

# Mechanism of REST/NRSF regulation of clustered protocadherin $\alpha$ genes

Yuanxiao Tang<sup>1</sup>, Zhilian Jia, Honglin Xu, Lin-tai Da<sup>1</sup> and Qiang Wu<sup>1\*</sup>

Center for Comparative Biomedicine, MOE Key Laboratory of Systems Biomedicine, State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic & Developmental Sciences, School of Life Sciences and Biotechnology, Institute of Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, China

Received February 26, 2021; Revised March 23, 2021; Editorial Decision March 25, 2021; Accepted March 26, 2021

## ABSTRACT

**Repressor element-1 silencing transcription factor (REST) or neuron-restrictive silencer factor (NRSF) is a zinc-finger (ZF) containing transcriptional repressor that recognizes thousands of neuron-restrictive silencer elements (NRSEs) in mammalian genomes. How REST/NRSF regulates gene expression remains incompletely understood. Here, we investigate the binding pattern and regulation mechanism of REST/NRSF in the clustered protocadherin (*PCDH*) genes. We find that REST/NRSF directionally forms base-specific interactions with NRSEs via tandem ZFs in an anti-parallel manner but with striking conformational changes. In addition, REST/NRSF recruitment to the *HS5-1* enhancer leads to the decrease of long-range enhancer-promoter interactions and downregulation of the clustered *PCDH $\alpha$*  genes. Thus, REST/NRSF represses *PCDH $\alpha$*  gene expression through directional binding to a repertoire of NRSEs within the distal enhancer and variable target genes.**

## INTRODUCTION

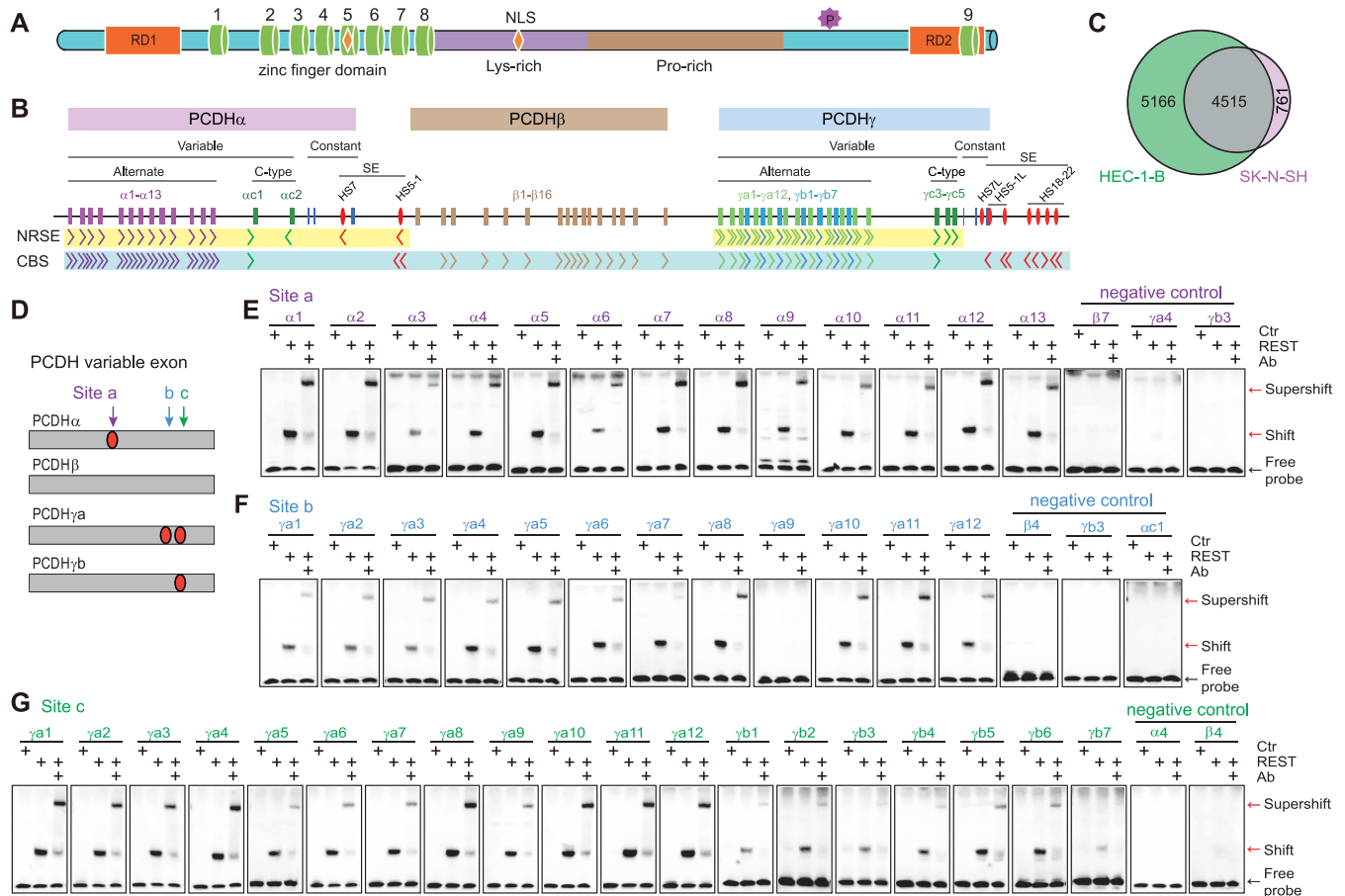
During early neurogenesis, the orderly acquisition and maintenance of neural identities are controlled epigenetically by derepression of neural genes through downregulating transcriptional repressors and corepressors (1). REST (repressor element-1 silencing transcription factor), also known as NRSF (neuron-restrictive silencer factor), is a crucial repressor for neural genes (2,3), reviewed in (4). Specifically, REST/NRSF represses the expression of numerous neural-specific genes in neural progenitor cells (NPCs) as well as non-neural tissues (5–8). In differentiated non-neural cells, REST/NRSF represses neural genes in collaboration with its corepressors (6,8–11). In embryonic stem cells, REST/NRSF is highly expressed (8,12). During transition to NPCs and finally to mature neurons,

REST/NRSF is degraded to minimal levels in NPCs and to an undetectable level in mature neurons (8,13). Recent studies revealed that REST/NRSF also has a protective role in genome stability (14).

REST/NRSF contains a central DNA-binding domain with eight tandem C2H2 ZFs and two repressor domains residing in the amino and carboxyl termini, respectively (Figure 1A) (2,3,15). REST/NRSF has been shown to bind to thousands of NRSEs which can be divided into three groups: canonical, noncanonical, and half-site only motifs (16,17). Intriguingly, canonical and noncanonical NRSEs contain very different gap sizes between the left- and right-half sites. ZF domains are small DNA-recognition units that are usually organized in tandem and there are >800 genes encoding ZF transcription factors in the human genome (18–21). The DNA-recognition mechanism by which these ZF transcription factors bind to distinct groups of vast number of genomic sites via conformation switch of ZF domains are largely unknown.

The clustered protocadherin (*PCDH*) genes encode a large number of cell-surface cadherin-like adhesion proteins which are thought to function as neuron identity codes in brain wiring and neuron discrimination (22–25). There are 53 highly-similar clustered *PCDH* genes organized into three closely-linked gene clusters (*PCDH $\alpha$* , *PCDH $\beta$*  and *PCDH $\gamma$* ) in the human 5q31 chromosomal region (Figure 1B) (22). The genomic organizations of *PCDH $\alpha$*  and *PCDH $\gamma$*  are similar in that each contains more than a dozen variable exons that can be spliced to a single set of three downstream constant exons. The variable exons can be grouped into alternate and C-type exons based on their locations and similarities (Figure 1B) (22). The *PCDH $\beta$*  cluster contains only variable exons but with no constant exon (22). Each variable exon has its own promoter (26,27). Two super-enhancers, one composed of *HS7* (DNase I hypersensitive site 7) and *HS5-1*, the other composed of *HS7L* (*HS7* like), *HS5-1L* (*HS5-1* like) and *HS18-22*, regulate the expression of the *PCDH $\alpha$*  and  *$\beta$*  clusters, respectively (Figure 1B) (25,28–33).

\*To whom correspondence should be addressed. Tel: +86 21 34204300; Fax: +86 21 34204300; Email: qiangwu@sjtu.edu.cn



**Figure 1.** REST/NRSF binding sites in the human *PCDH* clusters. (A) Schematic of the REST/NRSF protein. REST/NRSF protein contains a DNA binding domain of eight ZFs (cylinders), two repression domains (RD1 and RD2), lysine-rich and proline-rich domains, two nuclear localization signals (NLS, shown in an orange diamond shape), a phosphodegion, and a C-terminal ZF domain. (B) Schematic of the three human *PCDH* gene clusters. The *PCDH*  $\alpha$  and  $\gamma$  clusters share similar organization, with dozens of variable exons each of which is spliced to a single set of downstream constant exons. The variable exons are divided into alternate and C-type groups. The *PCDH*  $\beta$  cluster contains only variable exons. The super-enhancer of the *PCDH*  $\alpha$  cluster (red ellipses) is located between the *PCDH*  $\alpha$  and  $\beta$  clusters. The super-enhancer of the *PCDH*  $\beta$  clusters (red ellipses) is located downstream of *PCDH*  $\gamma$ . The locations and orientations of REST/NRSF sites (NRSE) and CTCF sites (CBS) are shown under the genes. (C) Venn diagram of REST/NRSF ChIP-nexus peaks of HEC-1-B and SK-N-SH cells. (D) Schematic of the NRSE locations within the four subgroups of the clustered *PCDH* variable exons. (E–G) EMSA experiments using REST/NRSF with a site ‘a’ NRSE probe of each member of the alternate *PCDH*  $\alpha$  (E), a site ‘b’ NRSE probe of each *PCDH*  $\gamma$  (F), or a site ‘c’ NRSE probe of each member of the alternate *PCDH*  $\gamma$  (G), using mock as a control (Ctr). Supershifted bands were detected with a specific antibody against human MYC (Ab) tag fused to the C-terminal of REST/NRSF.

CTCF (CCCTC-binding factor) is a master regulator of clustered *PCDH* genes (25,31,33–38). There are tandem arrays of forward CTCF-binding sites (CBSs) associated with the *PCDH* promoters and of reverse CBSs associated with the super-enhancers (Figure 1B) (31,33). Through CTCF/cohesin-mediated topological chromatin interactions between enhancers and promoters, clustered *PCDH* genes are stochastically and unbiasedly expressed (33,38). By contrast, REST/NRSF has been shown to repress expression of the clustered *PCDH* genes (29,39). However, the mechanism by which REST/NRSF represses these clustered genes remains unknown.

Here, we identified NRSEs in every alternate exon of the human *PCDH*  $\alpha$  and  $\gamma$  clusters, as well as in each C-type gene. By systematic EMSA (electrophoretic mobility shift assay) experiments, in conjunction with computational molecular dynamics, we found that REST/NRSF recognizes NRSEs via tandem ZF domains in a directional and

flexible manner. Moreover, through genetic experiments we found that REST/NRSF inhibits long-distance chromatin contacts between the *PCDH*  $\alpha$  *H5S-1* enhancer and its target promoters through preventing CTCF binding. Thus, REST/NRSF regulates the neural-specific expression of the clustered *PCDH*  $\alpha$  genes through modifying chromatin structures.

## MATERIALS AND METHODS

### Cell culture and transfection

HEK293T cells were cultured in Dulbecco’s modified Eagle’s medium (Hyclone) supplemented with 10% (v/v) FBS (Gibco), and 1% penicillin–streptomycin (Gibco). HEC-1-B and SK-N-SH cells were cultured in MEM (Hyclone), supplemented with 10% (v/v) FBS, 2 mM Glutamax (Gibco), 1 mM sodium pyruvate (Sigma), and 1% penicillin–streptomycin. Cells were maintained at 37°C in a

5% (v/v) CO<sub>2</sub> incubator. Cells were transfected by Lipofectamine 2000 (Invitrogen) when the cell confluency reached 70–90%.

### Animals

C57BL/6 and ICR mouse strains were housed at 23°C on a 12/12 h light–dark cycle (7:00 am–19:00 pm) in an SPF (specific pathogen free) facility. The zygotes were obtained from the oviducts of superovulated female C57BL/6 mice mated with the C57BL/6 stud male mice. All experiments were carried out in accordance with the protocol approved by the Institutional Animal Care and Use Committee of Shanghai Jiao Tong University (protocol#: 1602029).

### Plasmid construction

To generate plasmids for *in-vitro* expression of REST/NRSF, the coding sequences of REST/NRSF ZF1–8 were amplified by PCR using cDNA as templates and cloned into the pTNT vector (Promega) between the EcoRI and XbaI sites, with a MYC tag in the 3' terminal. A series of ZF-deleted and ZF-mutated REST/NRSF-expressing plasmids were constructed by PCR using ZF1–8 encoding region as templates and ligated into the pTNT vector between the EcoRI and XbaI sites, also with a MYC tag in the 3' terminal.

