

# Outward-oriented sites within clustered CTCF boundaries are key for intra-TAD chromatin interactions and gene regulation

Received: 9 May 2023

Accepted: 21 November 2023

Published online: 07 December 2023

 Check for updatesXiao Ge<sup>1,2,4</sup>, Haiyan Huang<sup>1,2,4</sup>, Keqi Han<sup>1,2</sup>, Wangjie Xu<sup>3</sup>, Zhaoxia Wang<sup>3</sup> & Qiang Wu<sup>1,2</sup>  

CTCF plays an important role in 3D genome organization by adjusting the strength of chromatin insulation at TAD boundaries, where clustered CBS (CTCF-binding site) elements are often arranged in a tandem array with a complex divergent or convergent orientation. Here, using *Pcdh* and *HOXD* loci as a paradigm, we look into the clustered CTCF TAD boundaries and find that, counterintuitively, outward-oriented CBS elements are crucial for inward enhancer-promoter interactions as well as for gene regulation. Specifically, by combinatorial deletions of a series of putative enhancer elements in mice in vivo or CBS elements in cultured cells in vitro, in conjunction with chromosome conformation capture and RNA-seq analyses, we show that deletions of outward-oriented CBS elements weaken the strength of long-distance intra-TAD promoter-enhancer interactions and enhancer activation of target genes. Our data highlight the crucial role of outward-oriented CBS elements within the clustered CTCF TAD boundaries in developmental gene regulation and have interesting implications on the organization principles of clustered CTCF sites within TAD boundaries.

The interphase genome is organized into highly dynamic structures including chromosome territory, compartment, topologically associated domain (TAD), and chromatin loop<sup>1</sup>. CTCF (CCCTC-binding factor) is a key architecture protein for interphase 3D genome organization and CBS (CTCF-binding site) elements throughout the mammalian genomes function as topological insulators to block aberrant enhancer activation and ensure proper gene expression, or to mark the boundaries between euchromatin and heterochromatin<sup>2–4</sup>. A prevailing model for interphase chromosome folding, known as loop extrusion, posits that the cohesin complex extrudes chromatin fibers bidirectionally into expanding loops until blocked asymmetrically by orientated CBS elements<sup>5–7</sup>. Specifically, continuous ATP-driven active loop extrusions lead to gradually larger-sized chromatin loops which

are stabilized by pairs of convergent or opposite CBS elements. This model explains TAD formation and is consistent with the observations that the two flanking boundaries of each TAD are associated with mostly convergent forward-reverse CBS elements<sup>8,9</sup>. However, it is puzzling that most TAD boundaries in mammalian genomes contain clustered CBS elements with complex orientations of both convergence and divergence.

TAD boundaries emerged as insulator elements that restrain inter-TAD chromatin interactions. Their dysfunction leads to abnormal development and progressive oncogenesis<sup>10–12</sup>. Mammalian clustered CTCF TAD boundaries are more evolutionarily conserved than other genomic regions<sup>13–15</sup>. High insulation score of TAD boundaries is associated with strong CTCF enrichment, and clustered enhancers,

<sup>1</sup>Center for Comparative Biomedicine, Ministry of Education Key Laboratory of Systems Biomedicine, State Key Laboratory of Systems Medicine for Cancer, Joint International Research Laboratory of Metabolic and Developmental Sciences, School of Life Sciences and Biotechnology, Institute of Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, China. <sup>2</sup>WLA Laboratories, Shanghai 201203, China. <sup>3</sup>Laboratory Animal Center, Instrumental Analysis Center, Shanghai Jiao Tong University, Shanghai 200240, China. <sup>4</sup>These authors contributed equally: Xiao Ge, Haiyan Huang.

✉ e-mail: [qiangwu@sjtu.edu.cn](mailto:qiangwu@sjtu.edu.cn)

known as super-enhancers, tend to be insulated by strong boundaries<sup>16,17</sup>. Divergent CBS elements are enriched at TAD boundaries; however, a divergent orientation signature is not strictly required for effective insulation. For example, the characteristics of specific CBS elements can outweigh CBS number and orientation<sup>18</sup>. Finally, insulation potency depends on the local context of CBS elements and robust enhancer-promoter communications can bypass CTCF insulation regardless of the intervening CBS strength<sup>19–22</sup>.

As a paradigm for investigating mechanisms of 3D genome folding, the mouse clustered protocadherins (*cPcdh*) comprise three sequential gene clusters, *Pcdh*  $\alpha$ ,  $\beta$ , and  $\gamma$ , encoding 58 isoforms and spanning -1 million bps (Fig. 1a)<sup>23</sup>. The *Pcdh*  $\alpha$  and  $\gamma$  clusters consist of 14 and 22 variable first exons, respectively, each of which is *cis*-spliced to a single set of cluster-specific constant exons; whereas the *Pcdh*  $\beta$  cluster consists of 22 single variable exons ( $\beta 1$ – $\beta 22$ ) with no constant exon. The *Pcdh* promoter choice is determined by CTCF/cohesin-mediated long-range enhancer-promoter looping interactions anchored by convergent or opposite CBS elements<sup>9,24,25</sup>. Their combinatorial expression patterns in single neurons generate enormous cell-surface molecular diversity for assembling complex networks of neuronal connectivity in the brain<sup>26,27</sup>. In addition, precise expression patterns of members of the *HoxD* gene cluster are central for limb development<sup>28,29</sup>. Here, using *Pcdh* and *HOXD* as model genes, we find that outward-oriented CBS elements within clustered CTCF boundaries are crucial for intra-TAD promoter-enhancer interactions and gene regulation. Our data shed interesting light on the organizational principles of the clustered CTCF TAD boundaries in 3D genome architecture and gene regulation.

## Results

### The downstream boundary of *cPcdh* superTAD comprises a cluster of CTCF sites with mixed orientations

During mouse neocortical development, the 58 *cPcdh* genes are organized into a large megabase-sized superTAD comprised of the *Pcdh* $\alpha$  and *Pcdh* $\beta$  TADs (Fig. 1a; Supplementary Fig. 1a–c)<sup>9,30</sup>. At the telomeric boundary of the *cPcdh* superTAD, there is a complex array of eight CTCF sites or CBS elements (*CBS a–h*, Fig. 1a–c). Among these eight CBS elements, *CBS a–e*, *CBS g*, and *CBS h* are in the reverse orientation (Supplementary Fig. 1d). By contrast, a single *CBS f* is in the forward orientation (Supplementary Fig. 1e). We performed chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) experiments with a specific antibody against CTCF using microdissected mouse neocortical tissues and found that CTCF is enriched at these elements (Fig. 1c). Interestingly, the *CBS b* and *c* elements form a single peak and *CBS f* and *g* elements form a single peak (Fig. 1c). To see whether these elements are anchors of cohesin loop extrusion, we performed ChIP-seq experiments with a specific antibody against Rad21, a subunit of cohesin, and found that cohesin is co-localized with CTCF at all of these CBS elements, suggesting that cohesin loop extrusion from this boundary may regulate *cPcdh* gene expression (Fig. 1b, c).

We first comprehensively mapped expression patterns of each member of the three *Pcdh* gene clusters at a series of 12 developmental stages by collecting embryos every other day and new pups every other 2 days. RNA-seq of microdissected neocortical tissues showed that each member of the three *Pcdh* gene clusters is expressed dynamically and that the *cPcdh* genes are most abundantly expressed at the newborn stage (Supplementary Fig. 2). Thus, we used the newborn mouse neocortical tissues in the following gene regulation studies.

Within the *Pcdh* $\alpha$  TAD, members of the *Pcdh* $\alpha$  cluster are activated by a super-enhancer comprised of *HSS-1* and *HS7* (Fig. 1b and Supplementary Fig. 3a, b)<sup>24,31,32</sup>. However, how members of the *Pcdh* $\beta$  clusters are regulated within the *Pcdh* $\beta$  TAD is not clear. To this end, we profiled the chromatin regulatory landscape by assaying for transposase-accessible chromatin with high-throughput sequencing

(ATAC-seq) and identified six highly accessible regions at the telomeric boundary of the *Pcdh* $\beta$  TAD, corresponding to the previously identified DNaseI hypersensitive sites of *HS7L*, *HSS-1aL*, *HSS-1bL*, *HS18*, *HS19-20*, and *HS21* (Fig. 1c and Supplementary Fig. 3c, d)<sup>24,33</sup>. Finally, we performed H3K27ac and H3K4me1 ChIP-seq and found, except the last one (*HS21*), all of the other five clustered *HS* elements are marked by H3K27ac and H3K4me1, suggesting that they form a super-enhancer for members of the *Pcdh* $\beta$  gene clusters (Fig. 1c).

