ChIP-isol-205.158lipor-438.957andCh5rose*Ful3351711c9999241d*