To prepare plasmids for generating probes used in EMSA experiments, the sequences containing putative NRSE motifs were amplified by PCR from the human genomic DNA and then subcloned into the pGEM-T Easy vector (Promega). To prepare plasmids for generating the mutated probes, the mutated sequences were constructed by PCR from the wild-type plasmids and then ligated into the pGEM-T Easy vector.

The plasmids for REST/NRSF knockdown and for GFP control were constructed by ligating annealed primer pairs into the pLKO.1 vector (Addgene) between the EcoRI and AgeI sites. Plasmids for sgRNA expression were constructed by inserting annealed primer pairs into a BsaI-linearized pGL3 vector under the control of the U6 promoter. All constructs were confirmed by Sanger sequencing and all primers are shown in Supplementary Table S1.

### Lentivirus packaging and infection

The pLKO.1-plasmids for REST/NRSF knockdown were co-transfected into HEK293T cells with the psPAX2 and pMD2.G helper plasmids (Addgene) using Lipofectamine 2000 (Life Technologies) to produce lentiviral particles. HEC-1-B cells and HEK293T cells at the confluency of 70–90% were infected with the virus in the presence of 8 µg/ml polybrene (Sigma). Puromycin (Sigma) was added at a final concentration of 2 µg/ml to select the infected cells. Fresh puromycin-containing medium was changed every other day. Cells were collected for assays at day 5 post infection. After harvesting by lysing cells with RIPA lysis buffer (50 mM Tris–HCl, pH 7.4, 150 mM NaCl, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS, 1 mM PMSF), the expression levels of REST/NRSF were measured by western blot.

### Western blot

Proteins were denatured and separated by SDS–PAGE and transferred to nitrocellulose membranes. These membranes were then incubated with mouse anti-MYC (Millipore), rabbit anti-REST/NRSF (Millipore), or rabbit anti-β-actin antibody (Abcam). Finally, the membranes were incubated with anti-mouse or anti-rabbit secondary antibodies and scanned using the Odyssey System (LI-COR Biosciences).

### Electrophoretic mobility shift assays (EMSA)

EMSA experiments were performed as described (31,36) with some modifications. Proteins used for EMSA experiments were synthesized *in-vitro* from pTNT plasmids using TNT T7 Quick Coupled Transcription/Translation System (Promega) according to the manufacturer's protocol. Briefly, ZF-deleted or ZF-mutated pTNT plasmids were mixed gently with TNT T7 Quick Master Mix, Methionine, and T7 TNT PCR Enhance by pipetting, and then incubated at 30°C for 60–90 min.

Probes were generated by PCR with high-fidelity polymerase using 5' biotin-labeled primers from template-containing plasmids and gel-purified. The primers used are listed in Supplementary Table S1. Probe concentration was measured with NanoDrop (Thermo). Each binding reaction contained equimolar of the biotin-labeled probes. Protein concentration was determined by western blot and each binding reaction contained the same amounts of proteins.

EMSA was performed with the LightShift Chemiluminescent EMSA Kit (Thermo) according to the manufacturer's manuals. Briefly, the *in-vitro*-synthesized proteins were precleared with the binding buffer containing 10 mM Tris–HCl, 250 mM KCl, 2.5 mM MgCl<sub>2</sub>, 0.1 mM ZnSO<sub>4</sub>, 1 mM DTT, 0.1% NP-40, 50 ng/µl poly (dI–dC), and 2.5% (v/v) glycerol on ice for 20 min. Fifty fmol of biotin-labeled probes were then added and the reactions were incubated at room temperature for 20 min. One µg of anti-MYC antibody was added into the binding reaction and incubated at room temperature for another 20 min for the supershift experiments. The binding reactions were electrophoresed on 5% nondenaturing polyacrylamide gels in ice-cold 0.5× TBE buffer (pH 8.0) and transferred to a nylon membrane. After crosslinking under UV-light for 10 min, the membranes were blocked in the Blocking Buffer by incubating for 15 min with gentle shaking and then incubated in the Stabilized Streptavidin–Horseradish Peroxidase Conjugate solution for 15 min with gentle shaking. After rinsing the membrane with 1× washing buffer briefly, the membrane was washed four times for 5 min each in 1× washing buffer with gentle shaking. Then, the membrane was incubated in substrate equilibration buffer for 5 min with gentle shaking, followed by incubation in the substrate working solution for 5 min without shaking and exposure using ChemiDoc XRS+ System (Bio-Rad).

### ChIP-nexus

ChIP-nexus experiments were performed as described (40) with some modifications. Briefly, ~2 × 10<sup>7</sup> cells were cross-linked with 1% formaldehyde for 10 min at room temperature, followed by quenching the crosslinking with glycine at

a final concentration of 0.125 M and then spun down. Subsequently, cell pellets were lysed twice with ice-cold ChIP buffer (10 mM Tris-HCl, pH 7.5, 0.15 M NaCl, 1% Triton X-100, 1 mM EDTA, 0.1% SDS, 0.1% sodium deoxycholate, 1× protease inhibitors) by incubating at 4°C for 10 min with slow rotation. Nuclei were spun down and then resuspended in 700 µl of the ChIP buffer. After incubating on ice for 10 min, the samples were sonicated using a Bioruptor Sonicator on high power (30 s on/30 s off) for 30 min to fragmentize DNA to sizes ranging from 100 to 10000 bp. The sheared chromatin solutions were immunoprecipitated with a specific antibody against REST/NRSF (Millipore, 4 µg for each reaction) by slow rotation at 4°C overnight. Antibody-precipitated complexes were incubated with 50 µl of Protein A/G Magnetic beads (Thermo) at 4°C for another 3 h the next day. The chromatin-enriched magnetic beads were then washed with the washing buffer A (10 mM TE, 0.1% Triton X-100), washing buffer B (150 mM NaCl, 20 mM Tris-HCl, pH 8.0, 5 mM EDTA, 5.2% sucrose, 1.0% Triton X-100, 0.2% SDS), washing buffer C (250 mM NaCl, 5 mM Tris-HCl, pH 8.0, 25 mM HEPES, 0.5% Triton X-100, 0.05% sodium deoxycholate, 0.5 mM EDTA), washing buffer D (250 mM LiCl, 0.5% NP-40, 10 mM Tris-HCl, pH 8.0, 0.5% sodium deoxycholate, 10 mM EDTA), and finally the Tris buffer (10 mM Tris-HCl, pH 7.5). Washing volumes were 1 ml per sample. After the washing buffer was added, the tubes were briefly inverted by hand to resuspend the beads for each washing. The beads were resuspended every ~15 min by gently tapping the tubes for all the following incubations.

The DNA-complex-coated beads were incubated with the end-repair enzyme mixture (NEB) at 20°C for 30 min to repair the DNA ends and then with Klenow exo- (NEB) in the NEB buffer 2 containing 0.2 mM dATP at 37°C for 30 min for dA tailing. The samples were then ligated with the annealed Nexus adaptors (Nex\_adaptor\_UBamHI: 5' phosphate GATCG GAAGA GCACA CGTCT GGATC CACGA CGCTC TTCC, Nex\_adaptor\_Barcode\_BamHI: 5' phosphate TCAGA GTCGA GATCG GAAGA GCGTC GTGGA TCCAG ACGTG TGCTC TTCCG ATCT) with 2× Blunt/TA ligase master mix (NEB) at 25°C for 1 h. The adaptors contained a pair of sequences for library amplification, a BamHI site for later linearization, a nine-nucleotide barcode containing five random bases and four fixed bases. Subsequently, the samples were treated with Klenow exo- at 37°C for 30 min to fill-in the ends of the adaptors and then trimmed with T4 DNA polymerase (NEB) at 12°C for 5 min. The blunt-ended DNA was then treated with Lambda Exonuclease (NEB) at 37°C for 60 min with constant rotation to digest one strand of the double-stranded DNA in a 5' to 3' direction until encountering a cross-linked protein. The samples were then digested by RecJ<sub>f</sub> exonuclease (NEB) for 60 min at 37°C to degrade the single-stranded DNA from the 5'-end and washed three times with the RIPA buffer (50 mM HEPES, pH 7.5, 1 mM EDTA, 0.7% sodium deoxycholate, 1% NP-40, 0.5 M LiCl).

The DNA-protein complex was eluted with 200 µl of elution buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS) at 65°C, shaken at 1000 rpm for 30 min, and then reverse-crosslinked at 65°C overnight. The next day, after

incubating the samples with 2 µl of RNase A (Thermo) at 37°C for 2 h and then with 8 µl of proteinase K (NEB) at 55°C for 4h, ssDNA was extracted with phenol:chloroform:isopentanol (25:24:1, v/v), and precipitated with 2.5× (v/v) of ethanol, 1/10 (v/v) of 3 M sodium acetate and 1.5 µl of glycogen (20 mg/ml, Thermo). The DNA pellet was resuspended in 10 µl of nuclease-free water.

After being denatured at 95°C for 5 min, the ssDNA was circularized with CircLigase (Epicentre) at 60°C for 1 h. The samples were annealed with the BamHI\_cut\_oligo (GAAGA GCGTC GTGGA TCCAG ACGTG) in the digesting buffer (Thermo) in a thermocycler using the annealing program (95°C for 1 min, slowly cooled down at 1% ramp to 25°C for 1 min, 25°C for 30 min, hold at 4°C) and then digested with BamHI (Thermo) at 37°C for 30 min. The linearized DNA were precipitated using ethanol and sodium acetate, then resuspended with 20 µl of water. The DNA was amplified by Q5 polymerase (NEB) with the Illumina primers with the PCR program (98°C for 30 s, 16 cycles of 98°C for 10 s and 65°C for 75 s, 65°C for 5 min, hold at 4°C). The PCR products with sizes from 150 bp to 300 bp were extracted from 2% agarose gel using MinElute Gel Extraction Kit (Qiagen).

ChIP-nexus library DNA was sequenced on an Illumina HiSeq X Ten platform, and reads were filtered for the presence of the fixed barcode CTGA starting from the sixth position of reads. The random and fixed barcode sequences were then removed. Adaptor sequences at the 3' end of reads were then trimmed using the Cutadapt tool. The trimmed reads were aligned to the human genome (hg19) using Bowtie2, and then analyzed by MACS2, MACE and MEME suites. The heatmaps were generated by an R package named pheatmap.

### RNA-seq

RNA-seq experiments were performed as previously described (31,33). Briefly, total RNA was extracted from cultured cells or mouse tissues using TRIzol Reagent (Ambion). RNA-seq experiments were performed using the NEBNext Ultra RNA Library Prep Kit for Illumina (NEB) according to the manufacturer's protocol. Briefly, mRNA was enriched from 1 µg of total RNA using NEBNext oligo (dT) magnetic beads, and fragmented by heating at 95°C for 10 min. After reverse transcription of the first stranded cDNA and synthesis of the second stranded cDNA, cDNA was purified using 1.8× AMPure XP beads (Beckman). The purified cDNA was end-repaired and then ligated with NEBNext Adaptors, followed by treatment with the USER enzyme (NEB). The ligated product was purified using the AMPure XP beads (Beckman). The purified cDNA product was then amplified with the Illumina primers by the Q5 enzyme (NEB) with the PCR program (98°C for 30 s, 14 cycles of 98°C for 10 s and 65°C for 75 s, 65°C for 5 min, hold at 4°C). RNA-seq libraries were sequenced on an Illumina HiSeq X Ten platform, and reads were aligned to the human or mouse genome using TopHat (v2.0.14). The expression levels were calculated using the Cufflinks software (v2.2.1) with default parameters. All RNA-seq experiments were performed with at least two biological replicates.

### Screening single-cell CRISPR clones by DNA-fragment editing

The generation of the CRISPR single-cell clones by DNA-fragment editing was performed as previously described (41,42). Briefly, HEC-1-B or HEK293T cells at ~80% confluency were transfected with Cas9 plasmids (0.6  $\mu$ g) and two *HS5-1*-NRSE-targeted sgRNA-expressing plasmids (1.2  $\mu$ g) by Lipofectamine 2000 (Life Technologies) in a 12-well plate. Two days post transfection, puromycin (Sigma) was added at a final concentration of 2  $\mu$ g/ml. Four days later, cells were changed to fresh medium without puromycin and cultured for another 2 days. The cells were then diluted and plated into 96-well plates with approximately one cell per well. Two weeks later, single-cell clones were marked manually under a microscope and screened for targeted deletion by PCR. At least two single-cell clones for each deletion were obtained. We screened for a total of 133 single-cell clones, and 4 homozygous clones were obtained and analyzed. Single-cell clones for each editing were confirmed by Sanger sequencing and the screening primers are shown in Supplementary Table S1.

### Generation of *HS5-1*-NRSE-deleted mice by CRISPR DNA-fragment editing

The generation of CRISPR mice was performed as previously described (33,41). Briefly, Cas9 mRNA and a pair of sgRNAs targeting the NRSE sequences were injected into zygotes. Cas9 mRNA for mouse zygote injection was *in-vitro* transcribed from the XbaI-linearized Cas9 plasmid with a T7 promoter using mMACHINE T7 Ultra Kit (Life Technologies). The sgRNAs were *in-vitro* transcribed from PCR products with the MEGAShortscript Kit (Life Technologies). The PCR products were amplified with a forward primer containing a T7 promoter followed by targeting sequences and a common reverse primer. Then, the transcribed RNA was purified with the MEGAclear Kit (Life Technologies) and eluted with the TE buffer (0.2 mM EDTA).