### The *Pcdh* $\beta$ TAD boundary comprises a super-enhancer in mice in vivo

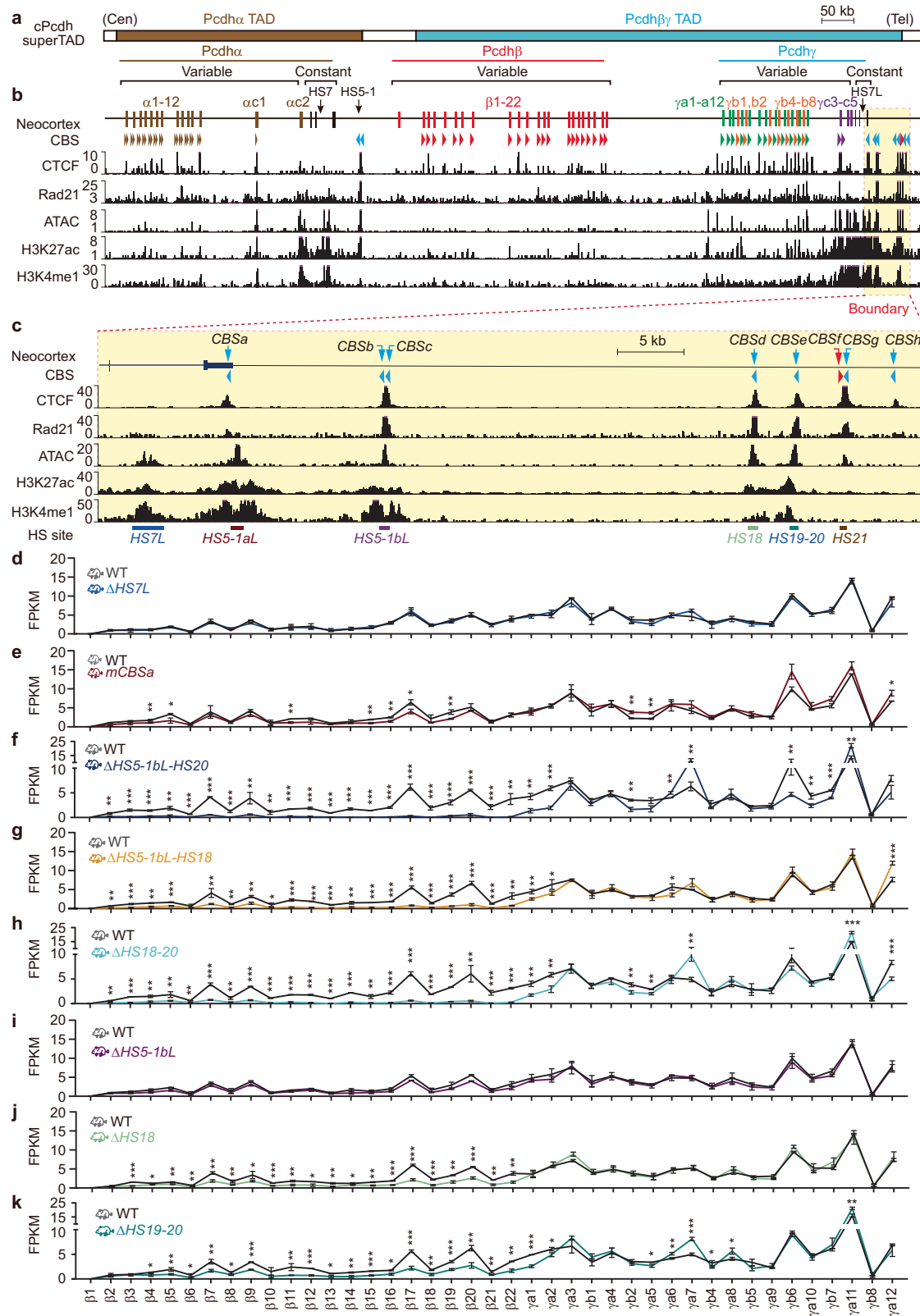
We dissected the putative enhancers in mice in vivo by CRISPR/Cas9-based DNA-fragment editing system (Supplementary Fig. 4a–c)<sup>9</sup> to test their functionality. We first deleted the *HS7L* element (Supplementary Fig. 4b, c) and found, surprisingly, it does not lead to decreased levels of *cPcdh* gene expression in neocortical and olfactory tissues (Fig. 1d, Supplementary Fig. 4e–g, and Supplementary Fig. 5a). Because *HSS-1aL* is in the *Pcdh* $\gamma$  constant coding region and its deletion will destroy the transcript integration, we mutated the *CBS a* element within *HSS-1aL* by homologous recombination using CRISPR with an ssODN donor. We found that *CBS a* mutation or deletion abolished CTCF enrichments (Supplementary Fig. 4d). Interestingly, it only results in a moderate decrease of expression levels of members of the *Pcdh* $\beta$ , but not  $\alpha$ , clusters (Fig. 1e, Supplementary Figs. 4e, 5b), suggesting a limited role of *CBS a* in tethering *HSS-1aL* to *Pcdh* $\beta$ .

We next deleted the entire region covering all of the other three putative enhancers (from *HSS-1bL* to *HS19-20*, *CBS b–e*) in mice and found that, in contrast to *HS7* and *HSS-1aL*, this large deletion leads to a significant decrease of levels of *Pcdh* $\beta$  gene expression, suggesting that enhancers for members of *Pcdh* $\beta$  clusters reside in this region (Fig. 1f and Supplementary Fig. 5c). We then dissected this region in great details and found that deletion of the large DNA fragment covering both *HSS-1bL* and *HS18* (*CBS b–d*) results in a significant decrease of levels of *Pcdh* $\beta$  gene expression (Fig. 1g and Supplementary Fig. 5d). In addition, deletion of the DNA fragment covering both *HS18* and *HS19-20* (*CBS d–e*) also results in a significant decrease of levels of *Pcdh* $\beta$  gene expression (Fig. 1h and Supplementary Fig. 5e).

Finally, we deleted each of these three elements individually in mice and found that, surprisingly, deletion of *HSS-1bL* (*CBS b–c*) has no effect on *Pcdh* $\beta$  gene expression (Fig. 1i and Supplementary Fig. 5f). By contrast, deletion of either *HS18* (*CBS d*) or *HS19-20* (*CBS e*) result in a significant decrease of levels of *Pcdh* $\beta$  gene expression (Fig. 1j, k and Supplementary Fig. 5g, h). These data suggest that these two H3K27ac-enriched elements play a leading role in activating *Pcdh* $\beta$  genes in vivo in the mouse brain. Consistently, 4C experiments revealed close contacts between *Pcdh* $\beta$  and *HS18-20* (Supplementary Fig. 6a, b). Unlike *HSS-1*, which is brain-specific, *HS18-20* are enriched with H3K27ac and H3K4me3 marks in multiple tissues (Supplementary Fig. 6c, d), explaining why members of the *Pcdh* $\beta$  gene clusters are expressed in both neural and non-neural tissues. Taken together, these data suggest that the downstream boundary of the *Pcdh* $\beta$  TAD is a super-enhancer comprising a cluster of enhancers among which *HS18* and *HS19-20* are the most important.

### *HSS-1bL* is a *Pcdh* $\gamma$ 3-specific insulator in mice in vivo

*Pcdh* $\gamma$ 3 is the most abundantly expressed isoform in the mouse neocortex and has a unique role in dendrite arborization (Supplementary Fig. 2e)<sup>34</sup>. However, the underlying mechanism remains unclear. To this end, we performed QHR-4C experiments and found that *Pcdh* $\gamma$ 3 is in close contacts with *HSS-1bL*, *HS18*, and *HS19-20* (Fig. 2a). Consistently, deletion of *HS18* and/or *HS19-20* results in a significant decrease of levels of *Pcdh* $\gamma$ 3 gene expression, suggesting that *HS18* and *HS19-20* are enhancers for *Pcdh* $\gamma$ 3 (Fig. 2b–d). Surprisingly, deletion of *HSS-1bL* in mice in vivo leads to a significant increase of levels of *Pcdh* $\gamma$ 3, but not  $\gamma$ 4 or  $\gamma$ 5, gene expression



(Fig. 2e), suggesting that *HSS-1bL* is a putative insulator specifically blocking the activation of *Pcdhyc3* by the *HS18* and *HS19-20* enhancers. To this end, we performed 4C experiments with *Pcdhyc3* as an anchor using neocortical tissues of the *HSS-1bL* deleted mice and found a significant increase of long-distance chromatin interactions between *Pcdhyc3* and *HS18-20* (Fig. 2f). To confirm this aberrantly increased

chromatin contacts upon deletion of *HSS-1bL*, we performed 4C with *HS18* or *HS19-20* as an anchor and indeed observed increased chromatin contacts with *Pcdhyc3* (Fig. 2g). In conjunction with no expression increase of all other members of the *Pcdhβy* gene clusters upon deletion of *HSS-1bL* (Fig. 1i), we conclude that *HSS-1bL* is a  $\gamma$ 3-specific insulator.

**Fig. 1 | The downstream *cPcdh* superTAD boundary comprises clustered enhancers for *Pcdhβγ* gene regulation.** **a** Schematics of the mouse *cPcdh* locus. The mouse *Pcdh*  $\alpha$ ,  $\beta$ , and  $\gamma$  gene clusters are closely-linked within the *cPcdh* superTAD, which is divided into two TADs of *Pcdh $\alpha$*  and *Pcdh $\beta\gamma$* . *Pcdh*  $\alpha$  and  $\gamma$  gene clusters contain 14 and 22 variable first exons, respectively, each of which is alternatively spliced to a single set of three downstream constant exons. The *Pcdh $\beta$*  gene cluster contains 22 variable exons but with no constant exon. Two *Pcdh $\alpha$*  enhancers, *HS5-1* and *HS7*, are indicated by vertical black arrows. **b** ChIP-seq of CTCF, Rad21, H3K27ac, and H3K4me1 as well as ATAC-seq profiles of the *cPcdh* locus in the mouse neocortex. Arrowheads indicate CBS elements with orientations. Within the *Pcdh $\alpha$*  TAD, each *Pcdh $\alpha$*  alternate promoter is flanked by two forward CBS elements, the downstream enhancer of *HS5-1* is flanked by two reverse CBS elements. Within the *Pcdh $\beta\gamma$*  TAD, each *Pcdh*  $\beta$  or  $\gamma$  promoter is associated with one forward CBS, except for  $\beta 1$ ,  $\gamma 4$ , and  $\gamma 5$  which have no CBS, as well as  $\gamma 3$  which has two

forward CBS elements. **c** Close-up of the *cPcdh* superTAD downstream boundary marked by a red dotted rectangle in **(a, b)**. This boundary comprises eight CBS elements, of which only *CBSf* is in the reverse orientation, and six ATAC-seq peaks, of which *HS7L*, *HS5-1bL*, *HS18*, and *HS19-20* are knocked out individually or in combinations in mice. *HS5-1aL* is mutated at the *CBSa* motif (*mCBSa*) due to its location in the coding region. **d–k** Expression levels of members of the *Pcdh*  $\beta$  and  $\gamma$  gene clusters in the mouse neocortex from  $\Delta$ *HS7L*, *mCBSa*,  $\Delta$ *HS5-1bL-HS20*,  $\Delta$ *HS5-1bL-HS18*,  $\Delta$ *HS18-20*,  $\Delta$ *HS5-1bL*,  $\Delta$ *HS18*, or  $\Delta$ *HS19-20* homozygous mice compared to their wild-type (WT) littermates. FPKM, fragments per kilobase of exon per million fragments mapped. Data as mean  $\pm$  SD; two-tailed Student's *t* test. \**p* < 0.05, \*\**p* < 0.01, \*\*\**p* < 0.001. For each mouse line, *n* = 4 biologically independent samples; for their WT littermate controls, *n* = 2. Source data are provided as a Source Data file.

