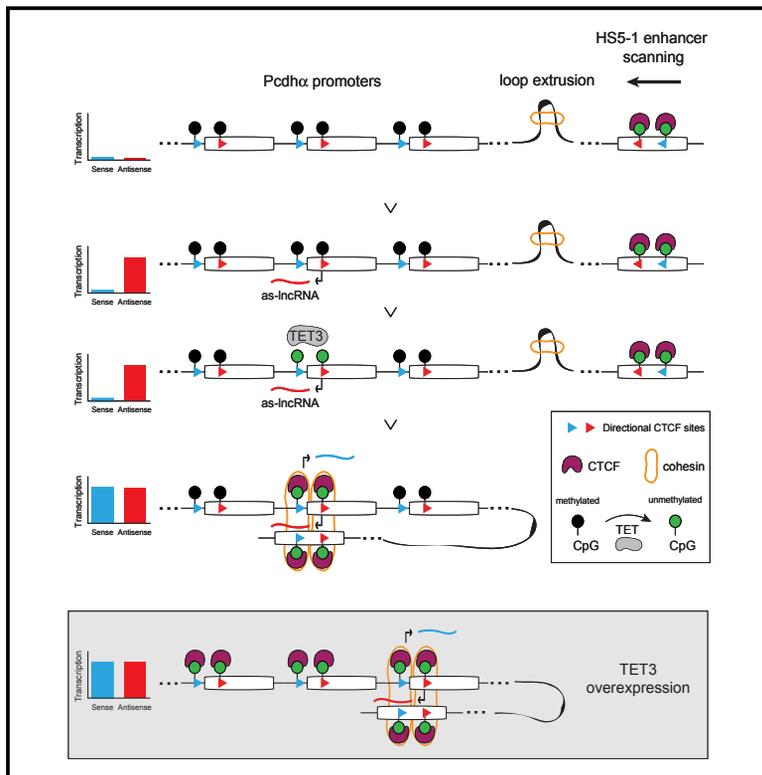


# Antisense lncRNA Transcription Mediates DNA Demethylation to Drive Stochastic Protocadherin $\alpha$ Promoter Choice

## Graphical Abstract



## Authors

Daniele Canzio, Chiamaka L. Nwakeze, Adan Horta, ..., Matthew D. Simon, Stavros Lomvardas, Tom Maniatis

## Correspondence

tm2472@cumc.columbia.edu

## In Brief

Coupling transcription of a long noncoding RNA to DNA demethylation ensures stochastic promoter choice for clustered Protocadherin  $\alpha$  genes, which is essential for the establishment of a neuronal surface identity code involved in circuit assembly.

## Highlights

- A conserved antisense promoter is located within each of the *Pcdh $\alpha$*  alternate exons
- Antisense lncRNA transcription leads to DNA demethylation of promoters and CBSs
- CTCF/cohesin drive the assembly of *Pcdh $\alpha$*  enhancer/promoter complex via loop extrusion
- Coupling lncRNA transcription to DNA demethylation ensures stochastic promoter choice



# Antisense lncRNA Transcription Mediates DNA Demethylation to Drive Stochastic Protocadherin $\alpha$ Promoter Choice

Daniele Canzio,<sup>1,2,6,7</sup> Chiamaka L. Nwakeze,<sup>1,2,6</sup> Adan Horta,<sup>1,2,6</sup> Sandy M. Rajkumar,<sup>1,2</sup> Eliot L. Coffey,<sup>3</sup> Erin E. Duffy,<sup>4</sup> Rachel Duffié,<sup>1,2</sup> Kevin Monahan,<sup>1,2</sup> Sean O’Keeffe,<sup>1,8</sup> Matthew D. Simon,<sup>4</sup> Stavros Lomvardas,<sup>1,2</sup> and Tom Maniatis<sup>1,2,5,9,\*</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA

<sup>2</sup>Mortimer B. Zuckerman Mind Brain and Behavior Institute, Columbia University, New York, NY 10027, USA

<sup>3</sup>Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA

<sup>4</sup>Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06516, USA

<sup>5</sup>New York Genome Center, New York, NY 10013, USA

<sup>6</sup>These authors contributed equally

<sup>7</sup>Present address: Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA 94158, USA

<sup>8</sup>Present address: Celmatix, New York, NY 10005, USA

<sup>9</sup>Lead Contact

\*Correspondence: [tm2472@cumc.columbia.edu](mailto:tm2472@cumc.columbia.edu)