Zygotes obtained from the oviducts of super-ovulated C57BL/6 mice were injected with Cas9 mRNAs (100 ng/ $\mu$ l) and a pair of sgRNAs (50 ng/ $\mu$ l each). After culturing with KSOM medium (Millipore) for 0.5 h at 37°C in a 5% CO<sub>2</sub> incubator, the injected embryos were transplanted into the oviducts of pseudo-pregnant ICR foster mothers. The F0 mice with *HS5-1* NRSE deletion were maintained and crossed to obtain F1 mice. F1 mice were genotyped again for heterozygous deletion. Heterozygous mice were then crossed to obtain homozygous mice. The wildtype littermates were used as controls.

For genotyping, mouse tails were lysed with 40  $\mu$ l of the alkaline lysis buffer (25 mM NaOH and 0.2 mM disodium EDTA, pH 12.0) at 98°C for 40 min, and then neutralized with equal volume of the neutralizing buffer (40 mM Tris-HCl, pH 5.0). 1  $\mu$ l of the solution containing genomic DNA was used as the template for PCR in a total volume of 20  $\mu$ l to screen for NRSE deletion with specific primers (Supplementary Table S1).

### ChIP-seq

ChIP was performed as previously described (31,33) with some modifications. Briefly, the P0 mouse tissues were dispersed by 0.0125% (w/v) (for brain) or 0.0625% (w/v) (for kidney) collagenase (Sigma) treatment in DMEM supplemented with 10% (v/v) FBS for 20 min at 37°C by rotating at 700 rpm. Cells were then filtered through a 100- $\mu$ m cell strainer (CORNING) to obtain single-cell suspension.

About  $5 \times 10^6$  cells were cross-linked with 1% formaldehyde for 10 min at room temperature, followed by quenching the crosslinking with glycine at a final concentration of 0.125 M and then spun down. Subsequently, cell pellets were lysed twice with the ChIP buffer (10 mM Tris-HCl, pH 7.5, 0.15 M NaCl, 1% Triton X-100, 1 mM EDTA, 0.1% SDS, 0.1% sodium deoxycholate, 1 $\times$  protease inhibitors) by incubating at 4°C for 10 min with slow rotation. Nuclei were spun down and resuspended in 400  $\mu$ l of the ChIP buffer (for human cells) or ChIP buffer with 0.4% SDS (for cells from mouse tissues), and then sonicated using the Vibra-cell ultrasonic processor (Sonics) (25% maximum, a train of 20 s on and 20 s off for 15 cycles). The sonicated solution was diluted with the ChIP buffer (1:5), pre-cleared with 40  $\mu$ l Protein A agarose beads (Millipore), and then immunoprecipitated with specific antibodies against REST/NRSF (Millipore, 4  $\mu$ g for each reaction), CTCF (Millipore, 2.5  $\mu$ g for each reaction), H3K4me3 (Millipore, 2.5  $\mu$ g for each reaction), and H3K27ac (Abcam, 2.5  $\mu$ g for each reaction). The protein-DNA complexes were enriched with 40  $\mu$ l Protein A/G magnetic beads (Thermo) by incubating at 4°C for another 3h with slow rotation, and washed once with 1 ml of ChIP buffer, once with 1 ml of ChIP buffer with 0.4M NaCl, once with ChIP buffer without NaCl, and once with ml of LiCl buffer (250 mM LiCl, 5% NP-40, 0.5% sodium deoxycholate), by incubating at 4°C for 10 min with slow rotation. Then, the samples were eluted with elution buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS) at 65°C, shaken at 1000 rpm for 30 min, and then reverse-crosslinked by heating at 65°C overnight. The DNA was purified with phenol-chloroform and precipitated with ethanol, and then used for library preparation.

Libraries were prepared according to the manual of Universal DNA Library Prep Kit for Illumina V2 (Vazyme). Briefly, DNA was end-repaired using End Prep Mix 3, ligated with index-containing adaptors using rapid DNA ligase in the rapid ligation buffer (66 mM Tris-HCl pH 7.6, 20 mM MgCl<sub>2</sub>, 2 mM DTT, 2 mM ATP, 7.5% PEG 6000), and amplified using primers and amplification mix to generate libraries. The library fragments of ~350 bp (insert plus adaptor and PCR primer sequences) were selected and isolated with Agencourt AMPure XP beads (Beckman). Library DNA was sequenced on an Illumina HiSeq X Ten platform. The reads were sorted by indexes, aligned to hg19 or mm9 using Bowtie2, and peak-called using MACS2. All of the ChIP-seq experiments were performed with at least two biological replicates.

### Quantitative high-resolution chromosome conformation capture copy (QHR-4C)

QHR-4C experiments were performed as previously described (33) with some modifications. Briefly, cells were har-

vested from mice as described in ChIP-seq experiments. Then, ~2 million cells were crosslinked with 1% formaldehyde for 10 min at room temperature and the crosslinking reaction was quenched by glycine at a final concentration of 0.125 M. After spun down and washed twice with 1× PBS, cell pellets were lysed twice with cold lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% Triton X-100, 5 mM EDTA, 0.5% NP-40, 1× protease inhibitors). The pellet was resuspended in 225 μl of water, 30 μl of 10× DpnII buffer, and 7.5 μl of 10% SDS and incubated at 37°C for 1 h with constant shaking at 900 rpm. After quenching by adding 37.5 μl of 20% Triton X-100 and incubating at 37°C for 1 h with shaking at 900 rpm, the nuclei were then digested with 10 μl of DpnII (10 unit/μl) *in situ* at 37°C with shaking at 900 rpm overnight. After enzyme inactivation at 65°C for 20 min, nuclei were centrifuged at 2500 g for 5 min, resuspended and ligated in 358 μl of water, 40 μl of 10× T4 ligation buffer, and 2 μl of T4 DNA ligase (400 unit/μl) overnight at 16°C. The ligated product was then reverse crosslinked by adding 10 μl of proteinase K (10 mg/ml) and heating at 65°C for 4 h, followed by adding 2 μl of RNase A (NEB) and incubating at 37°C for 45 min. DNA was then extracted using phenol-chloroform. 2 μl of glycogen (20 mg/ml) was added to facilitate DNA precipitation. The precipitated DNA was dissolved in 50 μl of water, and then sonicated to 200–600 bp using the Bioruptor system with a low energy setting at a train of 30 s on with 30s off for 12 cycles.

With the fragmented DNA as a template, a linearized amplification step was applied using a 5′ biotin-tagged primer (Supplementary Table S1) complementary to the viewpoint fragment in 100 μl of PCR system. The PCR product was denatured at 95°C for 5 min and immediately chilled on ice to obtain ssDNA. The ssDNA was enriched and purified with Streptavidin Magnetic Beads (Thermo) according to the manufacturer's instructions, and then ligated with annealed adaptors in 45 μl of the ligation system (20 μl of DNA-coated beads, 4.5 μl of 10× T4 ligation buffer, 10 μl of 30% PEG 8000, 1 μl of 50 μM adaptor, 0.9 μl of T4 DNA ligase, 8.6 μl of water). The beads were then washed twice to remove the free adaptors with the B/W buffer (5 mM Tris-HCl, 1 M NaCl, 0.5 mM EDTA, pH 7.5). The DNA-coated beads were then resuspended in 10 μl of water. With the DNA on beads as a template, the QHR-4C libraries were generated by PCR amplification with a pair of primers (Supplementary Table S1). Finally, the libraries were purified with a PCR purification kit (Qiagen) and sequenced on an Illumina HiSeq X Ten platform. Reads were sorted using indexes and barcodes, mapped to the mouse (NCBI37/mm9) or human (GRCh37/hg19) genomes using Bowtie2, and calculated using the r3Cseq program (version 1.20) in the R package (version 3.3.3). All of the QHR-4C experiments were performed with at least two biological replicates.

### Structural modeling and molecular dynamics (MD) simulations

Structural modeling and MD simulations were performed as previously described (43). Briefly, we first built the models of REST-*HS5-1*-NRSE and REST-*PCDHγa6*-NRSE

based on the crystal structure of the ZF protein PR/SET domain 9 (PRDM9) in complex with DNA (PDB id: 5v3g). The REST/NRSF structure and DNA was modeled by directly replacing the DNA nucleobases in 5v3g to the ones of *HS5-1*-NRSE or *PCDHγa6*-NRSE. In particular, the orientation of the REST-ZF6 in REST-*HS5-1*-NRSE in respect to DNA was constructed according to the formerly captured CTCF-ZF8 structure that displays a non-DNA-penetration conformation (44,45).

Then, the above two DNA-bound REST/NRSF complexes were subject to energy minimization, followed by MD simulations. The protein/DNA were described using the AMBER force field ff14SB, with the bsc1 corrections used for the DNA nucleotides. Each solute was immersed in a cubic box filled with the TIP3P waters, and the appropriate number of Na<sup>+</sup> ions were added to neutralize the whole system. The final structure includes a total of 49652 and 68845 atoms for REST-*PCDHγa6*-NRSE and REST-*HS5-1*-NRSE, respectively. Energy minimization was conducted using the steepest descent and conjugate gradient methods. Then, each complex was heated from 0 to 310 K within 200-ps MD simulations, followed by 200-ps equilibrium MD simulations by constraining all of the solute heavy-atoms. Finally, we performed three parallel 50-ns production MD simulations, each initiated from different velocities. The temperature was controlled by the Langevin thermostat at 310 K and the SHAKE algorithm was used to constrain the bond lengths involving hydrogen atoms. The non-bonded cutoff distance was set as 12 Å, and the long-range electrostatic interactions were treated by the particle mesh Ewald (PME) method. All the MD simulations were conducted using the AMBER 2018 package.

### DNA methylation analysis

About  $5 \times 10^5$  cells were harvested from mice and crosslinked with formaldehyde as described above in the ChIP-seq experiments. The harvested cells were then lysed and reverse-crosslinked by adding 400 μl of 1× TE buffer containing 5 μl of proteinase K (10 mg/ml) and heating at 65°C for 4 h. The genomic DNA was extracted from the lysed solution with phenol: chloroform: isopentanol (25:24:1, v/v) and resuspended in 40 μl of nuclease-free water. Then bisulfite conversions of the DNA were performed using the EpiTect Plus DNA Bisulfite Kit (Qiagen) according to the manual. Briefly, DNA was converted with bisulfite solution in the thermal cycler under the condition (95°C for 5 min, 60°C for 25 min, 95°C for 5 min, 60°C for 1 h 5 min, 95°C for 5 min, 60°C for 2 h 55 min). Then, the converted DNA was desulfonated and cleaned up with the MinElute DNA spin column, amplified using the bisulfite sequencing primers (Supplementary Table S1) with GoTaq G2 Green Master Mix (Promega) under the condition (95°C for 10 min, 38 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 30 s, and a final extension at 72°C for 5 min). The PCR products were gel-purified using MinElute Gel Extraction Kit (Qiagen) and ligated into the pGEM-T Easy vector (Promega) for Sanger sequencing. The results were analyzed using a quantification tool for methylation analysis (QUMA) (46).

## RESULTS

### A repertoire of NRSEs in clustered *PCDH* genes

To investigate mechanisms of neural-specific expression of the clustered *PCDH* genes in the brain, we performed REST/NRSF ChIP-nexus experiments in model cells of neuroblastoma SK-N-SH and endometrial carcinoma HEC-1-B (33,36,38) and found 5276 and 9681 REST/NRSF sites, respectively (Figure 1C). In the clustered *PCDH* loci, we found a repertoire of numerous REST/NRSF sites within variable regions of the *PCDH*  $\alpha$  and  $\gamma$ , but not  $\beta$ , clusters (Figure 1B, Supplementary Figure S1A). In particular, there are two REST/NRSF sites in the variable exons of *PCDH* $\gamma$ *a* (Supplementary Figure S1B). Sequence analyses identified three conserved REST/NRSF sites, designated as 'a', 'b' and 'c', within the *PCDH* variable exons (Figure 1D, Supplementary Figures S2 and S3). Specifically, site 'a' in *PCDH* $\alpha$ , site 'b' in *PCDH* $\gamma$ *a*, and site 'c' in the *PCDH* $\gamma$  alternate exons match the REST/NRSF consensus motif (Supplementary Figures S2 and S3) (16).