### Clustered *Pcdhs* are expressed from H3K9me3-enriched facultative heterochromatins of superTAD

We next mapped the chromatin landscape of a wide variety of histone modifications in the *cPcdh* superTAD and found that this region is enriched for both the repressive mark of H3K9me3 and the active mark of H3K36me3 (Fig. 3a), suggesting that *cPcdhs* are expressed from facultative heterochromatin region of the superTAD. In addition, expression of each member of the three *Pcdh* gene clusters is strongly correlated with active marks of H3K4me3 and H3K9ac (Fig. 3b and Supplementary Figs. 7, 8). Moreover, this region is also marked by the repressive histone marks of H3K9me2 in the mouse neocortex (Fig. 3c and Supplementary Fig. 9). Specifically, the coding region of each member of the *Pcdh* clusters is correlated with a peak of H3K9me2 (Supplementary Fig. 9) but not H3K9me3 (Supplementary Fig. 10). To further investigate mechanisms of *cPcdh* gene regulation, we analyzed methylated DNA immunoprecipitation sequencing (MeDIP-seq) data and found that each member of the H3K9me3-enriched *cPcdhs* is hypermethylated on the body of each variable exon (Fig. 3c, Supplementary Fig. 11a–d). We noted that *Pcdh $\beta 1$*  differs from other *cPcdh* members in that its DNA-methylation peak is much closer to the transcription start site (Supplementary Fig. 11e), explaining why it is silenced in the mouse neocortex (Fig. 3b).

### Outward-oriented *CBSf* is the key boundary CTCF site

All of the *HS5-1aL*, *HS5-1bL*, *HS18*, and *HS19-20* enhancers contain reverse-oriented CBS elements (*CBSa–e*, Fig. 1c) that are enriched for cohesin and are in close contacts with the forward-oriented CBS elements of the *Pcdh $\beta\gamma$*  promoters via CTCF/cohesin-mediated chromatin loops. In addition, there are two more reverse-oriented CBS elements, *CBSg* and *CBSH*, located immediately downstream of this super-enhancer. However, there is a single forward-oriented *CBSf* element located between the super-enhancer and the two reverse-oriented *CBSg* and *h* elements (Fig. 4a). To this end, we generated a *CBSf* deletion mouse line (Fig. 4a and Supplementary Fig. 12a) and confirmed the abolishment of CTCF and cohesin enrichments at the location of *CBSf* but not *CBSg* (Fig. 4a, b). In addition, CTCF and cohesin binding at other CTCF sites within this boundary is not altered by *CBSf* deletion (Supplementary Fig. 12b).

Remarkably, deletion of the forward-oriented *CBSf* element results in a significant decrease of expression levels of members of the *Pcdh $\beta\gamma$*  gene clusters (Fig. 4c and Supplementary Fig. 13a), demonstrating that the single forward-oriented *CBSf* plays a crucial role in the regulation of *Pcdh $\beta\gamma$*  gene expression. To investigate the underlying mechanism, we performed a series of QHR-4C experiments with members of the *Pcdh $\beta\gamma$*  gene clusters as anchors and found a significant decrease of long-distance chromatin contacts between the super-enhancer and members of the *Pcdh $\beta\gamma$*  gene clusters (Fig. 4d). In particular, we observed a significant increase of long-distance chromatin contacts with *CBSH* with a sharp edge at the location of *CBSf* (Fig. 4e). To confirm this, we performed reciprocal capture

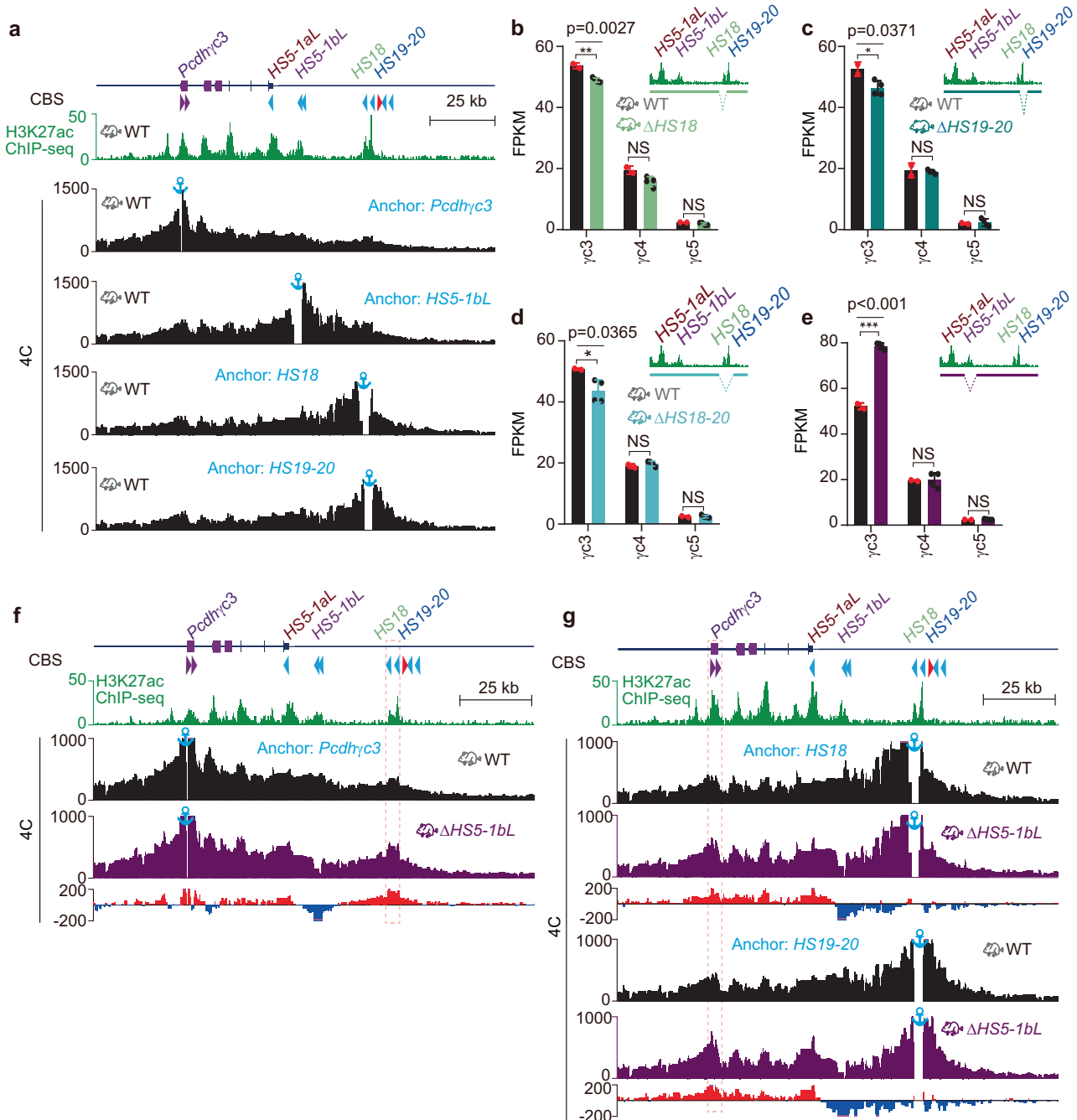
experiments with *CBSH* as an anchor and observed a significant increase of long-distance chromatin contacts with members of the *Pcdh $\beta\gamma$*  gene clusters (Supplementary Fig. 12c). In addition, we profiled the histone modifications and found decrease in enrichments of repressive marks, H3K9me3 and H3K9me2, but not of active marks, H3K36me3, H3K4me3, and H3K9ac (Supplementary Figs. 14–17).

Finally, combined deletion of both *CBSf* and *CBSg* elements (Supplementary Fig. 12d) results in a similar phenotype (Supplementary Figs. 18, 12e, 13b). As a control, single deletion of *CBSg* results in a less decrease in *Pcdh $\beta\gamma$*  expression (Supplementary Figs. 12f–h and 19). Taken together, we conclude that a single outward-oriented *CBSf* element within the clustered CTCF TAD boundary plays a key role in *Pcdh $\beta\gamma$*  gene regulation and long-distance intra-TAD chromatin interactions.

### Outward-oriented CBS elements within a clustered CTCF TAD boundary are crucial for *HOXD13* expression

To see whether it is a general phenomenon for a key role of the outward-oriented boundary CBS element in intra-TAD chromatin interactions and gene regulation, we investigated the *HOXD* left boundary, which has a complex array of six conserved clustered CBS elements (*CBSI-6*, Supplementary Fig. 20a–d), of the human centromeric TAD (C-DOM) (Fig. 5a, b)<sup>28,29</sup>. The human *HOXD* locus is located at the boundary between C-DOM and T-DOM (telomeric TAD) (Supplementary Fig. 21a, b). It contains nine *HOXD* genes, of which only *HOXD13* is associated with a cluster of reverse-oriented CBS elements (*CBS8-12*, Fig. 5c). By contrast, *HOXD8* and *HOXD9* have two forward-oriented CBS elements (*CBS13,14*) and *HOXD10-12* and *HOXD1-4* have no CBS element (Fig. 5a–c and Supplementary Fig. 20a). We analyzed Hi-C and found that there are strong long-distance chromatin interactions between the human *HOXD13* and the left boundary of C-DOM, consistent with the convergent rule of forward-reverse CBS elements (Supplementary Fig. 21a, b). Interestingly, within the left boundary, there are two outward-oriented CBS elements (*CBS3* and *CBS5*, Fig. 5b and Supplementary Fig. 21b).