<https://doi.org/10.1016/j.cell.2019.03.008>

## SUMMARY

Stochastic activation of clustered Protocadherin (Pcdh)  $\alpha$ ,  $\beta$ , and  $\gamma$  genes generates a cell-surface identity code in individual neurons that functions in neural circuit assembly. Here, we show that Pcdh $\alpha$  gene choice involves the activation of an antisense promoter located in the first exon of each Pcdh $\alpha$  alternate gene. Transcription of an antisense long noncoding RNA (lncRNA) from this antisense promoter extends through the sense promoter, leading to DNA demethylation of the CTCF binding sites proximal to each promoter. Demethylation-dependent CTCF binding to both promoters facilitates cohesin-mediated DNA looping with a distal enhancer (HS5-1), locking in the transcriptional state of the chosen Pcdh $\alpha$  gene. Uncoupling DNA demethylation from antisense transcription by Tet3 overexpression in mouse olfactory neurons promotes CTCF binding to all Pcdh $\alpha$  promoters, resulting in proximity-biased DNA looping of the HS5-1 enhancer. Thus, antisense transcription-mediated promoter demethylation functions as a mechanism for distance-independent enhancer/promoter DNA looping to ensure stochastic Pcdh $\alpha$  promoter choice.

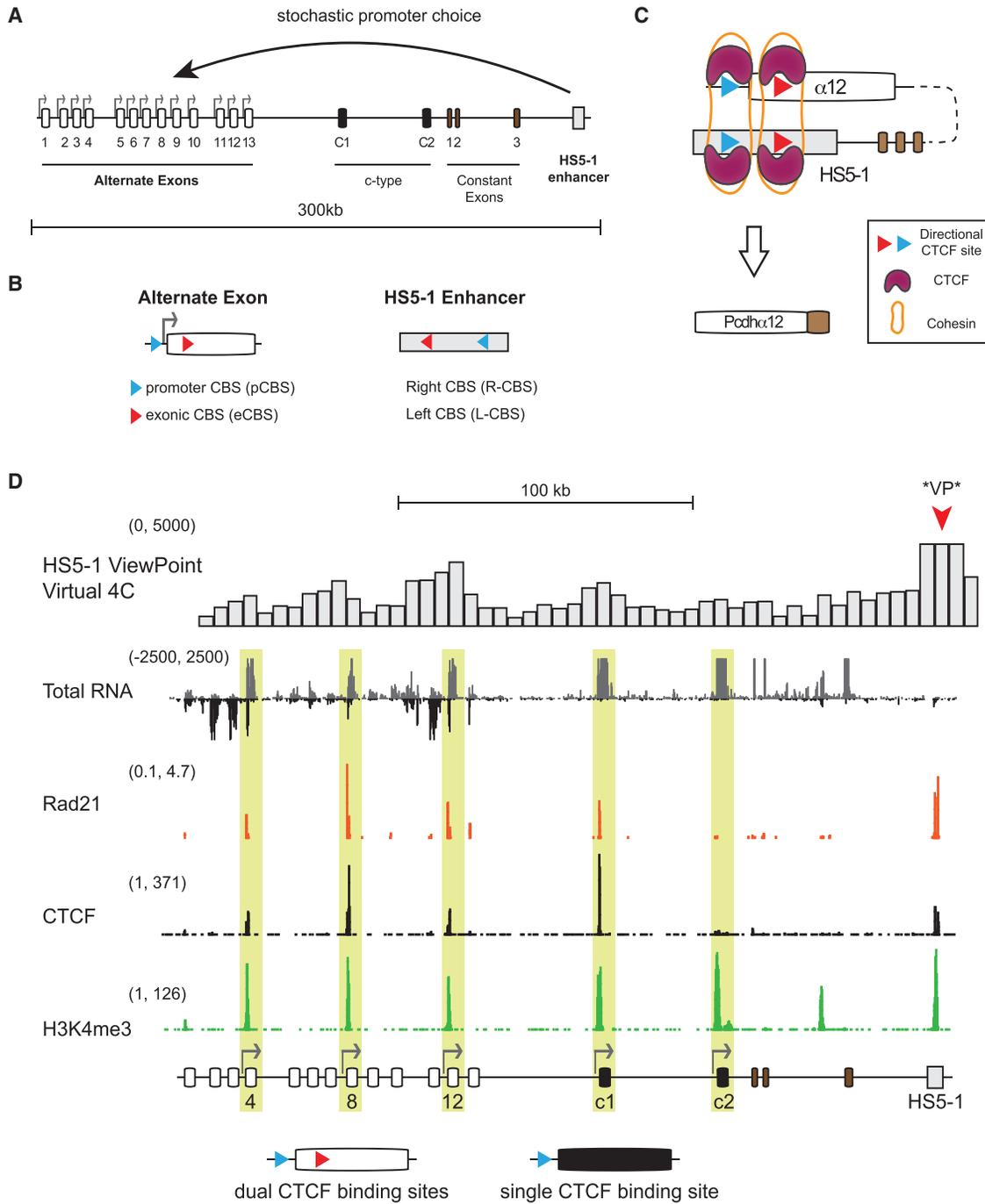
## INTRODUCTION

During brain development, individual neurons differentiate into distinct functional cell types, respond to a plethora of guidance molecules, and project into specific regions of the nervous system to form complex neural circuits. A key aspect of this process is the ability of neurites of individual neurons (axons and dendrites) to distinguish between themselves and neurites from

other neurons (self versus non-self; Zipursky and Grueber, 2013). This process, known as self-avoidance, requires a unique combination of cell-surface homophilic recognition molecules that function as a molecular identity code (Zipursky and Grueber, 2013; Zipursky and Sanes, 2010). In mammals, this identity code is generated by random transcription of clustered Protocadherin (Pcdh) genes (Lefebvre et al., 2015; Mountoufaris et al., 2018) by means of a poorly understood mechanism of stochastic and combinatorial promoter choice (Esumi et al., 2005; Tasic et al., 2002; Wang et al., 2002; Wu and Maniatis, 1999). Pcdh genes have a unique genomic