To investigate the recognition mechanisms of these *PCDH* sites, we performed comprehensive EMSA experiments using recombinant REST/NRSF proteins. For site 'a', we designed a repertoire of probes for each alternate member of the *PCDH* $\alpha$  cluster and three probes in the same location for members of the *PCDH* $\beta$  $\gamma$  clusters as negative controls. EMSA experiments revealed that site 'a' in the *PCDH* $\alpha$  cluster, but not in the *PCDH*  $\beta$  or  $\gamma$  cluster, is recognized by REST/NRSF *in vitro*. Specifically, there is a shifted band with probes for each *PCDH* $\alpha$  alternate exon and a supershifted band when incubated with a specific antibody (Figure 1E). Similarly, for site 'b', only members of the *PCDH* $\gamma$ *a* subfamily (except *PCDH* $\gamma$ *a*9) are recognized by REST/NRSF (Figure 1F). Consistent with the nonbinding to the *PCDH* $\gamma$ *a*9 probe (Figure 1F), the first and fourth positions of the *PCDH* $\gamma$ *a*9 site had been mutated from consensus 'G' and 'C' to non-consensus 'A' and 'T' nucleotides during the evolution, respectively (Supplementary Figure S3A). For site 'c', members of both the *PCDH* $\gamma$ *a* and *PCDH* $\gamma$ *b* subfamilies are recognized by REST/NRSF *in vitro* (Figure 1G). These data suggest that there is one NRSE in each alternate exon of *PCDH* $\alpha$  (site 'a') and *PCDH* $\gamma$ *b* (site 'c'), two NRSEs in each alternate exon of *PCDH* $\gamma$ *a* (sites 'b' and 'c', except *PCDH* $\gamma$ *a*9), and no NRSE in the exons of the *PCDH* $\beta$  cluster. Consistently, each variable exon of the *PCDH* $\beta$  cluster is an independent gene whereas every variable exon of the *PCDH* $\alpha$  and *PCDH* $\gamma$ *a* clusters is alternatively spliced to the respective constant exons. Together, these observations suggest that members of the *PCDH* $\beta$  cluster may be regulated differently from *PCDH* $\alpha$  $\gamma$ .

The five C-type *PCDH* exons are distinct from alternate exons both in their encoded protein sequences and regulatory mechanisms (22,36,38,47,48). We performed EMSA experiments for each member of the C-type exons and found that each contains a REST/NRSF site (Supplementary Figure S4A and B). We also systematically examined candidate recognition sites in the super-enhancers of the *PCDH*  $\alpha$  and  $\beta$  $\gamma$  clusters by EMSA experiments and found that *HS5-1* and *HS7* each contains a REST/NRSF site (Supplementary Figure S4C and D). However, in the super-

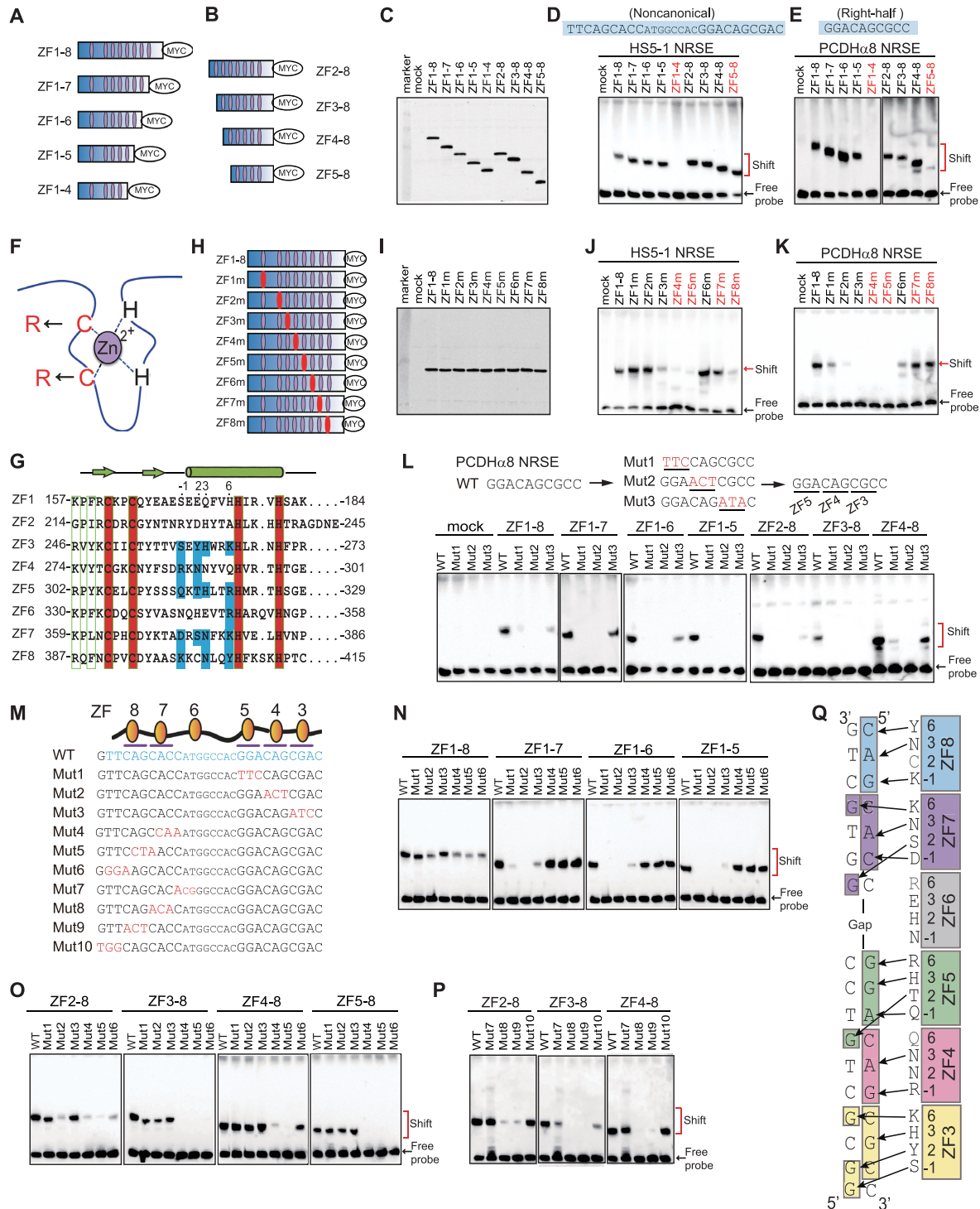
enhancer of the *PCDH* $\beta$  $\gamma$  clusters, we could not find any authentic REST/NRSF site (Supplementary Figure S4C and E).

In summary, in the *PCDH* $\alpha$  cluster, both variable exons and the downstream super-enhancer contain NRSEs; but in the *PCDH* $\beta$  $\gamma$  clusters, only *PCDH* $\gamma$  but neither *PCDH* $\beta$  nor their downstream super-enhancer contains NRSE, consistent with the neural-specific expression of *PCDH* $\alpha$ , but not *PCDH*  $\beta$  or  $\gamma$  genes (Supplementary Figure S4F) (49).

### Directional REST/NRSF binding to *PCDH* NRSEs

To investigate REST/NRSF DNA-recognition mechanisms, we generated a set of truncated REST/NRSFs through the sequential deletion of ZF domains from either N- or C-terminus (Figure 2A–C). We found that, for the series of C-terminal deletion mutants, truncation up to ZF6 (ZF1–5) does not perturb REST/NRSF binding, but truncation up to ZF5 (ZF1–4) abolishes REST/NRSF binding to the noncanonical *HS5-1* site, suggesting that ZF5 plays an important role in the recognition of the *HS5-1* site (Figure 2D). The same result was also observed using the right-half only site of *PCDH* $\alpha$ 8, suggesting that ZF5 is also essential for the recognition of the right-half site (Figure 2E). For the series of N-terminal deletion mutants, truncation up to ZF4 (ZF5–8) does not perturb the REST/NRSF binding to the noncanonical *HS5-1* site (Figure 2D), but remarkably results in a significant decrease in binding levels to the right-half only site of *PCDH* $\alpha$ 8 (Figure 2E), suggesting that ZF4 probably recognizes nucleotides within the right-half site. This is consistent with previous studies on a C-terminal truncated form of REST/NRSF (50). Finally, we tested the binding ability of ZF-truncated REST/NRSF to two canonical sites and found that ZF5 is also required for their binding (Supplementary Figure S5A).

To further investigate the differential roles of ZF domains in NRSE recognition, we substituted the two Cys residues of each REST/NRSF tandem ZF domain with two Arg residues (Figure 2F–I). Substitution of Cys in ZF1 or ZF2 with Arg residues results in a significant decrease in the REST/NRSF binding to the right-half only site of *PCDH* $\alpha$ 8, but does not perturb the binding to the non-canonical and canonical NRSEs (Figure 2J and K, Supplementary Figure S5B), suggesting that ZF1 and ZF2 are required for the binding to the right-half only site but are dispensable for REST/NRSF binding to the noncanonical and canonical NRSEs with both left- and right-half sites. In addition, substitution of Cys in ZF3, ZF4 or ZF5 with Arg residues affects the REST/NRSF binding to all NRSEs tested, including noncanonical, canonical, and right-half only sites (Figure 2J and K, Supplementary Figure S5B). This suggests that ZF3–5 recognize the right-half sites. Finally, substitution of Cys in ZF7 or ZF8 with Arg residues does not affect REST/NRSF binding to the right-half only site of *PCDH* $\alpha$ 8, but results in a significant decrease in REST/NRSF binding levels to the noncanonical and canonical NRSEs (Figure 2J and K, Supplementary Figure S5B), suggesting that ZF7–8 recognize nucleotides within the left-half site. Together, these REST/NRSF truncation and mutation experiments suggest that ZF3–5 recognize the right-half while ZF7–8 recognize the left-half sites.



**Figure 2.** Directional base-specific binding of REST/NRSF to the clustered *PCDH* NRSEs. (A, B) Schematics of truncated REST/NRSF, fusing with a MYC tag, with sequential deletions of ZFs from either N or C terminus. Purple ellipses denote zinc-finger domains. (C) Western blot of REST/NRSF with sequential deletions of ZFs. (D, E) EMSA using the truncated proteins with biotinylated probes of noncanonical *HS5-1* NRSE (D) or of the right-half only *PCDHα8* NRSE (E). The sequences of the motifs are shown above the gel. (F) Schematic of point mutations in each ZF domain. Two key cysteine residues were mutated to two arginine residues. (G) Sequence alignment of the eight tandem Cys2-His2 zinc fingers of REST/NRSF, with the residues involved in chelating Zn<sup>2+</sup> highlighted in the red background and residues involved in base contacts with NRSE motifs highlighted in the blue background. The -1, 2, 3, and 6 positions of the α-helix of each ZF domain are indicated above the sequences. (H) Diagram of the REST/NRSF mutants in each ZF domain (highlighted by red ellipses). (I) Western blot of ZF-mutated proteins. (J, K) EMSA using ZF mutated proteins with a biotinylated probe of the noncanonical *HS5-1* NRSE motif (J) or of the right-half only *PCDHα8* NRSE motif (K). (L) EMSA of the wild-type (WT) *PCDHα8* NRSE and its mutants using a set of ZF-deleted REST/NRSFs. (M) Sequences of the *HS5-1* NRSE WT (highlighted by blue) and mutant highlighted by red) motifs. (N–P) EMSA for ZF-deleted proteins using a biotinylated probe of the *HS5-1* NRSE Mut1-Mut6 (N, O) or Mut7-Mut10 (P), with the WT probe as a control. (Q) A model of base-specific contacts of ZF3–8 with DNA duplexes.



To further investigate the directionality of REST/NRSF DNA recognition, we sequentially mutated triple nucleotides (Mut1, Mut2 and Mut3) of the right-half only site of *PCDH $\alpha$ 8* and found that these mutations abolish the binding of the C-terminal truncated REST/NRSF (Figure 2L). For the N-terminal truncated proteins, Mut3 abolishes the binding of ZF2–8 and ZF3–8, but not ZF4–8, suggesting that ZF3 binds to the three nucleotides corresponding to Mut3 (Figure 2L). Considering the essential role of ZF4 and ZF5 in binding to the right-half site (Figure 2D and E, J and K), we conclude that they recognize the triple nucleotides corresponding to Mut2 and Mut1, respectively.

We next generated a series of mutated NRSE probes of the *PCDH HS5–1* noncanonical site with sequential substitution of triple nucleotides (Figure 2M) and performed comprehensive EMSA experiments using REST/NRSF with sequential ZF deletions. Similar to the right-half only site, EMSA experiments demonstrated that mutations of the right-half (Mut1–3), but not left-half (Mut4–6), of the noncanonical site affect the REST/NRSF binding only if it contains ZF1–5 (Figure 2N). By contrast, mutations of the left-half (Mut4–6), but not right-half (Mut1–3), of the noncanonical site appear to affect the REST/NRSF binding only if it contains ZF5–8 (Figure 2O), again suggesting that ZF3–5 bind to the right-half site while ZF7–8 bind to the left-half site of NRSEs. Moreover, we generated an additional series of triple nucleotide mutations of the left-half site with a different phase (Mut7–10) (Figure 2M). Remarkably, we found that Mut8 and Mut9 almost abolish the REST/NRSF binding completely but Mut7 and Mut10 have much less effect (Figure 2P), suggesting that the six nucleotides (CAGCAC) corresponding to Mut8 and Mut9 play a major role. In conjunction with the fact that ZF6 mutation does not appear to affect the REST/NRSF binding (Figure 2J, Supplementary Figure S5B), we conclude that the six conserved nucleotides of the left-half site are recognized by ZF7–8. Finally, we mutated sequences downstream of the Mut3 site (Mut11 and Mut12) and found that they do not affect REST/NRSF binding, suggesting that ZF2 and ZF1 have no major role in binding the right-half only site of *PCDH $\alpha$ 8* (Supplementary Figure S5C). Taken together, we conclude that REST/NRSF base-specifically recognizes DNA duplexes in an antiparallel manner via tandem ZF3–8 (Figure 2Q).