We screened single-cell clones for deletion of both *CBS3* and *CBS5* elements and obtained three homozygous double-knockout cell clones ( $\Delta$ *CBS3* + 5, Supplementary Fig. 20g). We first confirmed the loss of CTCF enrichments at both *CBS3* and *CBS5* (Fig. 5a, b). We then performed 4C with *HOXD13* as an anchor and found that there is a significant increase of chromatin interactions with the left boundary upon deletions of the two outward-oriented CBS elements (Fig. 5d). These increased chromatin interactions were confirmed by reciprocal 4C experiments with *CBS2* or *CBS4* as an anchor (Fig. 5e, f). By contrast, there is a significant decrease of chromatin interactions with proximal sequences (Fig. 5d), which is confirmed by reciprocal 4C experiments with *GT2* or *LNPK* as an anchor (Supplementary Fig. 21c, d). In addition, we performed RNA-seq experiments and found that deletion of both outward-oriented CBS elements results in a significant decrease of expression levels of *HOXD13* but not *HOXD12-1*, indicating that the two



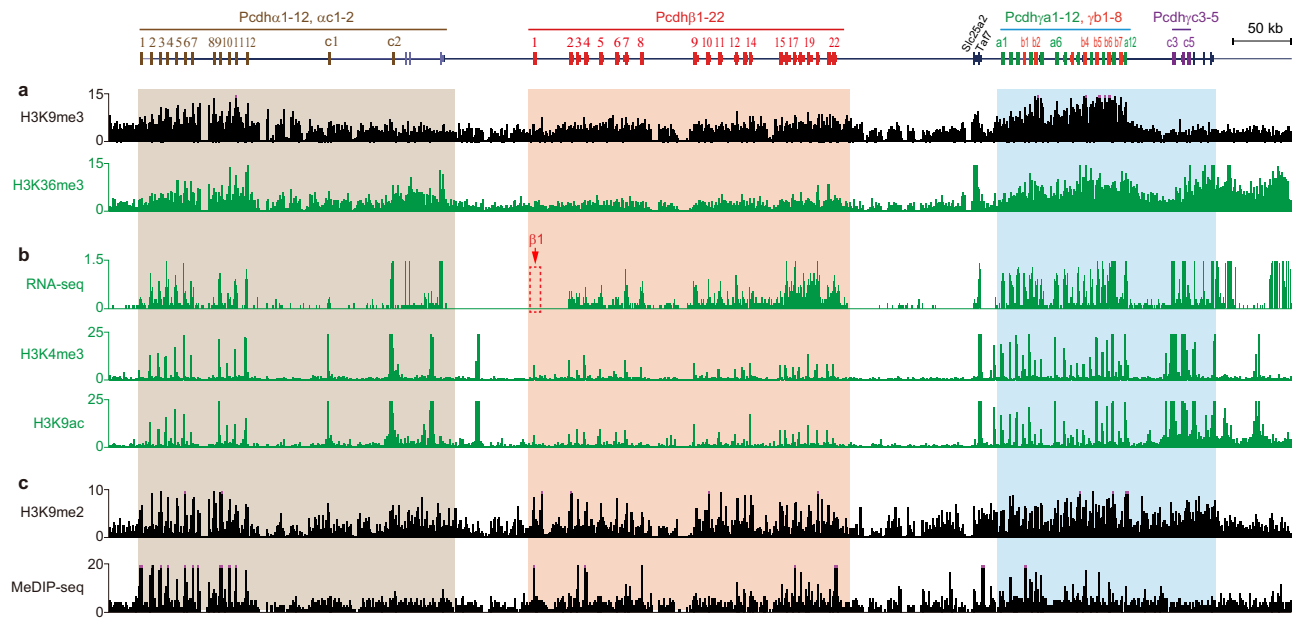
**Fig. 2 | *HSS-1bL* is an insulator specific for *Pcdhyc3*.** **a** 4C profiles using a repertoire of elements as anchors showing close contacts between *Pcdhyc3* and *HSS-1bL*, *HS18*, or *HS19-20* in wild-type (WT) mice. **b–e** Expression levels of the *Pcdhy* c-type genes in the mouse neocortex from  $\Delta$ *HS18* (**b**),  $\Delta$ *HS19-20* (**c**),  $\Delta$ *HS18-20* (**d**), or  $\Delta$ *HS5-1bL* (**e**) homozygous mice compared to their wild-type littermates. Note the significant increase of expression levels of *Pcdhyc3* upon deletion of *HSS-1bL*. Data as mean  $\pm$  SD; two-tailed Student's *t* test. \**p* < 0.05, \*\**p* < 0.01, \*\*\**p* < 0.001. For each mouse line, *n* = 4 biologically independent samples; for their WT littermate

controls, *n* = 2. Source data are provided as a Source Data file. **f** 4C profiles using the *Pcdhyc3* promoter as an anchor, showing increased chromatin interactions between *Pcdhyc3* and its *HS18-20* enhancers in  $\Delta$ *HS5-1bL* homozygous mice compared to their WT littermates. Differences ( $\Delta$ *HS5-1bL* versus WT) are shown under the 4C profiles. **g** 4C profiles using *HS18* or *HS19-20* as an anchor, showing increased chromatin interactions between *Pcdhyc3* and *HS18-20* in  $\Delta$ *HS5-1bL* mice compared to their WT littermates. Differences ( $\Delta$ *HS5-1bL* versus WT) are shown under the 4C profiles.

outward-oriented CBS elements play a key role in *HOXD13* gene regulation (Fig. 5g). However, different from the *Pcdh $\beta$*  genes which are associated with CBS elements, there is no CBS element associated with the LNPk gene within C-DOM and its expression is not altered upon deletion of outward-oriented CBS elements (Fig. 5g). Finally, single deletions of the *CBS3* or *CBS5* element demonstrated a cooperation role between the two outward-oriented CBS elements in *HOXD13* gene regulation (Supplementary Fig. 22).

## Discussion

The enormous connectivity of billions of neurons is intricately related to the precise cell-specific expression patterns of *cPcdh* genes in the brain. Specifically, the isoform-promiscuous *cis* heterodimerization and isoform-specific *trans* homo-dimerization of combinatorially expressed *cPcdhs* mediate cell-specific recognition between neuronal membranes, resulting in ingenious dendrite self-avoidance and axonal coexistence<sup>27,35–42</sup>. Intriguingly, chromatin



**Fig. 3 | The *cPcdh* genes are expressed from H3K9me3/H3K36me3-enriched facultative heterochromatin domains. a** H3K9me3 and H3K36me3 ChIP-seq profiles of the three *Pcdh* clusters in mouse neocortex, showing colocalization of both the active mark of H3K36me3 and inactive mark of H3K9me3. **b** RNA-seq as well as H3K4me3 and H3K9ac ChIP-seq profiles of the *cPcdh* clusters in mouse

neocortex, showing a strong correlation between each *cPcdh* gene and the enrichment of active marks. **c** The H3K9me2 ChIP-seq and MeDIP-seq profiles of the *cPcdh* clusters, showing enrichments of the repressive mark of H3K9me2 and DNA methylation at most *cPcdh* genes in the mouse neocortex.

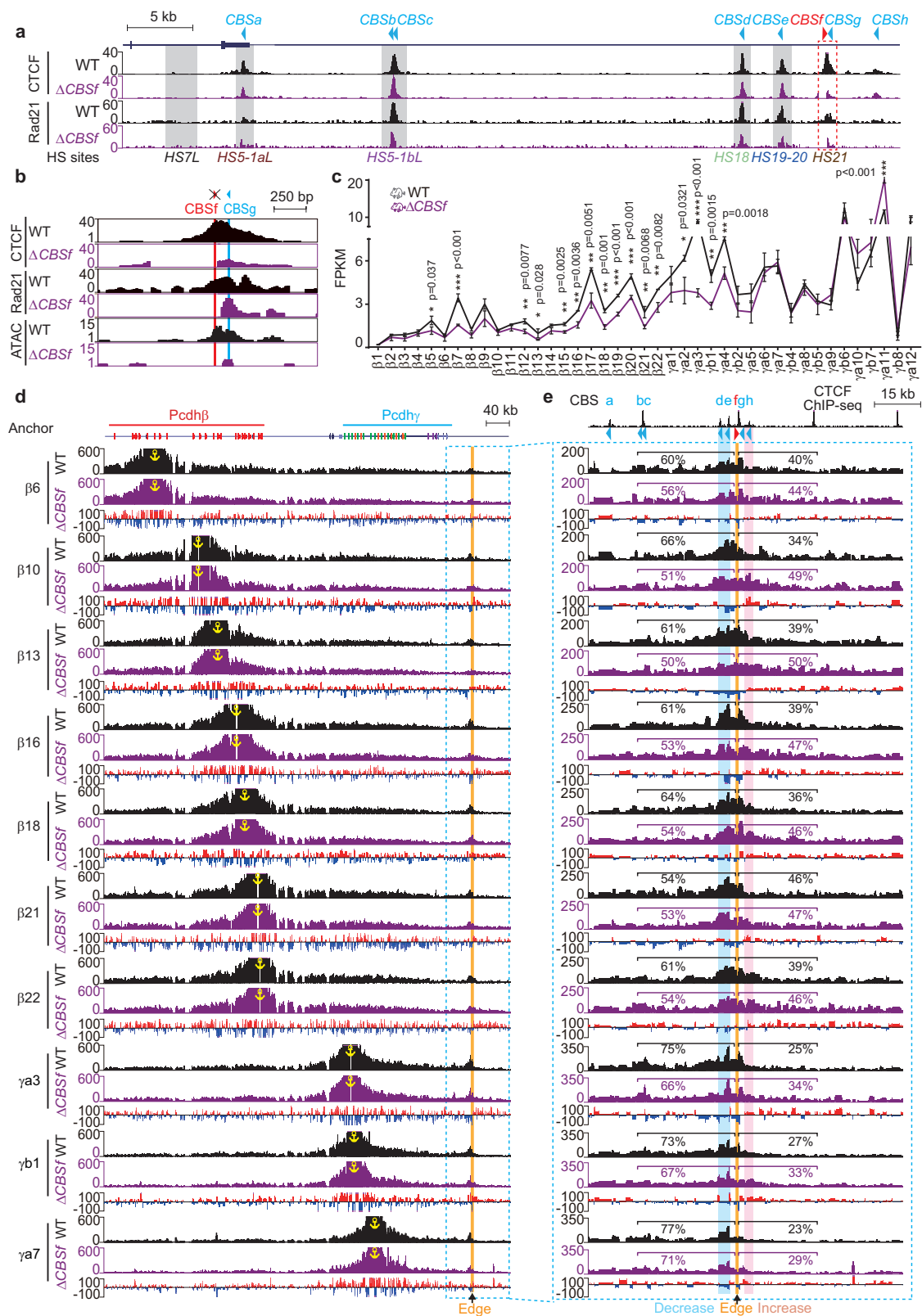
establishes a stable *cPCDH* expression pattern for single neurons in the brain during human fetal development<sup>43</sup>. In addition, the proper spatiotemporal expression of the *HoxD* genes is central for limb development<sup>29,44</sup>. Here, we first comprehensively mapped *cPcdh* expression patterns in the neocortex during mouse brain development. We then dissected, through sequential and combinatorial deletion of candidate enhancer elements, how *cPcdhs* are regulated by a downstream clustered CTCF boundary super-enhancer. In addition, we further deleted outward-oriented CBS elements within *Pcdh* and *HOXD* clustered CTCF TAD boundaries, which are associated with dozens of human pathogenic copy number variations (CNVs) (Supplementary Data 1). In conjunction with chromosome conformation and gene expression analyses, we found that outward-oriented CBS elements within clustered CTCF TAD boundaries are crucial for intra-TAD promoter-enhancer interactions and proper gene regulation (Fig. 5h). Unlike inward-oriented CBS elements, which directly make intra-TAD spatial contacts, such as between enhancers and promoters, the outward-oriented CBS elements may block extruding cohesin sliding from the outside of the TAD to prevent aberrant inter-TAD chromatin interactions (Fig. 5h).