arrangement consisting of three closely linked gene clusters ( $\alpha$ ,  $\beta$ , and  $\gamma$ ) that, together, span nearly 1 million bp of genomic DNA. The  $\alpha$  and  $\gamma$  clusters are organized into variable (alternate and c-type) and constant regions, reminiscent of the organization of immunoglobulin and T cell receptor gene clusters (Wu and Maniatis, 1999; Figure 1A, Pcdh $\alpha$ ).

Neuron-specific expression of individual Pcdh $\alpha$  genes requires long-range DNA looping between individual Pcdh $\alpha$  promoters and a transcriptional enhancer, called HS5-1 (hypersensitivity site 5-1) (Guo et al., 2012, 2015; Kehayova et al., 2011; Monahan et al., 2012; Ribich et al., 2006; Figure 1A). Conserved transcriptional promoter sequences are located immediately proximal to every Pcdh $\alpha$  exon (Tasic et al., 2002), and the HS5-1 enhancer is located downstream of the constant exons, between the Pcdh  $\alpha$  and the  $\beta$  clusters (Ribich et al., 2006; Figures 1A, 1B, and S1). These stochastic enhancer/promoter interactions occur independently on each of the two allelic chromosomes in diploid cells and require the binding of the CCCTC-binding protein (CTCF) and the cohesin protein complex (Guo et al., 2012; Hirayama et al., 2012; Kehayova et al., 2011; Monahan et al., 2012; Figure 1C). CTCF is an 11 zinc finger (ZF) domain protein that, together with the cohesin complex, plays a central role as an insulator of chromatin domains and mediates genome-wide enhancer/promoter interactions (Ghirlando and Felsenfeld, 2016; Ong and Corces, 2014). All Pcdh $\alpha$  alternate exons contain two CTCF binding sites (CBSs): one in the promoter (pCBS) and one in the protein coding sequence of





**Figure 1. Transcription of Sense and Antisense RNA from *Pcdhα* Alternative Exons**

(A) Genomic organization of the human *Pcdhα* gene cluster. The stochastic expression of 13 alternate exons is driven by with their own promoter (arrows) and is equally likely to be activated. c1 and c2, c-type exons; 1–3, constant exons encoding the intracellular domain of Pcdh proteins.

(B) Location and relative orientations (arrows) of the promoter and the exonic CBSs (alternate exons) and the left and right CBSs (HS5-1 enhancer).

(C) Example of a *Pcdhα* promoter/HS5-1 enhancer complex bound to CTCF and cohesin.