### ChIP-nexus peaks with multiple NRSE motifs

We previously found that some CTCF peaks contains more than one CBS element (51). We also noted that some REST/NRSF peaks contain more than one NRSE. For example, one peak in the first exon of the neural protocadherin *CELSR3* gene contains four tandem half-site motifs in the configuration of left-11bp-right-9bp-left-11bp-right (Figure 3A). These four tandem half-site motifs either function as four half-site only NRSEs or as two noncanonical NRSEs separated by 9 bp (Figure 3A). These sites have previously been shown to play an important role in the regulation of neural expression of the protocadherin *CELSR3* gene (52).

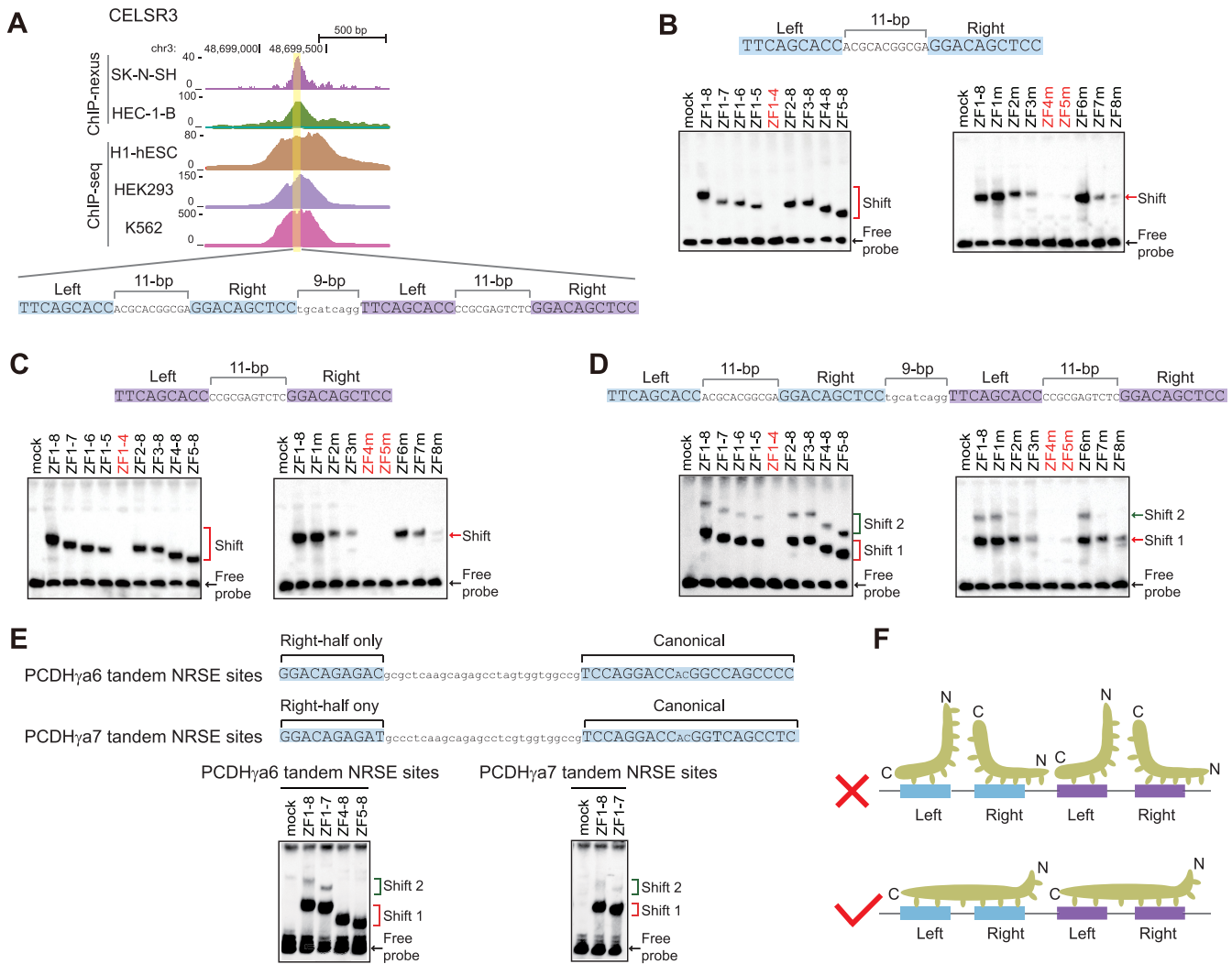
To investigate how REST/NRSF binds to these composite tandem sites, we prepared two probes each matching

one of the hypothetical noncanonical NRSEs as well as a probe containing all of the four tandem sites (Figure 3B–D). EMSA experiments showed only a single shifted band for either left or right noncanonical NRSE (Figure 3B and C); by contrast, we observed two shifted bands for the probe containing all four tandem sites (Figure 3D). Two similarly shifted bands were also observed for the *PCDH $\gamma$ a* tandem site (Figure 3E, Supplementary Figure S3). Taken together, these observations suggest that some REST/NRSF ChIP-nexus peaks contain more than one NRSE in tandem and that each NRSE is bound by one REST/NRSF protein molecule in only one orientation (Figure 3F).

### Distinct ZF6 conformations for canonical and noncanonical NRSEs

Analysis of REST/NRSF binding sites of ChIP-nexus data revealed that the gaps between left- and right-half sites are either 2 bp in canonical NRSEs or 7–9 bp in noncanonical NRSEs, but intriguingly could not be 3–6 bp (Figure 4A). To investigate how REST/NRSF tolerates the flexible length of gaps in NRSE motifs, we performed molecular dynamics (MD) experiments using ZF3–8 and the site ‘c’ canonical NRSE motif of *PCDH $\gamma$ a6* with a 2-bp gap (Supplementary Figure S3C and D). We found that these six ZFs wrap around the DNA duplex with the N-terminus of each ZF  $\alpha$ -helix inserted into the major groove and that this binding complex can remain stable during the MD simulations (Figure 4B), even though EMSA experiments demonstrated that ZF6 is not essential for binding to canonical NRSEs (Supplementary Figure S5A and B). We then performed MD experiments using ZF3–8 and the noncanonical NRSE motif of *HS5–1* with an 8-bp gap. We found that ZF3–5 and ZF7–8 wrap around the DNA duplex in the major groove and form base-specific contacts; however, in this case ZF6 flips out of the DNA major groove and remains parallel to the axis of DNA during the whole MD simulation to span the 8-bp gap (Figure 4C), consistent with no alteration of binding upon ZF6 mutation (Figure 2J). Thus, REST/NRSF recognizes canonical and noncanonical NRSEs with remarkable conformation changes.

We noted that in the 3232 canonical NRSE motifs with 2-bp gaps, the ninth position is a highly conserved base of ‘C’ (Figure 4A); however, this ‘C’ is barely conserved in the noncanonical NRSE motifs with variable large gaps of 7–9 bp (Figure 4A). Nevertheless, our EMSA experiments demonstrated that ZF7–8 and ZF3–5 contact with the left- and right-half sites, respectively, of both canonical and noncanonical NRSEs in a base-specific manner (Figures 1–3, Supplementary Figure S5). To investigate why the ninth position ‘C’ is more conserved in canonical than noncanonical NRSEs, we simulated base-recognition of ZF6 by MD in the two different conformations. We found that, in the canonical NRSEs, the side chain of Arg349 in ZF6 can form stable hydrogen bonds with the base ‘G’ in the complementary strand (Figure 4B, Supplementary Figure S5D). In the noncanonical NRSEs, however, ZF6 is not a reader of any NRSE base but is positioned parallel to the DNA axis to span the much larger gap (Figure 4C). This explains why the ninth position ‘C’ of canonical NRSEs is much more conserved than that of the noncanonical NRSEs (Figure 4A).



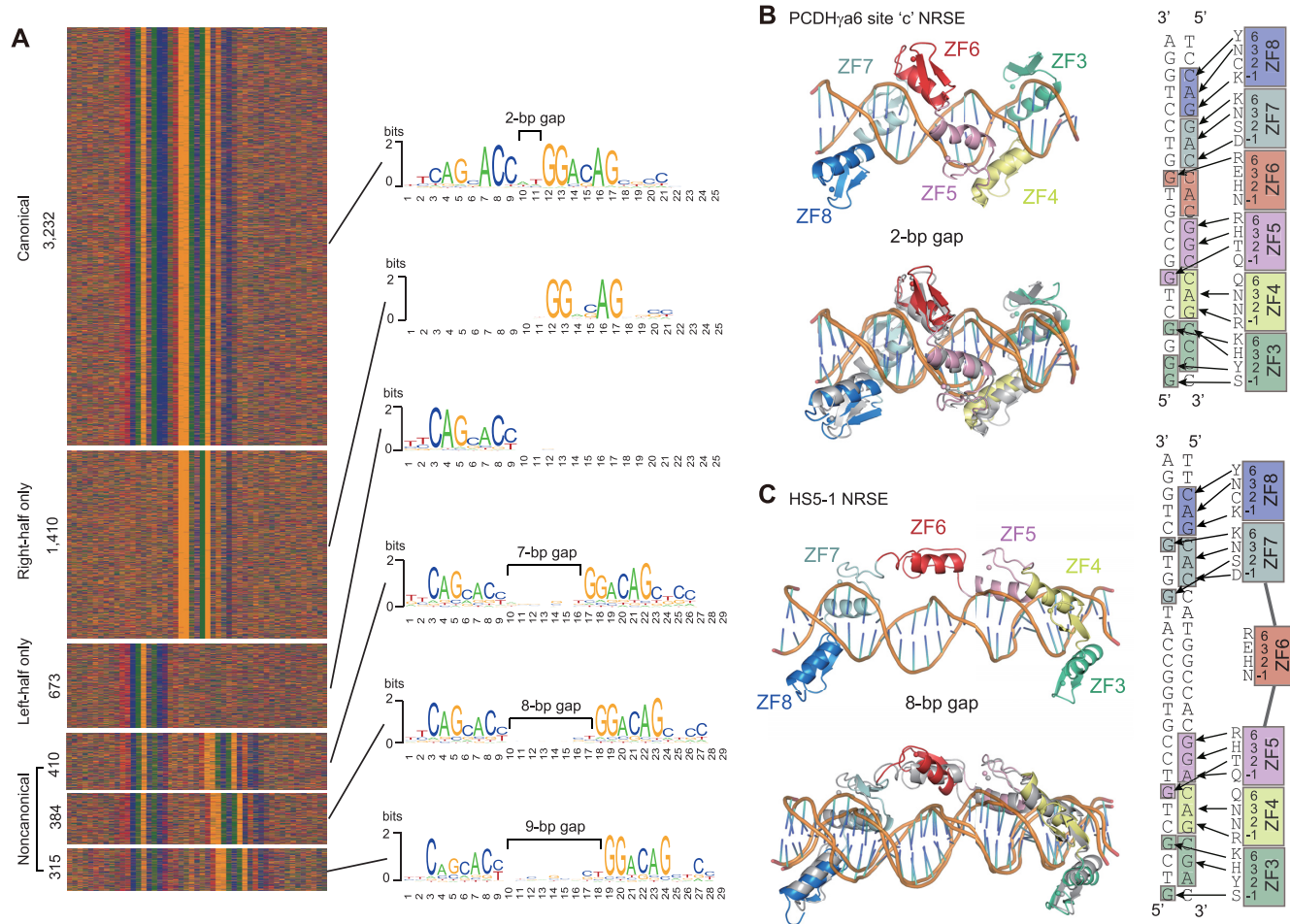
**Figure 3.** Recognition of tandem NRSE motifs by REST/NRSF. (A) Comparison of the REST/NRSF ChIP-nexus and ChIP-seq data of the human protocadherin *CELSR3* gene containing four tandem NRSE motifs. (B–D) EMSA for probes of the first two NRSE motifs (B), the last two NRSE motifs (C), or all of the four tandem NRSE motifs (D) with a repertoire of truncated or mutated REST/NRSFs, using WT as a control. (E) EMSA for probes of the *PCDH $\gamma$ 6* or *PCDH $\gamma$ 7* tandem NRSE motifs with truncated REST/NRSFs, using WT as a control. (F) Schematic of the stereo-hindrance and proper configuration of REST/NRSF recognition of tandem NRSE motifs.