The characteristics of the clustered CTCF TAD boundaries and their exact nature of insulation activities are under intense investigation<sup>13,18,20,22,29,44–48</sup>. The CBS elements within a clustered CTCF TAD boundary between *Pax3* and its distal enhancer cooperate redundantly for robust insulation<sup>18</sup>. In addition, various combinations of clustered CBS elements in distinct orientations could function as topological insulators when inserted between distal enhancers and their target promoters<sup>20,22,49</sup>. Moreover, the CBS elements within a clustered CTCF TAD boundary located between the C-DOM and T-DOM (telomeric domain) of *HoxD* have various functions for controlling its spatiotemporal expression patterns during limb development<sup>29,44</sup>. Here, we show that the outward-oriented CBS elements of the left boundary of C-DOM are crucial for maintaining intra-TAD chromatin interactions and *HOXD13* gene expression (Fig. 5). Finally, consistent with the crucial role of outward-oriented CBS

elements, an inversion leading to increased numbers of outward-oriented CBS elements within a clustered CTCF TAD boundary results in enhanced insulation activity and altered gene expression<sup>50</sup>.

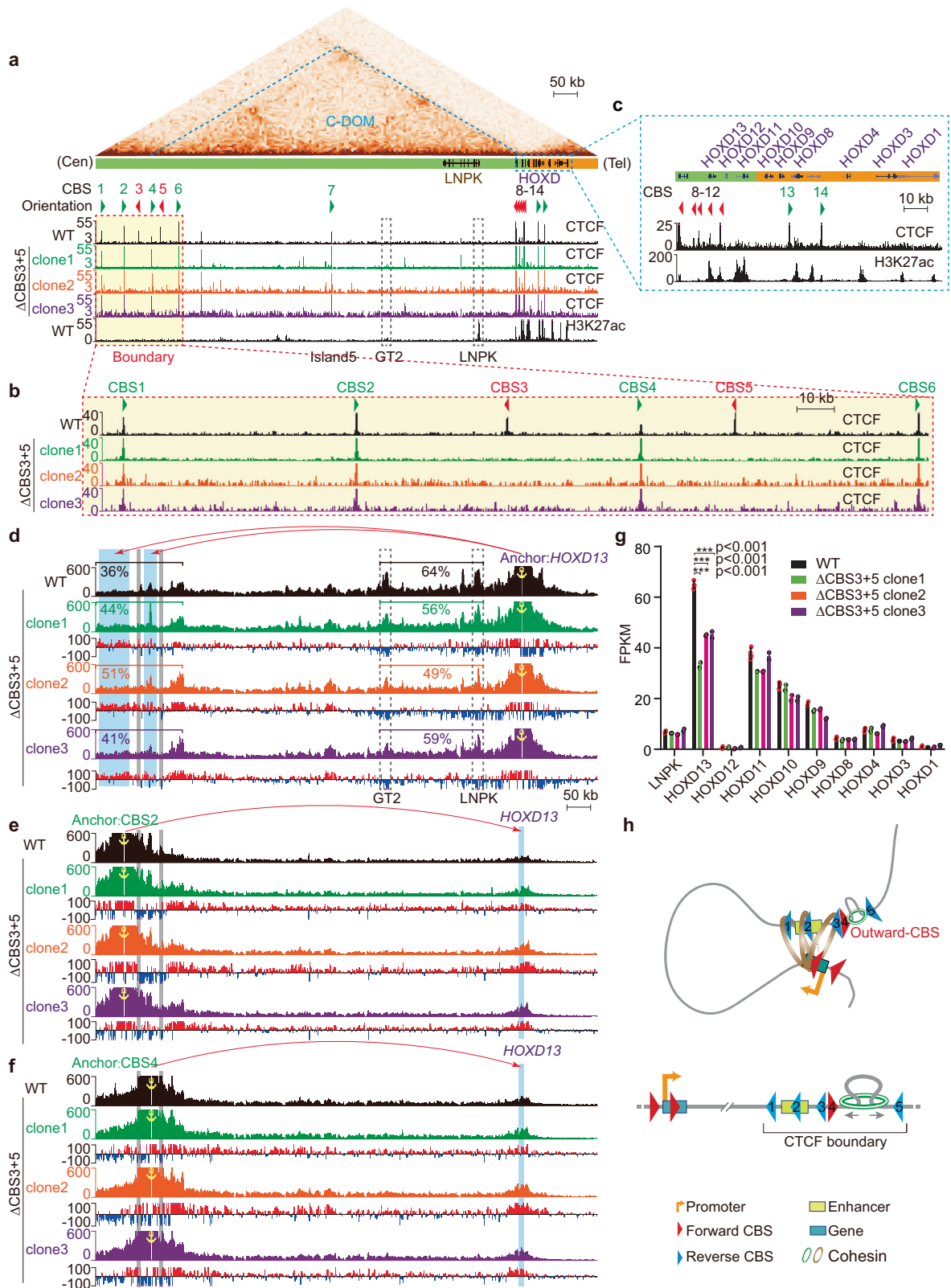
CTCF plays a key role in the formation of insulation at TAD boundaries, where in many cases multiple CBS elements are arranged in a clustered array with complex and mixed orientations, such as the telomeric downstream boundary of the *Pcdhβy* TAD and the centromeric upstream boundary of the C-DOM of the *HOXD* cluster. According to the convergent rule, the inward-oriented CBS elements within the clustered CTCF TAD boundaries are essential for intra-TAD chromatin interactions<sup>8,9</sup> (Fig. 5h). Thus, chromatin contacts such as long-distance enhancer-promoter interactions primarily occur within the TADs<sup>51,52</sup>. However, recent studies provide increasing evidence for inter-TAD communications between neighboring TADs<sup>50</sup>. Although TADs are likely a statistical model from population cells, there is clear evidence of clustering of CBS elements at TAD boundaries<sup>13,15,48</sup>. On one hand, the outward-oriented CBS elements could be crucial for intra-TAD chromatin interactions. On the other hand, the outward-oriented CBS elements function as crucial insulators blocking aberrant chromatin interactions from neighboring TADs as we observed a sharp edge for increased chromatin interactions beyond the outward CBS element (Fig. 4d, e). In the case of clustered CTCF TAD boundary of *Pcdhβy*, the extruding cohesin complexes from the downstream TAD cannot pass through the forward *CBSf* element. However, upon deletion of the outward *CBSf* element, these upward-extruding cohesin complexes pass through the entire TAD boundary and eventually anchored at variable promoters, leading to an increase of chromatin interactions between *Pcdhβy* genes and *CBSf* region.

The structure of TADs and loops is closely related to gene expression. Their formation is thought to be resulted from CTCF-anchored active cohesin “loop extrusion”. Specifically, cohesin-extruded loops anchored between convergent forward-reverse CBS elements can enhance or stabilize enhancer-promoter spatial contacts, which may or may not activate gene expression<sup>53–56</sup>. In addition, both the spatial and linear distances between distal enhancers and target promoters could be crucial for gene regulation<sup>57–59</sup>. Our study



**Fig. 4 | Outward-oriented *CBSf* regulates *Pcdhβy* by blocking their aberrant looping from the outside of the TAD boundary.** **a** CTCF and Rad21 ChIP-seq profiles of the *Pcdhβy* downstream TAD boundary in *CBSf*-deleted ( $\Delta CBSf$ ) homozygous mice compared to their wild-type (WT) littermates. **b** Close-up of CTCF and Rad21 ChIP-seq as well as ATAC-seq profiles in  $\Delta CBSf$  mice compared to their WT littermates. **c** RNA-seq showing decreased expression levels of *Pcdhβy* upon *CBSf* deletion. Data as mean  $\pm$  SD, \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ; two-tailed Student's *t*

test. For *CBSf* deletion mouse line,  $n = 4$  biologically independent samples; for their WT littermate controls,  $n = 2$ . Source data are provided as a Source Data file. **d**, **e** 4C profiles using a repertoire of *Pcdhβy* promoters as anchors, showing decreased chromatin interactions with *HSI8-20* enhancers (highlighted in blue, **(e)**) and increased chromatin interactions beyond *CBSf* (highlighted in pink, **(e)**). Note a sharp edge for the transition from decrease to increase at the location of *CBSf*. Differences ( $\Delta CBSf$  versus WT) are shown under the 4C profiles.



on *Pcdhyc3* regulation is consistent with this idea. *Pcdhyc3*, the most abundantly expressed *cPcdh* isoform in the mouse neocortex (Supplementary Fig. 2e), plays a crucial role in dendrite arborization<sup>34</sup>. It is puzzling why *H5S-1bl* deletion leads to a significant increase of expression levels of only  $\gamma 3$ , but not other members, of *cPcdh* genes, suggesting that it is a  $\gamma 3$ -specific insulator (Figs. 1i, 2e and Supplementary Fig. 4e). It has been shown that gene activation is not only

dependent on promoter-associated CBS elements, but also on the density of enhancers of the genomic context<sup>32,57</sup>. Consistently, only  $\gamma 3$ , but not  $\gamma 4$  and  $\gamma 5$ , is associated with two forward-oriented CBS elements (Fig. 2a) despite the fact that, compared with other members of *PcdhBy* clusters, all of the three *Pcdhyc* genes are located close to the downstream distal enhancers of *H5S18-20*, which contain two reverse-oriented CBS elements (Fig. 1b). A “double clamp”