(D) Sense (gray) and antisense (black) RNA (total RNA and cRNA-seq) at the *Pcdhα* cluster in SK-N-SH cells. CTCF, Rad21, and H3K4me3 (ChIP-seq) relative to active exons (yellow) is shown. *Pcdhα*c2, active, but not bound by CTCF or Rad21. Virtual 4C (cHi-C) on top (HS5-1 as a viewpoint) is shown. x axis, linear sequence of the genomic organization of the *Pcdhα* cluster. Numbers on the left-hand side of each track are minimum and maximum densities in reads per million. See also Figure S1.

the first exon (eCBS; Guo et al., 2012; Monahan et al., 2012; Figure 1B). The two binding sites are separated by approximately 1,000 bp, and similarly spaced CBSs are located in the HS5-1 enhancer (L-CBS and R-CBS; Guo et al., 2012; Monahan et al., 2012; Figure 1B). Interestingly, the CBSs in Pcdh $\alpha$  promoters and the HS5-1 enhancer are in opposite relative orientations, and inversion of the HS5-1 enhancer results in a significant decrease in Pcdh $\alpha$  gene cluster expression, demonstrating the functional importance of this arrangement (Guo et al., 2015). This opposite relative orientation of promoter and enhancer CBSs appears to be a general feature of eukaryotic chromosomes genome-wide (Guo et al., 2015; Rao et al., 2014) and has been proposed to play a critical role in promoting the spatial interaction between genes and transcriptional regulatory elements by a mechanism known as loop extrusion (Fudenberg et al., 2016). In the context of the Pcdh $\alpha$  gene cluster, the loop-extrusion model predicts that the HS5-1 enhancer, bound by CTCF and the cohesin proteins, scans the Pcdh $\alpha$  exons until it finds the exon bound by CTCF. However, this possibility has yet to be demonstrated.

A critical insight into the formation of Pcdh $\alpha$  enhancer/promoter complexes was provided by the observation that there is an inverse relationship between Pcdh $\alpha$  gene expression and DNA methylation of the Pcdh $\alpha$  promoters (Tasic et al., 2002; Toyoda et al., 2014). Specifically, the CTCF/cohesin complex associates exclusively with transcriptionally active promoters, which are characterized by hypomethylation of the CBSs and of the DNA sequences located between the two CBSs (Guo et al., 2012). By contrast, CBSs and the DNA between them are hypermethylated in inactive promoters, thus preventing CTCF/cohesin binding (Guo et al., 2012). Although DNA methylation of the CBSs is likely to play an important role in the mechanism of stochastic Pcdh $\alpha$  promoter choice, the temporal relationship between promoter DNA methylation and promoter choice is not known. For example, it is not known whether promoter DNA methylation is the ground state upon which promoter choice operates or whether all promoters are initially unmethylated and methylation of the inactive promoters occurs subsequent to stochastic promoter choice.

Here, we provide evidence that the ground state of Pcdh $\alpha$  promoter DNA is methylated and transcriptionally repressed in immature cells destined to become olfactory sensory neurons. Stochastic promoter demethylation occurs by a remarkable mechanism in which transcription of an antisense long noncoding RNA (lncRNA) is initiated from a promoter located within the downstream protein coding region of each Pcdh $\alpha$  exon. Transcription through the upstream sense promoter results in its demethylation, binding of CTCF, and DNA looping to the HS5-1 enhancer. The binding of CTCF marks the promoter for engagement by the HS5-1 enhancer through DNA loop extrusion (Fudenberg et al., 2016), thus eliminating enhancer/promoter proximity bias.

## RESULTS

### Transcription of Sense and Antisense RNA from Clustered Pcdh $\alpha$ Alternate Exons

The mechanism of stochastic promoter choice in the Pcdh $\alpha$  gene cluster cannot be studied *in vivo*, as each neuron expresses