### REST/NRSF inhibits long-distance enhancer-promoter contacts

Most NRSEs are located distal from promoters (16,17). To investigate how these NRSEs regulate gene expression, we focused on the *HS5-1* noncanonical site and perturbed either the REST/NRSF protein or the NRSE *cis*-element. We first knocked down REST/NRSF by designing three different shRNAs and found that REST/NRSF knockdown results in a significant increase in the expression levels of *PCDH $\alpha$ 6* and  *$\alpha$ 12* genes in HEC-1-B cells (Figure 5A and B). In addition, we found that deposition of active chromatin marks, H3K4me3 and H3K27ac, is also significantly increased in both the *HS5-1* enhancer and *PCDH $\alpha$*  promoters (Figure 5C–E), suggesting that REST/NRSF regulates *PCDH $\alpha$*  expression through epigenetic modifications or the deposition of H3K4me3 and H3K27ac is the consequence of transcription upon REST/NRSF knockdown

(53). Because the expression of *PCDH $\alpha$*  genes depends on their long-distance chromatin interactions with the distal enhancer (36), to see whether the 3D chromatin architecture of the locus is altered, we performed quantitative high-resolution chromosome conformation capture followed by next-generation sequencing (QHR-4C) experiments (33) with either *HS5-1* or *PCDH $\alpha$ 12* as a viewpoint. Remarkably, we found that there is a subtle and reproducible increase in long-distance chromatin interactions between the *HS5-1* enhancer and its target *PCDH $\alpha$*  genes upon REST/NRSF knockdown (Figure 5F and G), which explains the increased expression levels of the *PCDH $\alpha$*  genes in HEC-1-B cells (Figure 5B). Similar effects were also observed in HEK293T cells upon REST/NRSF knockdown (Supplementary Figure S6A–C).

We then deleted the NRSE within the *HS5-1* enhancer in HEC-1-B and HEK293T cells by screening single-cell

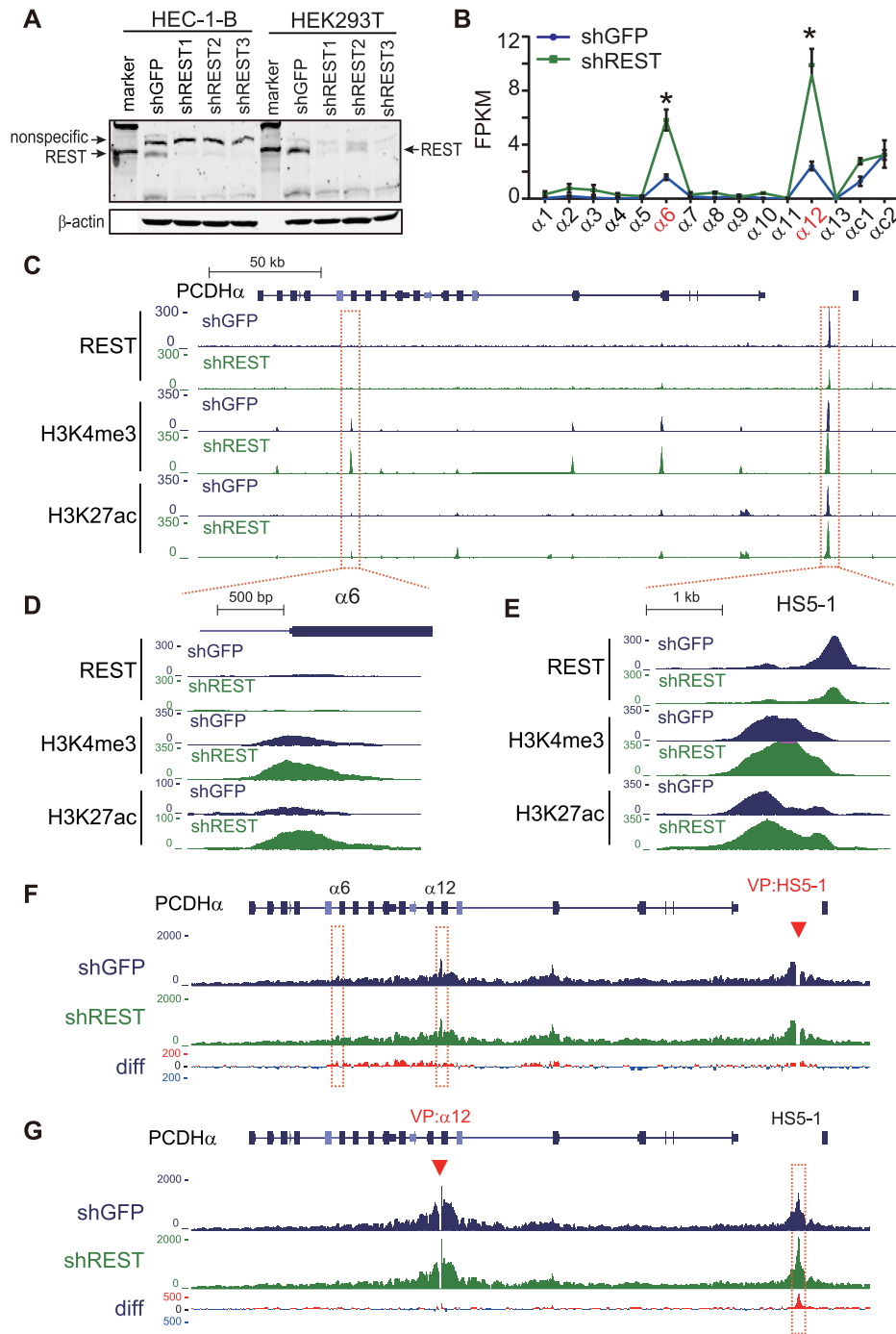


**Figure 4.** Flexible ZF6 determines two binding models of REST/NRSF with canonical and noncanonical NRSEs. (A) Six types of NRSEs and their motifs revealed by ChIP-nexus. (B, C) The molecular dynamics (MD) simulations for the base-specific contacts of ZF3–8 of REST/NRSF with site ‘c’ canonical NRSE of *PCDHγa6* with 2-bp gaps (B), or with *HS5-1* noncanonical NRSE with 8-bp gaps (C). In both (B) and (C), the upper panel is the initial structure of the MD simulations while the lower panel is the comparison of the initial and final structures. Base-specific contacts of ZF3–8 with the two NRSEs are shown on the right side of the panels.

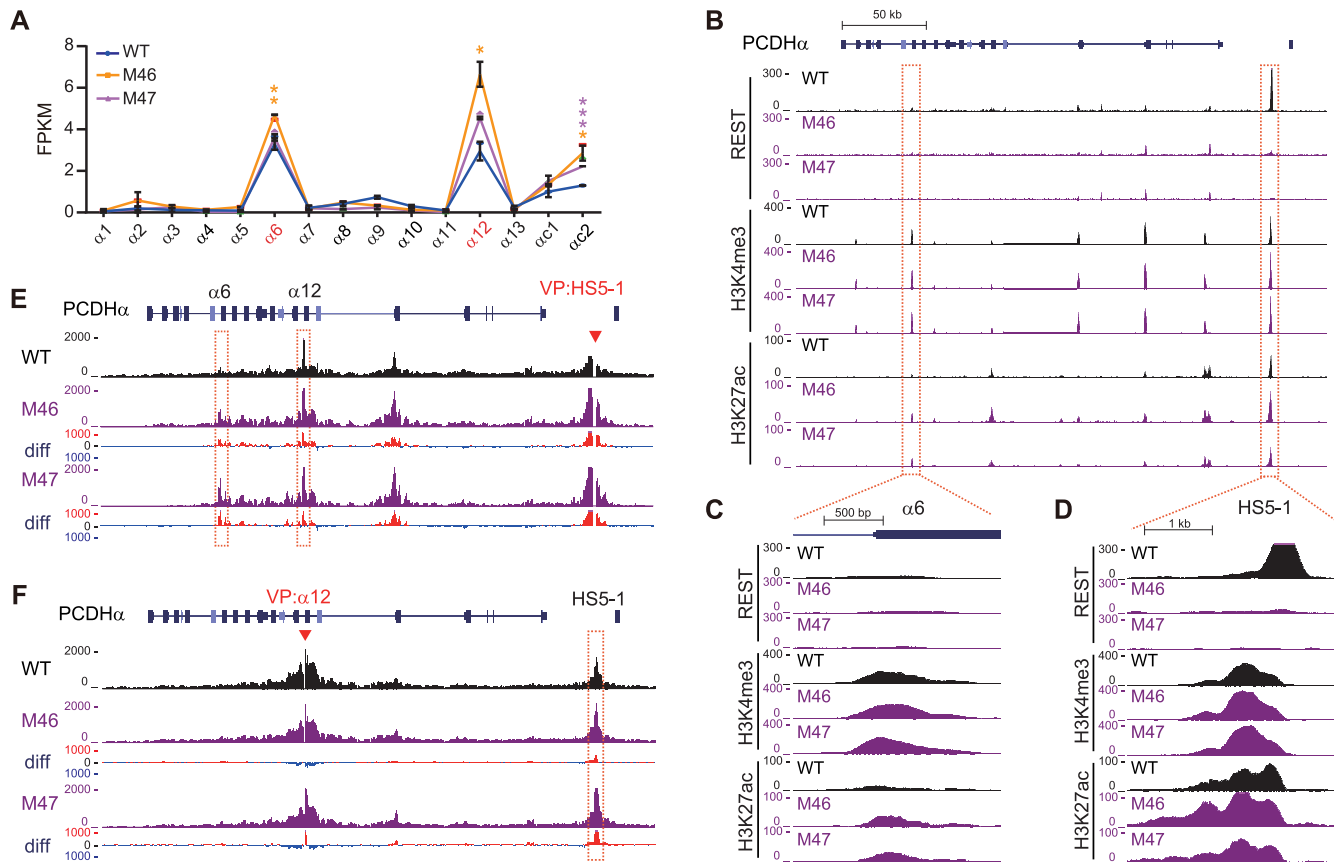
CRISPR clones through DNA-fragment editing (41,42). We obtained two single-cell homozygous clones with NRSE deletion in each cell line (Supplementary Figure S6D–G). RNA-seq revealed that deletion of the *HS5-1* NRSE results in a significant increase of expression levels of the *PCDHα6* and *α12* genes (Figure 6A, Supplementary Figure S6H). In addition, the deposition of H3K4me3 and H3K27ac histone marks is also significantly increased in both the *HS5-1* enhancer and *PCDHα* target promoters in these single-cell CRISPR clones (Figure 6B–D, Supplementary Figure S6I–K). Similar to the REST/NRSF knock-down, deletion of the *HS5-1* NRSE also results in a significant increase in long-distance chromatin interactions between the *HS5-1* enhancer and *PCDHα* target promoters (Figure 6E and F). Together, these data suggest that *HS5-1*-bound REST/NRSF suppresses the expression of *PCDHα* genes and may indirectly modulate histone tails and long-distance chromatin interactions.

#### NRSE deletion reshapes *Pcdhα* 3D chromatin structure *in vivo*

To see whether *HS5-1*-bound REST/NRSF modulates histone tails and chromatin architectures *in vivo*, we deleted the *HS5-1* NRSE in mice through CRISPR pronuclear injection with dual sgRNAs (Figure 7A, Supplementary Figure S7A and B). ChIP-seq showed that the binding of REST/NRSF to *HS5-1* is abolished upon NRSE deletion (Supplementary Figure S7C). We first confirmed that there are significantly higher levels of REST/NRSF expression in kidney than in cortical tissues by Western blot and RNA-seq (Figure 7B and C). Interestingly, there is a significant increase in expression levels of members of the *Pcdhα* cluster in kidney but not in cortical tissues upon deletion of the *HS5-1* NRSE (Figure 7D and E). In addition, the deposition of the active histone mark of H3K4me3 is significantly increased in *HS5-1* and in members of the *Pcdhα* cluster but not in *Pcdhβ1* in kidney tissues (Figure 7F).



**Figure 5.** REST/NRSF knockdown enhances *HS5-1* activity and long-distance enhancer-promoter contacts. (A) Western blot showing the REST/NRSF knockdown efficiency of the three shRNAs in HEC-1-B and HEK293T cells. (B) RNA-seq showing the increased *PCDHα* expression upon REST/NRSF knockdown in HEC-1-B cells. Data are presented as mean ± SEM. \*  $P < 0.05$ . (C–E) ChIP-seq with antibodies against REST/NRSF, H3K4me3 and H3K27ac in control and REST/NRSF knockdown cells. Note the decrease of REST/NRSF and the increase of H3K4me3 and H3K27ac upon REST/NRSF knockdown. (F, G) QHR-4C interaction profiles of the *PCDHα* locus using *HS5-1* (F), or *PCDHα12* (G) as a viewpoint (VP, arrowheads) for the control and REST/NRSF knockdown cells. Differences (shREST versus shGFP) are shown under the 4C profiles.



**Figure 6.** *HS5-1* NRSE deletion results in increased *HS5-1* enhancer activity and long-distance enhancer-promoter contacts. (A) RNA-seq results showing the increased *PCDHα6* and *α12* expression upon *HS5-1* NRSE deletion in HEC-1-B cells. M46 and M47 are two CRISPR single-cell clones with *HS5-1* NRSE deletion. Data are presented as mean  $\pm$  SEM. \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ . (B–D) ChIP-seq with a specific antibody against REST/NRSF, H3K4me3, or H3K27ac in WT and *HS5-1* NRSE deletion cells. (E, F) QHR-4C interaction profiles of the *PCDHα* locus using *HS5-1* (E), or *PCDHα12* (F) as a viewpoint (VP, arrowheads) for the WT and *HS5-1* NRSE deletion single-cell clones. Differences (deletion vs WT) are shown under the 4C profiles.