**Fig. 5 | Double knockout of outward-oriented CBS3 and CBS5 elements down-regulates HOXD13 gene expression.** **a** CTCF and H3K27ac ChIP-seq profiles of the *HOXD* centromeric regulatory TAD (C-DOM) in wild-type (WT) and in *CBS3* and *CBS5* double knockout ( $\Delta CBS3+5$ ) homozygous single-cell clones. The left boundary (red-dotted triangle) comprises a cluster of CTCF sites, *CBS1-6*, of which *CBS3* and *CBS5* are in the outward orientation. Arrowheads indicate CBS elements with orientations. **b** Close-up of CTCF ChIP-seq profiles at the C-DOM left boundary marked by red dotted rectangle in (a). Note the abolishment of CTCF peaks at the locations of *CBS3* and *CBS5* in the three single-cell deletion clones. **c** Close-up of CTCF and H3K27ac ChIP-seq profiles at the *HOXD* cluster in (a). Note that the five reverse-oriented CBS elements are associated with *HOXD13* but not with other *HOXD* genes. **d** 4C profiles using *HOXD13* as an anchor in  $\Delta CBS3+5$  single-cell clones compared to WT clones, showing increased chromatin interactions with the

left boundary and decreased chromatin interactions with the H3K27ac-enriched *GT2* or *LNPK* element upon double knockout. **e, f** 4C profiles using *CBS2* (e) or *CBS4* (f) as an anchor confirming increased chromatin interactions with *HOXD13*. **g** Expression levels of *HOXD* genes in three  $\Delta CBS3+5$  single-cell clones compared to WT clones. Data as mean  $\pm$  SD, \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ; two-tailed Student's *t* test. For each deletion clone,  $n = 2$  biologically independent samples; for their WT controls,  $n = 4$ . Source data are provided as a Source Data file. **h** A boundary model illustrating the role of outward-oriented CTCF sites (CBS elements) in intra-TAD chromatin contacts and gene regulation. TAD boundary normally contains clustered CTCF sites with complex orientations of both divergence and convergence. The outward-oriented CTCF sites function as an important barrier for outside cohesin to ensure proper and balanced intra-TAD chromatin interactions and gene expression.

interaction mechanism of the two forward-reverse CBS pairs, similar to the activation of members of the *Pcdha* gene cluster<sup>24</sup>, may be utilized to activate  $\gamma c3$ . This explains why only  $\gamma c3$ , but not  $\gamma c4$  and  $\gamma c5$ , expression levels are enhanced upon deletion of the *HSS-1bl* (*CBSbc*) insulator.

*HOXD13* showed strong interactions with two H3K27ac-marked regions of *LNPK* and *GT2*. We observed a significant decrease of chromatin interactions with *LNPK* and *GT2* upon removal of outward-oriented CBS elements (Fig. 5d), explaining the decreased levels of *HOXD13* gene expression. This suggests that the outward-oriented CBS boundary elements of the C-DOM may ensure proper interactions with *LNPK* and *GT2* for precise control of *HOXD13* expression patterns. Our data highlight a significant role of the outward-oriented CBS elements in providing optimal insulation for restraining aberrant inter-TAD interactions and maintain proper intra-TAD promoter-enhancer interactions in the *Pcdh* and *HoxD* clusters. We analyzed genome-wide CBS patterns at TAD boundaries and found that 85.7% have outward-oriented CBS elements (Supplementary Data 2, 3). However, whether it could be generalized to the entire genome awaits further studies.

We observed that H3K9me3 marks correlate with the monoallelic *cPcdh* gene expression and that H3K9me2 marks correlate with both monoallelic and biallelic *cPcdh* gene expression<sup>26</sup>. The correlation between H3K9me3 enrichments and DNA methylation levels in both monoallelic and biallelic *cPcdh* genes may be coordinated by UHRF1<sup>60</sup>. Interestingly, almost every member of the *cPcdh* genes is marked by both repressive (H3K9me3 and H3K9me2) and active (H3K36me3) marks (Fig. 3), suggestive of a facultative heterochromatic state related to monoallelic *cPcdh* gene expression. Whether the overlap of H3K9me3 and H3K36me3 is resulted from allelic differences of the epigenome and/or admixtures of distinct subpopulation cells awaits further exploration.

## Methods

### Cell culture

Human HEK293T cells were cultured in DMEM (Hyclone) supplemented with 10% (v/v) FBS (Gibco) and 10 U/ml penicillin-streptomycin (Gibco, 15140122). Human HEC-1-B cells were cultured in MEM (Hyclone) supplemented with 200 mM L-glutamine, Earle's balanced salt solution (Hyclone, SH30024.01), 1 mM sodium pyruvate (Sigma, SLBP4879V), 10% fetal bovine serum, and 10 U/ml penicillin-streptomycin. All cells were incubated at 37 °C and 5% CO<sub>2</sub> in a humidified incubator, and passaged every 3 days.

### CRISPR screening of CBS-deletion single-cell clones

*CBS3* and *CBS5* within the clustered CTCF TAD boundary of the *HOXD* locus were sequentially knocked out by two steps of CRISPR/Cas9-mediated DNA fragment editing (Supplementary Data 4)<sup>9,61</sup>. In the first step, *CBS3* or *CBS5* was deleted in HEK293T cells to generate homozygous *CBS3* or *CBS5* knockout cell clones. In the second step, a *CBS3* knockout cell clone was used to delete *CBS5* and a *CBS3* knockout cell

clone was used to delete *CBS3*, to generate cell clones homozygous for *CBS3* and *CBS5* double knockout ( $\Delta CBS3+5$ ).

For each step of CRISPR editing, cells were grown to ~80% confluence in 6-well tissue culture plates and transfected with 1.5  $\mu$ g of pcDNA3.1-Cas9 and 1.5  $\mu$ g dual sgRNA expression plasmids per well by Lipofectamine 3000 (Invitrogen, L3000015). Two days after transfection, 2  $\mu$ g/ml puromycin was added to the cell growth medium and maintained for 5 days to select the transfected cells. The transfected cells were suspended into single-cell solutions, diluted and plated into 96 well plates. Two weeks later, single-cell clones were picked and genotyped by PCR with specific primers (Supplementary Data 5). Positive clones were Sanger sequenced.

The sgRNA expression plasmids were constructed as previously described<sup>9,61</sup> with paired oligonucleotides (Supplementary Data 5). For each sgRNA, a forward oligonucleotide with the (ACCG-5') overhang and a reverse complementary oligonucleotide with the (AAAC-5') overhang were annealed and cloned into the Bsa I site of pGL3-U6-sgRNA-PGK-Puro vector for sgRNA transcription<sup>62</sup>. All sgRNA-expressed plasmids were confirmed by Sanger sequencing.

### In vitro synthesis of Cas9 mRNA and sgRNAs for micro-injection

A T7 promoter-driven Cas9 expression plasmid<sup>63</sup> was linearized by *Xba* I, purified and in-vitro transcribed using the T7 Ultra Kit (ThermoFisher, AM1345). The transcribed Cas9 mRNA was treated with Turbo DNase to remove template DNA and added with poly-A tail. The sgRNA template was generated by PCR amplification with specific primers (Supplementary Data 5) to contain a T7 promoter, a 20 nt target sequence, and a scaffold region<sup>49</sup>. The amplified product was purified and transcribed using the MEGAshortscript T7 Kit (ThermoFisher, AM1354). The transcribed sgRNA was treated with Turbo DNase. Both sgRNA and Cas9 mRNA were purified by the MEGAclear Kit (ThermoFisher, AM1908), eluted into elution buffer and stored at -80 °C before usage.

### Generation of CBS-deletion mice

Mice were maintained in an SPF mouse facility at 23 °C, with a humidity between 40 and 60%, with a 12 h (7:00–19:00) light/12 h (19:00–7:00) dark cycle. All the mouse experiments were approved by the Institutional Animal Care and Use Committee (IACUC) of Shanghai Jiao Tong University (Protocol#: 1602029).

C57BL/6J mice were used as embryo donors and ICR mice were used as foster mothers. 6-week-old C57BL/6J female mice were super-ovulated and mated with the sterilized C57BL/6J male mice to produce enough embryos. Zygotes were collected from the oviducts of the super-ovulated C57BL/6J female mice, washed with 200  $\mu$ g/ml hyaluronidase (Sigma, V900833) and incubated in the M2 medium (Sigma, M7167) in a 5% CO<sub>2</sub> incubator at 37 °C for 1 h. Viable embryos were then microinjected with a solution containing 100 ng/ $\mu$ l of Cas9 mRNA and 50 ng/ $\mu$ l each of two sgRNAs targeting each fragment and recovered in a 5% CO<sub>2</sub> incubator at 37 °C for 1 h. The injected live embryos were transplanted into the oviducts of the pseudo-pregnant ICR female

mice under a stereoscopic microscope, with each ICR mouse receiving 25–30 embryos. Post-surgery mice were maintained at a 37 °C heating plate for 1 h for recovery before being transferred back for regular housing until the birth of the F0 mice. Mice exhibiting signs of dystocia at day 20 received C-section rescue.

The chimeric F0 mice were screened 10 days later for desired deletions by PCR genotyping with specific primers (Supplementary Data 5) and the chimeric mutant mice were confirmed by Sanger sequencing. The F0 mice of 2-month old with desired deletions were mated with wild-type C57BL/6J mice to generate heterozygous F1 mice. The F1 mice were genotyped by PCR. The targeted F1 male and female mice were crossed to generate the F2 homozygous mice, which were genotyped (deletions of *HS7L*, *CBSf*, *CBSfg*, or *CBSg* were shown in Supplementary Figs. 4, 12, the other mouse lines were genotyped previously<sup>49</sup>) and used for downstream experiments. The wild-type F2 littermates were used as controls.