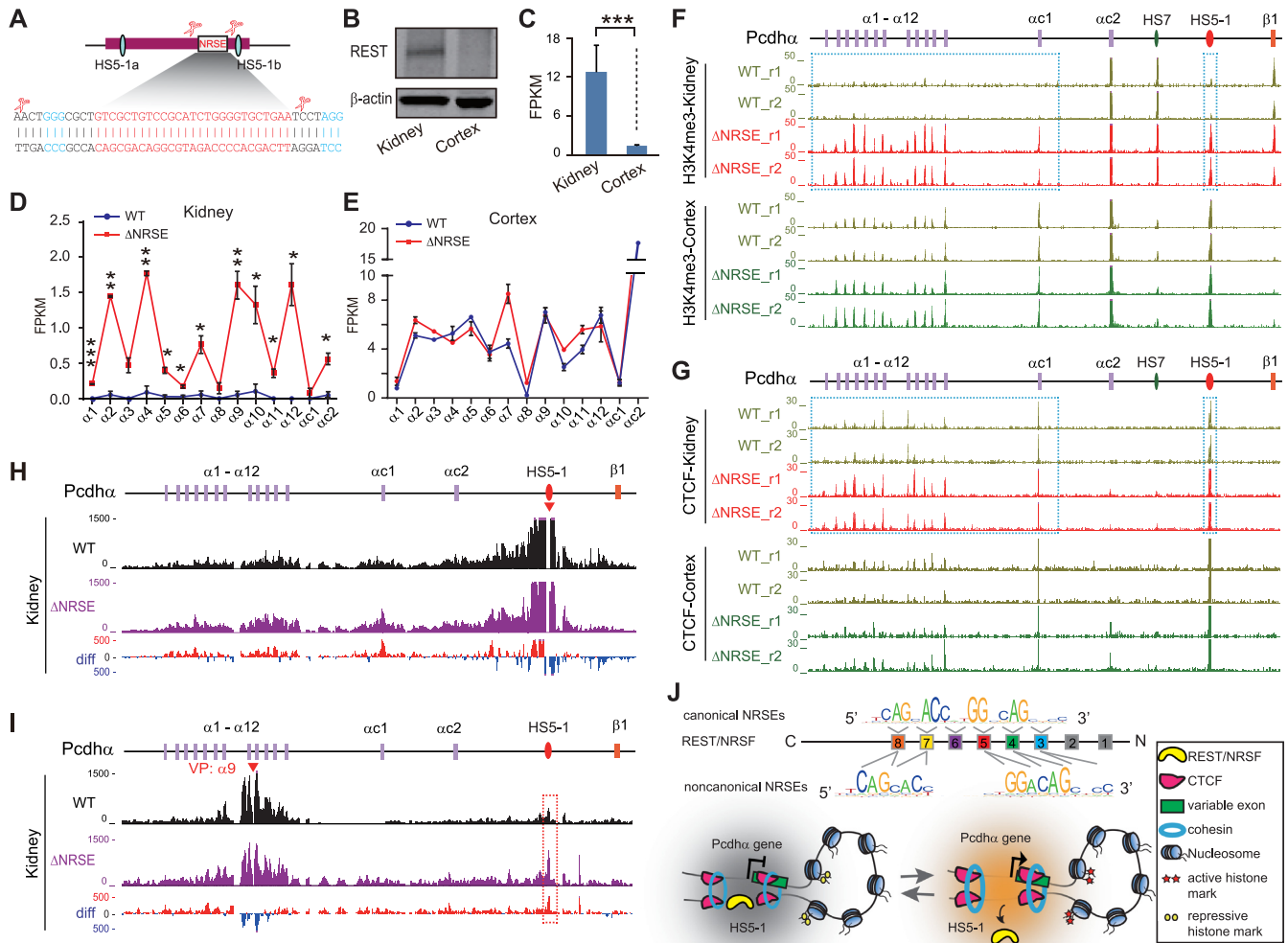
To investigate why the distal sites were affected, we performed CTCF ChIP-seq experiments and found that there is a significant increase of CTCF enrichments in *HS5-1* and in members of the *Pcdhα* cluster in kidney tissues (Figure 7G, Supplementary Figure S7D and E). In addition, there is a significant decrease of CpG methylation in kidney but not cortical tissues (Supplementary Figure S8). Finally, QHR-4C experiments demonstrated that there is a significant increase in long-distance chromatin interactions between the distal *HS5-1* enhancer and *Pcdhα* target genes upon NRSE deletion (Figure 7H and I). Considering the conserved organization of the *Pcdh* locus as well as conserved locations of CTCF (31) and REST/NRSF sites between mice and humans (Supplementary Figure S9), these data suggest that REST/NRSF represses expression of *Pcdhα* target genes and may indirectly modulate higher-order chromatin structures (Figure 7J).

## DISCUSSION

The clustered *PCDH*s participate in a wide variety of neurodevelopmental processes such as dendritic self-avoidance, axonal even spacing and tiling, spine morphogenesis and synaptogenesis, and neuronal migration and survival

(23,54). Through stochastic and combinatorial expression, the clustered *PCDH* genes encode countless assemblies of cell-surface molecules to endow each neuron with a unique identity code (33,38,47,55–58). Thus, the spatiotemporal expression patterns of the clustered *PCDH* genes must be precisely regulated during brain development. REST/NRSF is a key repressor of neural genes in neuronal progenitors and its derepression is central for neurogenesis and neuronal differentiation (2,3,8). Here, we found that REST/NRSF recognizes diverse *PCDH* NRSEs in an antiparallel manner via base-specific contacts with tandem ZF domains. In addition, MD simulations revealed that REST/NRSF endures different gap sizes of canonical and noncanonical NRSEs by adopting distinct conformations for ZF6. Finally, through genetic approaches, we demonstrated that the mechanism by which enhancer-bound REST/NRSF represses the expression of distant *PCDHα* genes and may modulate higher-order chromatin structures.

The stochastic expression of clustered *PCDHα* genes is achieved through promoter choice determined by CTCF/cohesin-mediated chromatin interactions between the *HS5-1* enhancer and its target promoters (29,31,36,38). In the *PCDHβγ* clusters, topological chromatin interac-



**Figure 7.** NRSE deletion reshapes the *Pcdha* 3D chromatin structure in mice *in vivo* via CTCF enrichments. (A) Schematic of the *HS5-1* enhancer showing the precise deletion of NRSE through CRISPR DNA-fragment editing with dual sgRNAs. The PAM sites are highlighted. The two CTCF binding sites (*HS5-1a* and *HS5-1b*) are marked with cyan ovals. (B, C) The expression levels of REST/NRSF in the kidney and cortical tissues of mice tested by western blot (B) or RNA-seq (C). (D, E) RNA-seq showing the significant increase of the *Pcdha* expression levels in mouse kidney (D), but not in cortical (E), tissues upon NRSE deletion. Data are presented as mean  $\pm$  SEM. \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ . (F, G) ChIP-seq of H3K4me3 (F) and CTCF (G) in the *Pcdha* cluster in the kidney and cortical tissues of WT and *HS5-1* NRSE-deleted ( $\Delta$ NRSE) mice. Note the significant increases of H3K4me3 and CTCF in kidney upon NRSE deletion. (H, I) QHR-4C interaction profiles of the *Pcdha* locus using *HS5-1* (H) or *Pcdha9* (I) as a viewpoint (VP, arrowheads) for the kidneys of WT and  $\Delta$ NRSE mice. Differences (deletion versus wild-type) are shown under the 4C profiles. (J) A model depicting the DNA-recognition and mechanism of REST/NRSF repressing *PCDH $\alpha$*  gene expression through CTCF/cohesin-mediated higher-order chromatin structures.

tions between tandem CTCF sites determine the balanced expression of members of the *PCDH $\beta\gamma$*  genes (33). In contrast, how *PCDH* genes are silenced in neural progenitors and nonneural tissues remains largely unknown. We showed here that the binding of REST/NRSF to the *HS5-1* enhancer inhibits the expression of the *PCDH $\alpha$*  genes through CTCF/cohesin-mediated chromatin interactions. We identified NRSEs in each member of the *PCDH $\alpha$*  and  $\gamma$  gene clusters. These sites may be related to the repression of non-chosen members of clustered *PCDH $\alpha$*  or  $\gamma$  genes in single cells in the brain because REST/NRSF has been shown to directly repress neural gene expression (6,59,60).

REST/NRSF binds to diverse NRSEs *in vivo* with distinct and hierarchical affinities according to their sequence variations (16,61,62). It is known that ZF domains are

important for REST/NRSF binding to DNA duplexes (16,50); however, it is puzzling why there exist two major classes of canonical and noncanonical NRSEs with distinct gap sizes in the human genome (16). By a combination of EMSA and molecular dynamics experiments, we found that, in the case of canonical NRSEs, each ZF domain of tandem ZF3–8 is inserted into the major groove of DNA duplexes and recognizes standard 3 basepairs (Figure 4B). By contrast, in the case of noncanonical NRSEs, ZF7–8 and ZF3–5 are inserted into the major grooves of the left- and right-half sites. However, in this case, ZF6 is flipped out of the major groove and positioned nearly parallel to the axis of the DNA duplexes, serving as a spacer element to tolerate variable distances between the two half sites. Thus, ZF6 functions as bridge-like spacer for the flexible gaps connect-

ing the left- and right-half sites of the noncanonical NRSE motifs (Figure 4C). This explains the long-standing mystery that REST/NRSF recognizes both canonical and non-canonical classes of NRSE motifs across the entire human genome (16).

MD demonstrated that, for the canonical NRSEs, the Arg349 of the ZF6 forms base-specific contacts with the 'G' base of the complementary strand at the 9<sup>th</sup> position (Figure 4B, Supplementary Figure S5D), explaining the observed conservation of the 'C' nucleotide at the ninth position of canonical NRSE motifs (Figure 4A). For the noncanonical NRSEs, however, once ZF6 has flipped out of the major groove, it cannot form base-specific contacts with any nucleobase. Instead, the flexible linkers between ZF5 and ZF6, and between ZF6 and ZF7, allow REST/NRSF to tolerate variable gap sizes between the left- and right-half sites of noncanonical NRSE motifs (Figure 4C). Finally, because of the biophysical hindrance of the flipped ZF6, the gap distance between the left- and right-half sites cannot be 3–6 bp, consistent with the fact that the flexible gaps of the non-canonical NRSEs cannot be <7 bp in the human genome (Figure 4). Together, our data shed significant insights into mechanisms by which REST/NRSF recognizes diverse genomic DNA sites and regulates gene expression.

## DATA AVAILABILITY

High-throughput sequencing files (ChIP-nexus, ChIP-seq, RNA-seq and QHR-4C) have been deposited into the NCBI Gene Expression Omnibus (GEO) database with the accession number GSE150254.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank all members of our laboratory for helpful discussions and the three anonymous reviewers for their critiques.

## FUNDING

National Natural Science Foundation of China [31630039, 91940303 to Q.W.]; Ministry of Science and Technology of China [2017YFA0504203, 2018YFC1004504]; Science and Technology Commission of Shanghai Municipality [19JC1412500].

*Conflict of interest statement.* None declared.