### ATAC-seq

ATAC-seq was performed following the omni-ATAC procedure<sup>64</sup> with some modifications. Briefly, the neocortex was microdissected and digested with 0.00625% collagenase followed by filtration through a 100 µm cell strainer to generate single-cell suspensions. The suspended cells were washed twice with PBS solution and then counted. 10<sup>5</sup> cells were aliquoted and resuspended with 50 µl of the resuspension buffer (RSB) supplemented with 0.1% NP40, 0.1% Tween-20, and 0.01% digitonin followed by incubation on ice for 3 min for lysis. The nuclei were then washed with 1 ml of RSB supplemented with only 0.1% Tween-20 and resuspended in 50 µl of transposition mix, which includes 25 µl of 2× TD buffer, 3 µl transposase, 16 µl PBS, 0.5 µl 1% digitonin, 0.5 µl 10% Tween-20 and 5 µl H<sub>2</sub>O, followed by incubation at 37 °C for 30 min in a shaker at 1000 rpm. Transposition products were immediately cleaned up with 1× Ampure XP Beads (Beckman, A63881), and the purified DNA was used for library construction by the TruePrep DNA Library Prep Kit V2 for Illumina (Vazyme Biotech, TD502). The ATAC-seq libraries were pooled and sequenced at the 2 × 150 bp mode using an Illumina NovaSeq 6000 platform.

### RNA-seq

Total RNA was extracted from cultured cells or the microdissected mouse neocortex using the Trizol reagent (ThermoFisher, 1596026). For each sample, 1 µg of total RNA was used to enrich mRNA using Poly(A) Magnetic Isolation Module (NEB, E7600). The mRNA was purified twice, eluted into the fragmentation buffer with random primers for fragmentation at 94 °C for 15 min, and reverse-transcribed into the first strand cDNA. The second strand was then synthesized by the second-strand DNA polymerase. The generated double-stranded cDNA was purified with Ampure XP beads (Beckman, A63881), end repaired, 3'-adenylated, and 5'-phosphorylated. The Illumina sequencing adapters were then ligated to both ends of cDNA and excess adapters were removed using Ampure XP beads (Beckman). The purified cDNA was amplified by PCR with barcoding to generate RNA-seq libraries. Pooled multiplexed libraries were sequenced at the 2 × 150 bp mode on an Illumina NovaSeq 6000 platform. All experiments were performed with at least two biological replicates.

Fetal sex is determined based on the expression of *Xist* and *Ddx3y*. For expression mapping at different embryonic stages during cortical development, mouse neocortices were isolated from both males (*Xist*<sup>+</sup>, *Ddx3y*) and females (*Xist*<sup>-</sup>, *Ddx3y*<sup>-</sup>). As we did not observe sex-specific differences in *cPcdh* expression at any developmental stages (Supplementary Fig. 2f, g), we did not discriminate between male and female mouse neocortices for all the following experiments.

### ChIP-seq

Mouse neocortical tissues was microdissected and dissociated into single cells using 0.00625% collagenase. 2 × 10<sup>6</sup> of mouse neocortical or

cultured human cells were aliquoted and cross-linked using 1% formaldehyde in PBS containing 10% FBS for 10 min at room temperature followed by two washes with PBS containing proteinase inhibitors (Roche, 04693132001). Cells were then resuspended and lysed twice with 1 ml of prechilled lysis buffer containing proteinase inhibitors (Roche) for 10 min with slow rotations. For cortical cells, additional SDS was added to the lysis solution to adjust the final SDS concentration to 0.4% for better sonication. Cells were then sonicated at 25% power for a train of 15 s ON and 30 s OFF for 20 cycles with the Bioruptor system to break the genomic DNA into 0.1–10 kb fragments. After sonication, the SDS-free lysis buffer was added to cortical samples to dilute the SDS concentration to 0.1% to allow the binding of antibodies to proteins. After removal of the insoluble debris by centrifuging at 14,000 *g* for 10 min, the lysate was pre-cleaned with protein-A agarose beads (Millipore, 16–157) for 1 h with slow rotations to remove non-specific binding and then incubated with specific antibody (Supplementary Data 6) overnight with slow rotations at 4 °C. The antibody-protein-DNA complex was isolated by adding protein A-agarose beads and incubated for 4 h with slow rotations followed by sequential washes with the low-salt buffer, high-salt buffer, salt-free buffer, LiCl buffer, and TE buffer. The beads were eluted in the elution buffer (0.1 M NaHCO<sub>3</sub>, 1% SDS) and de-cross-linked by proteinase K at 65 °C overnight. The DNA was finally purified and used for the construction of DNA library using the VAHTS Universal DNA Library Prep Kit for Illumina V3 (Vazyme Biotech, ND607). ChIP-seq libraries were sequenced at the 2 × 150 bp mode on an Illumina NovaSeq 6000 platform.

### QHR-4C

For QHR-4C<sup>49</sup> experiments, mouse neocortices were microdissected and dissociated with 0.00625% collagenase to obtain single cells. Mouse neocortical or cultured human cells were cross-linked in 2% formaldehyde for 10 min at room temperature and quenched by adding excess pre-chilled glycine. The fixed cells were incubated on ice for 5 min, and permeabilized twice with the permeabilization buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1% Triton X-100, and protease inhibitors), with slow rotations for 10 min at 4 °C. The permeabilized cells were then digested with *Dpn II* for 16 h at 37 °C while shaking at 900 rpm followed by incubation at 65 °C for 20 min to inactive *Dpn II*. Proximity ligation was then performed for 16 h at 16 °C by adding T4 DNA ligase in 1× T4 ligation buffer. The ligated product was de-cross-linked by proteinase K at 65 °C overnight, treated with RNase A at 37 °C for 45 min, and purified with phenol-chloroform followed by ethanol precipitation. DNA was sonicated to 0.2–1 kb fragments using the Bioruptor system at 33% power for a train of 15 s ON and 30 s OFF for four cycles.

The anchor fragments were linear-amplified using a specific 5' biotin-tagged primer (Supplementary Data 5) with 120 cycles (95 °C 20 s for denaturing, 60 °C 20 s for annealing, and 72 °C 90 s for elongation). The amplified products were finally denatured at 95 °C for 5 min and then immediately chilled on ice to prevent reannealing. The biotin-tagged single-stranded DNA (ssDNA) was pulled down with Streptavidin Magnetic beads (ThermoFisher, 65001) and ligated with adapters which were generated by annealing two partially complementary single-stranded oligonucleotides (Adapter-U and Adapter-L, Supplementary Data 5) by T4 DNA ligase in 1× T4 ligation buffer containing PEG8000 at 16 °C for 12 h. The on-bead ligation products were washed twice with the B/W buffer (5 mM Tris-HCl pH 7.5, 1 M NaCl, and 0.5 mM EDTA) to remove excess adapters and used as the template for PCR amplification of the 4C libraries with barcoded anchor-specific P5-forward primers and indexed P7-reverse primers (Supplementary Data 5). Multiplexed libraries from the same anchor with different combinations of barcodes and indexes were pooled for purification with a PCR purification kit (Qiagen, 11732668001). 4C libraries were sequenced at the 2 × 150 bp mode on an Illumina NovaSeq 6000 platform.

### In situ Hi-C

For in situ Hi-C<sup>8</sup> experiments,  $5 \times 10^6$  cells were cross-linked with formaldehyde and incubated in a lysis buffer (10 mM Tris-HCl pH8.0, 10 mM NaCl, 0.2% NP-40) to obtain nuclei. Nuclei were permeabilized with SDS, quenched by adding Triton X-100, and digested with *Mbo* II (NEB, R0147M) overnight at 37 °C, followed by heat inactivation at 62 °C. The ends of restriction fragments were filled in with biotin-14-dATP (Thermo, 19524016), dCTP, dTTP, and dGTP using the Klenow fragment of DNA polymerase I (NEB, M0210L), and ligated using T4 DNA ligase (NEB, M0202S). After ligation, the nuclei were de-cross-linked overnight at 68 °C. The ligated DNA was precipitated with ethanol and fragmented by sonication. 300–500 bp DNA fragments were selected using AMPure XP beads (Beckman, A63881). The biotinylated DNA was pulled down with M-280 Streptavidin beads (Thermo, 11206D), washed, and used for library construction on bead. The DNA was end-repaired, tailed with dA using Klenow exo<sup>-</sup> (NEB, M0212S), and ligated with Illumina U-type adapters, followed by adapter linearization with a USER enzyme (NEB, M5505S). The obtained DNA was used as templates for PCR amplification with 10–12 cycles. PCR products were purified with AMPure XP beads (Beckman, A63881) and sequenced at the  $2 \times 150$  bp mode on an Illumina Nova-Seq 6000 platform.

### High-throughput sequencing data analysis

RNA-seq reads were aligned against the mouse (GRCm38) or human (GRCh38) genome using Hisat2<sup>65</sup> (version 2.0.4) to generate the sequence alignment map (SAM) files. SAM files were then converted into the binary versions (BAM files) using Samtools (version 1.15.1). BAM files were used as input for Cufflinks<sup>66</sup> (version 2.1.1) to calculate FPKM values.

ChIP-seq and ATAC-seq reads were aligned to the mouse (mm9) or human (hg19) genome using Bowtie2 (version 2.3.5) to generate SAM files, which were converted into BAM files using Samtools (version 1.15.1). BAM files were indexed and converted into BedGraph files.

For 4 C data, P7 reads were initially demultiplexed based on the unique barcode-index combinations. The anchor primer sequences were trimmed and PCR duplicate reads were removed using FastUniq (version 1.1) of the UNIX system. The obtained unique reads were aligned against the mouse (mm9) or human (hg19) genome using Bowtie2 (version 2.3.5) to generate SAM files, which were converted into BAM files. The BAM files were used as input for r3Cseq (version 1.20) in the R package (3.3.3) to calculate the reads per million (RPM) values.

Hi-C reads were pre-processed with HiC-Pro (v3.0.0)<sup>67</sup>, including reads alignment to the human hg19 reference genome, interaction matrix generation, and ICE normalization. 10 kb ICE normalized interaction matrix was generated. TADs and TAD boundaries were called at a 10-kb resolution using the directionality index (DI) method<sup>51</sup>. Insulation scores were calculated at a 10-kb resolution as previously described<sup>68</sup>.