## REFERENCES

- Hsieh, J. and Gage, F.H. (2005) Chromatin remodeling in neural development and plasticity. *Curr. Opin. Cell Biol.*, **17**, 664–671.
- Chong, J.A., Tapia-Ramirez, J., Kim, S., Toledo-Aral, J.J., Zheng, Y., Boutros, M.C., Altshuler, Y.M., Frohman, M.A., Krainer, S.D. and Mandel, G. (1995) REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell*, **80**, 949–957.
- Schoenherr, C.J. and Anderson, D.J. (1995) The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science*, **267**, 1360–1363.
- Ooi, L. and Wood, I.C. (2007) Chromatin crosstalk in development and disease: lessons from REST. *Nat. Rev. Genet.*, **8**, 544–554.
- Chen, Z.F., Paquette, A.J. and Anderson, D.J. (1998) NRSF/REST is required *in vivo* for repression of multiple neuronal target genes during embryogenesis. *Nat. Genet.*, **20**, 136–142.
- Lunyak, V.V., Burgess, R., Prefontaine, G.G., Nelson, C., Sze, S.H., Chenoweth, J., Schwartz, P., Pevzner, P.A., Glass, C., Mandel, G. *et al.* (2002) Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science*, **298**, 1747–1752.
- Bruce, A.W., Donaldson, I.J., Wood, I.C., Yerbury, S.A., Sadowski, M.I., Chapman, M., Gottgens, B. and Buckley, N.J. (2004) Genome-wide analysis of repressor element 1 silencing transcription factor/neuron-restrictive silencing factor (REST/NRSF) target genes. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 10458–10463.
- Ballas, N., Grunseich, C., Lu, D.D., Speh, J.C. and Mandel, G. (2005) REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell*, **121**, 645–657.
- Andres, M.E., Burger, C., Peral-Rubio, M.J., Battaglioli, E., Anderson, M.E., Grimes, J., Dallman, J., Ballas, N. and Mandel, G. (1999) CoREST: a functional corepressor required for regulation of neural-specific gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 9873–9878.
- Naruse, Y., Aoki, T., Kojima, T. and Mori, N. (1999) Neural restrictive silencer factor recruits mSin3 and histone deacetylase complex to repress neuron-specific target genes. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 13691–13696.
- Huang, Y., Myers, S.J. and Dingledine, R. (1999) Transcriptional repression by REST: recruitment of Sin3A and histone deacetylase to neuronal genes. *Nat. Neurosci.*, **2**, 867–872.
- Singh, S.K., Kagalwala, M.N., Parker-Thornburg, J., Adams, H. and Majumder, S. (2008) REST maintains self-renewal and pluripotency of embryonic stem cells. *Nature*, **453**, 223–227.
- Huang, Z., Wu, Q., Guryanova, O.A., Cheng, L., Shou, W., Rich, J.N. and Bao, S. (2011) Deubiquitylase HAUSP stabilizes REST and promotes maintenance of neural progenitor cells. *Nat. Cell Biol.*, **13**, 142–152.
- Nechiporuk, T., McGann, J., Mullendorff, K., Hsieh, J., Wurst, W., Floss, T. and Mandel, G. (2016) The REST remodeling complex protects genomic integrity during embryonic neurogenesis. *Elife*, **5**, e09584.
- Tapia-Ramirez, J., Eggen, B.J., Peral-Rubio, M.J., Toledo-Aral, J.J. and Mandel, G. (1997) A single zinc finger motif in the silencing factor REST represses the neural-specific type II sodium channel promoter. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 1177–1182.
- Johnson, D.S., Mortazavi, A., Myers, R.M. and Wold, B. (2007) Genome-wide mapping of *in vivo* protein-DNA interactions. *Science*, **316**, 1497–1502.
- Jothi, R., Cuddapah, S., Barski, A., Cui, K. and Zhao, K. (2008) Genome-wide identification of *in vivo* protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res.*, **36**, 5221–5231.
- Miller, J., McLachlan, A.D. and Klug, A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.*, **4**, 1609–1614.
- Bellefroid, E.J., Marine, J.C., Ried, T., Lecocq, P.J., Riviere, M., Amemiya, C., Poncelet, D.A., Coulie, P.G., de Jong, P., Szpirer, C. *et al.* (1993) Clustered organization of homologous KRAB zinc-finger genes with enhanced expression in human T lymphoid cells. *EMBO J.*, **12**, 1363–1374.
- Najafabadi, H.S., Mnaimneh, S., Schmitges, F.W., Garton, M., Lam, K.N., Yang, A., Albu, M., Weirauch, M.T., Radovani, E., Kim, P.M. *et al.* (2015) C2H2 zinc finger proteins greatly expand the human regulatory lexicon. *Nat. Biotechnol.*, **33**, 555–562.
- Zuo, Z., Billings, T., Walker, M., Petkov, P., Fordyce, P. and Stormo, G.D. (2019) Why do long zinc finger proteins have short motifs? bioRxiv doi:10.1101/637298, 15 May 2019, preprint: not peer reviewed.
- Wu, Q. and Maniatis, T. (1999) A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell*, **97**, 779–790.
- Mountoufaris, G., Canzio, D., Nwakeze, C.L., Chen, W.V. and Maniatis, T. (2018) Writing, reading, and translating the clustered protocadherin cell surface recognition code for neural circuit assembly. *Annu. Rev. Cell Dev. Biol.*, **34**, 471–493.
- Brasch, J., Goodman, K.M., Noble, A.J., Rapp, M., Mannepalli, S., Bahna, F., Dandey, V.P., Bepler, T., Berger, B., Maniatis, T. *et al.* (2019)

- Visualization of clustered protocadherin neuronal self-recognition complexes. *Nature*, **569**, 280–283.
25. Wu, Q. and Jia, Z. (2021) Wiring the brain by clustered protocadherin neural codes. *Neurosci. Bull.*, **37**, 117–131.
  26. Tasic, B., Nabholz, C.E., Baldwin, K.K., Kim, Y., Rueckert, E.H., Ribich, S.A., Cramer, P., Wu, Q., Axel, R. and Maniatis, T. (2002) Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing. *Mol. Cell*, **10**, 21–33.
  27. Wang, X., Su, H. and Bradley, A. (2002) Molecular mechanisms governing Pcdh-gamma gene expression: evidence for a multiple promoter and cis-alternative splicing model. *Genes Dev.*, **16**, 1890–1905.
  28. Ribich, S., Tasic, B. and Maniatis, T. (2006) Identification of long-range regulatory elements in the protocadherin-alpha gene cluster. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 19719–19724.
  29. Kehayova, P., Monahan, K., Chen, W. and Maniatis, T. (2011) Regulatory elements required for the activation and repression of the protocadherin-alpha gene cluster. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 17195–17200.
  30. Yokota, S., Hirayama, T., Hirano, K., Kaneko, R., Toyoda, S., Kawamura, Y., Hirabayashi, M., Hirabayashi, T. and Yagi, T. (2011) Identification of the cluster control region for the protocadherin-beta genes located beyond the protocadherin-gamma cluster. *J. Biol. Chem.*, **286**, 31885–31895.
  31. Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y. *et al.* (2015) CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell*, **162**, 900–910.
  32. Lu, Y.J., Shou, J., Jia, Z.L., Wu, Y.H., Li, J.H., Guo, Y. and Wu, Q. (2019) Genetic evidence for asymmetric blocking of higher-order chromatin structure by CTCF/cohesin. *Protein Cell*, **10**, 914–920.
  33. Jia, Z., Li, J., Ge, X., Wu, Y., Guo, Y. and Wu, Q. (2020) Tandem CTCF sites function as insulators to balance spatial chromatin contacts and topological enhancer-promoter selection. *Genome Biol.*, **21**, 75.
  34. Golan-Mashiach, M., Grunspan, M., Emmanuel, R., Gibbs-Bar, L., Dikstein, R. and Shapiro, E. (2011) Identification of CTCF as a master regulator of the clustered protocadherin genes. *Nucleic Acids Res.*, **40**, 3378–3391.
  35. Hirayama, T., Tarusawa, E., Yoshimura, Y., Galjart, N. and Yagi, T. (2012) CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. *Cell Rep.*, **2**, 345–357.
  36. Guo, Y., Monahan, K., Wu, H., Gertz, J., Varley, K.E., Li, W., Myers, R.M., Maniatis, T. and Wu, Q. (2012) CTCF/cohesin-mediated DNA looping is required for protocadherin alpha promoter choice. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 21081–21086.
  37. Monahan, K., Rudnick, N.D., Kehayova, P.D., Pauli, F., Newberry, K.M., Myers, R.M. and Maniatis, T. (2012) Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of protocadherin-alpha gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 9125–9130.
  38. Canzio, D., Nwakeze, C.L., Horta, A., Rajkumar, S.M., Coffey, E.L., Duffy, E.E., Duffie, R., Monahan, K., O’Keeffe, S., Simon, M.D. *et al.* (2019) Antisense lncRNA transcription mediates DNA demethylation to drive stochastic protocadherin alpha promoter choice. *Cell*, **177**, 639–653.
  39. Tan, Y.P., Li, S., Jiang, X.J., Loh, W., Foo, Y.K., Loh, C.B., Xu, Q., Yuen, W.H., Jones, M., Fu, J. *et al.* (2010) Regulation of protocadherin gene expression by multiple neuron-restrictive silencer elements scattered in the gene cluster. *Nucleic Acids Res.*, **38**, 4985–4997.
  40. He, Q., Johnston, J. and Zeitlinger, J. (2015) ChIP-nexus enables improved detection of *in vivo* transcription factor binding footprints. *Nat. Biotechnol.*, **33**, 395–401.
  41. Li, J., Shou, J., Guo, Y., Tang, Y., Wu, Y., Jia, Z., Zhai, Y., Chen, Z., Xu, Q. and Wu, Q. (2015) Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. *J. Mol. Cell. Biol.*, **7**, 284–298.
  42. Shou, J., Li, J., Liu, Y. and Wu, Q. (2018) Precise and predictable CRISPR chromosomal rearrangements reveal principles of Cas9-mediated nucleotide insertion. *Mol. Cell*, **71**, 498–509.
  43. Da, L.T., Shi, Y., Ning, G. and Yu, J. (2018) Dynamics of the excised base release in thymine DNA glycosylase during DNA repair process. *Nucleic Acids Res.*, **46**, 568–581.
  44. Yin, M., Wang, J., Wang, M., Li, X., Zhang, M., Wu, Q. and Wang, Y. (2017) Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. *Cell Res.*, **27**, 1365–1377.
  45. Hashimoto, H., Wang, D., Horton, J.R., Zhang, X., Corces, V.G. and Cheng, X. (2017) Structural basis for the versatile and methylation-dependent binding of CTCF to DNA. *Mol. Cell*, **66**, 711–720.
  46. Kumaki, Y., Oda, M. and Okano, M. (2008) QUMA: quantification tool for methylation analysis. *Nucleic Acids Res.*, **36**, W170–W175.
  47. Thu, C.A., Chen, W.V., Rubinstein, R., Cheve, M., Wolcott, H.N., Felsevalyi, K.O., Tapia, J.C., Shapiro, L., Honig, B. and Maniatis, T. (2014) Single-cell identity generated by combinatorial homophilic interactions between alpha, beta, and gamma protocadherins. *Cell*, **158**, 1045–1059.
  48. Garrett, A.M., Bosch, P.J., Steffen, D.M., Fuller, L.C., Marcucci, C.G., Koch, A.A., Bais, P., Weiner, J.A. and Burgess, R.W. (2019) CRISPR/Cas9 interrogation of the mouse Pcdhg gene cluster reveals a crucial isoform-specific role for Pcdhgc4. *PLoS Genet.*, **15**, e1008554.
  49. GTEx Consortium (2013) The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.
  50. Shimajo, M., Lee, J.H. and Hersh, L.B. (2001) Role of zinc finger domains of the transcription factor neuron-restrictive silencer factor/repressor element-1 silencing transcription factor in DNA binding and nuclear localization. *J. Biol. Chem.*, **276**, 13121–13126.
  51. Zhai, Y.N., Xu, Q., Guo, Y. and Wu, Q. (2016) Characterization of a cluster of CTCF-binding sites in a protocadherin regulatory region. *Yi Chuan*, **38**, 323–336.
  52. Jia, Z., Guo, Y., Tang, Y., Xu, Q., Li, B. and Wu, Q. (2014) Regulation of the protocadherin Celsr3 gene and its role in *globus pallidus* development and connectivity. *Mol. Cell. Biol.*, **34**, 3895–3910.
  53. Martin, B.J.E., Brind’Amour, J., Kuzmin, A., Jensen, K.N., Liu, Z.C., Lorincz, M. and Howe, L.J. (2021) Transcription shapes genome-wide histone acetylation patterns. *Nat. Commun.*, **12**, 210.
  54. Jia, Z. and Wu, Q. (2020) Clustered protocadherins emerge as novel susceptibility loci for mental disorders. *Front. Neurosci.*, **14**, 587819.
  55. Esumi, S., Kakazu, N., Taguchi, Y., Hirayama, T., Sasaki, A., Hirabayashi, T., Koide, T., Kitsukawa, T., Hamada, S. and Yagi, T. (2005) Monoallelic yet combinatorial expression of variable exons of the protocadherin-alpha gene cluster in single neurons. *Nat. Genet.*, **37**, 171–176.
  56. Kaneko, R., Kato, H., Kawamura, Y., Esumi, S., Hirayama, T., Hirabayashi, T. and Yagi, T. (2006) Allelic gene regulation of Pcdh-alpha and Pcdh-gamma clusters involving both monoallelic and biallelic expression in single Purkinje cells. *J. Biol. Chem.*, **281**, 30551–30560.
  57. Schreiner, D. and Weiner, J.A. (2010) Combinatorial homophilic interaction between gamma-protocadherin multimers greatly expands the molecular diversity of cell adhesion. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 14893–14898.
  58. Hirano, K., Kaneko, R., Izawa, T., Kawaguchi, M., Kitsukawa, T. and Yagi, T. (2012) Single-neuron diversity generated by Protocadherin-beta cluster in mouse central and peripheral nervous systems. *Front. Mol. Neurosci.*, **5**, 90.
  59. Schoenherr, C.J., Paquette, A.J. and Anderson, D.J. (1996) Identification of potential target genes for the neuron-restrictive silencer factor. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 9881–9886.
  60. Zuccato, C., Tartari, M., Crotti, A., Goffredo, D., Valenza, M., Conti, L., Cataudella, T., Leavitt, B.R., Hayden, M.R., Timmusk, T. *et al.* (2003) Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. *Nat. Genet.*, **35**, 76–83.
  61. Zhang, C., Xuan, Z., Otto, S., Hover, J.R., McCorkle, S.R., Mandel, G. and Zhang, M.Q. (2006) A clustering property of highly-degenerate transcription factor binding sites in the mammalian genome. *Nucleic Acids Res.*, **34**, 2238–2246.
  62. Bruce, A.W., Lopez-Contreras, A.J., Flicek, P., Down, T.A., Dhimi, P., Dillon, S.C., Koch, C.M., Langford, C.F., Dunham, I., Andrews, R.M. *et al.* (2009) Functional diversity for REST (NRSF) is defined by *in vivo* binding affinity hierarchies at the DNA sequence level. *Genome Res.*, **19**, 994–1005.