For genome-wide CBS patterns at TAD boundaries, we analyzed in situ HiC data. CBS position and orientation at TAD boundaries were determined by FIMO. We refer to a CTCF motif as “forward” (>) when it is present on the forward strand of the chromosome and ‘reverse’ (<) when it is on the reverse strand. We defined TAD boundary regions as  $\pm 50$  kb from the center of each boundary. TAD boundary regions were intersected with CTCF motifs using “bedtools intersect” to identify CBS numbers and orientations for each TAD boundary.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The raw high-throughput sequencing data and processed data generated in this study have been deposited in the GEO database under accession code [GSE210817](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE210817). Source data are provided with this paper.

### References

- Mirny, L. & Dekker, J. Mechanisms of chromosome folding and nuclear organization: their interplay and open questions. *Cold Spring Harb. Perspect. Biol.* **14**, 1–20 (2022).
- Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–1211 (2009).
- Hansen, A. S., Cattoglio, C., Darzacq, X. & Tjian, R. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* **9**, 20–32 (2018).
- Wu, Q., Liu, P. & Wang, L. Many facades of CTCF unified by its coding for three-dimensional genome architecture. *J. Genet. Genom.* **47**, 407–424 (2020).
- Nichols, M. H. & Corces, V. G. A CTCF code for 3D genome architecture. *Cell* **162**, 703–705 (2015).
- Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. USA* **112**, E6456–E6465 (2015).
- Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
- Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
- Guo, Y. et al. CRISPR Inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162**, 900–910 (2015).
- Lupianez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
- Flavahan, W. A. et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2016).
- Kloetgen, A. et al. Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. *Nat. Genet.* **52**, 388–400 (2020).
- Kentepozidou, E. et al. Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome Biol.* **21**, 5 (2020).
- Nanni, L., Ceri, S. & Logie, C. Spatial patterns of CTCF sites define the anatomy of TADs and their boundaries. *Genome Biol.* **21**, 197 (2020).
- McArthur, E. & Capra, J. A. Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. *Am. J. Hum. Genet.* **108**, 269–283 (2021).
- Gong, Y. et al. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nat. Commun.* **9**, 542 (2018).
- Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
- Anania, C. et al. In vivo dissection of a clustered-CTCF domain boundary reveals developmental principles of regulatory insulation. *Nat. Genet.* **54**, 1026–1036 (2022).
- Despang, A. et al. Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat. Genet.* **51**, 1263–1271 (2019).
- Huang, H. et al. CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nat. Genet.* **53**, 1064–1074 (2021).
- Taylor, T. et al. Transcriptional regulation and chromatin architecture maintenance are decoupled functions at the Sox2 locus. *Genes Dev.* **36**, 699–717 (2022).
- Chakraborty, S. et al. Enhancer-promoter interactions can bypass CTCF-mediated boundaries and contribute to phenotypic robustness. *Nat. Genet.* **55**, 280–290 (2023).
- Wu, Q. et al. Comparative DNA sequence analysis of mouse and human protocadherin gene clusters. *Genome Res.* **11**, 389–404 (2001).

24. Guo, Y. et al. CTCF/cohesin-mediated DNA looping is required for protocadherin alpha promoter choice. *Proc. Natl. Acad. Sci. USA* **109**, 21081–21086 (2012).
25. Allahyar, A. et al. Enhancer hubs and loop collisions identified from single-allele topologies. *Nat. Genet.* **50**, 1151–1160 (2018).
26. Esumi, S. et al. Monoallelic yet combinatorial expression of variable exons of the protocadherin-alpha gene cluster in single neurons. *Nat. Genet.* **37**, 171–176 (2005).
27. Lv, X. et al. Patterned cPCDH expression regulates the fine organization of the neocortex. *Nature* **612**, 503–511 (2022).
28. Andrey, G. et al. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* **340**, 1234167 (2013).
29. Amandio, A. R. et al. Sequential in cis mutagenesis in vivo reveals various functions for CTCF sites at the mouse HoxD cluster. *Genes Dev.* **35**, 1490–1509 (2021).
30. Jiang, Y. et al. The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain. *Nat. Genet.* **49**, 1239–1250 (2017).
31. Ribich, S., Tasic, B. & Maniatis, T. Identification of long-range regulatory elements in the protocadherin-alpha gene cluster. *Proc. Natl. Acad. Sci. USA* **103**, 19719–19724 (2006).
32. Kehayova, P., Monahan, K., Chen, W. & Maniatis, T. Regulatory elements required for the activation and repression of the protocadherin-alpha gene cluster. *Proc. Natl. Acad. Sci. USA* **108**, 17195–17200 (2011).
33. Yokota, S. et al. Identification of the cluster control region for the protocadherin-beta genes located beyond the protocadherin-gamma cluster. *J. Biol. Chem.* **286**, 31885–31895 (2011).
34. Steffen, D. M. et al. A unique role for protocadherin gammaC3 in promoting dendrite arborization through an Axin1-dependent mechanism. *J. Neurosci.* **43**, 918–935 (2023).
35. Schreiner, D. & Weiner, J. A. Combinatorial homophilic interaction between gamma-protocadherin multimers greatly expands the molecular diversity of cell adhesion. *Proc. Natl. Acad. Sci. USA* **107**, 14893–14898 (2010).
36. Lefebvre, J. L., Kostadinov, D., Chen, W. V., Maniatis, T. & Sanes, J. R. Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. *Nature* **488**, 517–521 (2012).
37. Thu, C. A. et al. Single-cell identity generated by combinatorial homophilic interactions between alpha, beta, and gamma protocadherins. *Cell* **158**, 1045–1059 (2014).
38. Toyoda, S. et al. Developmental epigenetic modification regulates stochastic expression of clustered protocadherin genes, generating single neuron diversity. *Neuron* **82**, 94–108 (2014).
39. Mountoufaris, G. et al. Multicluster Pcdh diversity is required for mouse olfactory neural circuit assembly. *Science* **356**, 411–414 (2017).
40. Ing-Esteves, S. et al. Combinatorial effects of alpha- and gamma-protocadherins on neuronal survival and dendritic self-avoidance. *J. Neurosci.* **38**, 2713–2729 (2018).
41. Brasch, J. et al. Visualization of clustered protocadherin neuronal self-recognition complexes. *Nature* **569**, 280–283 (2019).
42. Canzio, D. et al. Antisense lncRNA transcription mediates DNA demethylation to drive stochastic protocadherin alpha promoter choice. *Cell* **177**, 639–653.e15 (2019).
43. Almenar-Queralt, A. et al. Chromatin establishes an immature version of neuronal protocadherin selection during the naive-to-primed conversion of pluripotent stem cells. *Nat. Genet.* **51**, 1691–1701 (2019).
44. Rodriguez-Carballo, E. et al. The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev.* **31**, 2264–2281 (2017).
45. Narendra, V., Bulajic, M., Dekker, J., Mazzoni, E. O. & Reinberg, D. CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes Dev.* **30**, 2657–2662 (2016).
46. Barutcu, A. R., Maass, P. G., Lewandowski, J. P., Weiner, C. L. & Rinn, J. L. A TAD boundary is preserved upon deletion of the CTCF-rich Firre locus. *Nat. Commun.* **9**, 1444 (2018).
47. Chang, L. H. et al. Multi-feature clustering of CTCF binding creates robustness for loop extrusion blocking and topologically associating domain boundaries. *Nat Commun.* **14**, 5615 (2023).
48. Okhovat, M. et al. TAD Evolutionary and functional characterization reveals diversity in mammalian TAD boundary properties and function Preprint at <https://doi.org/10.1101/2023.03.07.531534> (2023).
49. Jia, Z. et al. Tandem CTCF sites function as insulators to balance spatial chromatin contacts and topological enhancer-promoter selection. *Genome Biol.* **21**, 75 (2020).
50. Galupa, R. et al. Inversion of a topological domain leads to restricted changes in its gene expression and affects interdomain communication. *Development* **149**, dev200568 (2022).
51. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
52. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
53. Alexander, J. M. et al. Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *Elife* **8**, e41769 (2019).
54. Benabdallah, N. S. et al. Decreased enhancer-promoter proximity accompanying enhancer activation. *Mol. Cell* **76**, 473–484.e7 (2019).
55. Fulco, C. P. et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
56. Hsieh, T. S. et al. Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nat. Genet.* **54**, 1919–1932 (2022).
57. Kubo, N. et al. Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nat. Struct. Mol. Biol.* **28**, 152–161 (2021).
58. Rinzema, N. J. et al. Building regulatory landscapes reveals that an enhancer can recruit cohesin to create contact domains, engage CTCF sites and activate distant genes. *Nat. Struct. Mol. Biol.* **29**, 563–574 (2022).
59. Zuin, J. et al. Nonlinear control of transcription through enhancer-promoter interactions. *Nature* **604**, 571–577 (2022).
60. Butz, S. et al. DNA sequence and chromatin modifiers cooperate to confer epigenetic bistability at imprinting control regions. *Nat. Genet.* **54**, 1702–1710 (2022).
61. Li, J. et al. Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. *J. Mol. Cell Biol.* **7**, 284–298 (2015).
62. Shen, B. et al. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods* **11**, 399–402 (2014).
63. Chang, N. et al. Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell Res.* **23**, 465–472 (2013).
64. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
65. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
66. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).

67. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
68. Crane, E. et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).

## Acknowledgements

We are grateful for technical supports from Jingwei Li, Leyang Wang, and Mo Zhang. This work was supported by grants from the National Natural Science Foundation of China (32330016), the National Key R&D Program of China (2022YFC3400200), and the Science and Technology Commission of Shanghai Municipality (21DZ2210200).

## Author contributions

Q.W. conceived the research. X.G. and K.H. performed experiments. H.H. and X.G. analyzed data. Z.W. and W.X. supervised the mouse experiments. X.G., H.H., and Q.W. wrote the manuscript with inputs from all authors.

## Competing interests

The authors declare no competing interests

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-43849-0>.

**Correspondence** and requests for materials should be addressed to Qiang Wu.

**Peer review information** *Nature Communications* thanks Hongjie Yao and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023