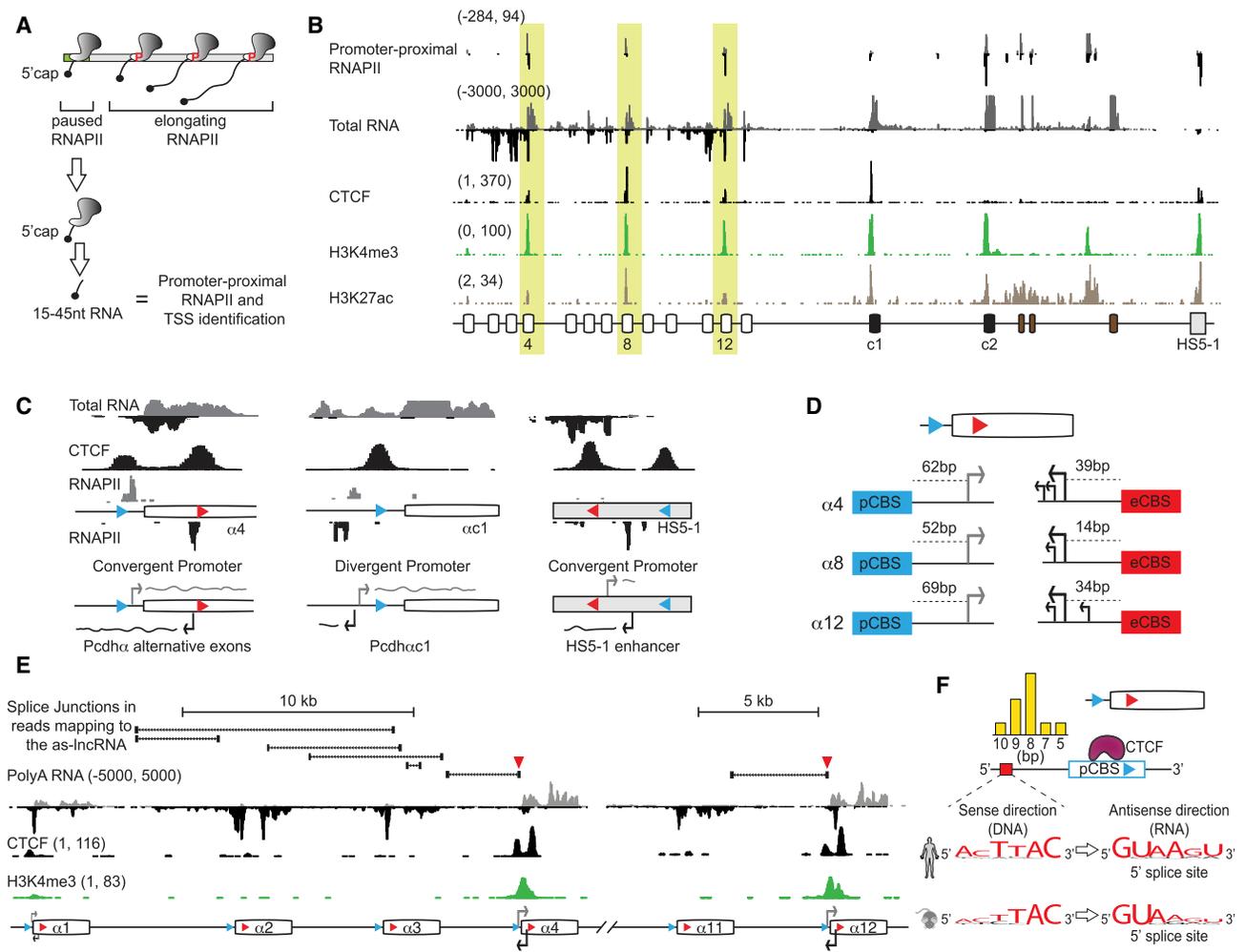
a distinct repertoire of Pcdh $\alpha$  alternate exons. We therefore made use of the well-characterized human neuroblastoma cell line SK-N-SH, which stably expresses a distinct repertoire of Pcdh $\alpha$  isoforms through multiple cell divisions:  $\alpha$ 4;  $\alpha$ 8;  $\alpha$ 12;  $\alpha$ c1; and  $\alpha$ c2 (Guo et al., 2012; Figure 1D). This stochastic pattern of expression in cell culture is indistinguishable from that observed in single neurons *in vivo* (Esumi et al., 2005; Moun-toufaris et al., 2017). SK-N-SH cells thus provide a multicellular “avatar” for studying single cell expression of Pcdh $\alpha$  genes and internal controls for exons that are transcriptionally silent.

The low level of expression of Pcdh genes provides an additional challenge to the study of Pcdh $\alpha$  promoter choice. To optimize the analysis of Pcdh RNA precursors (pre-mRNA) and messenger (mRNA) RNAs in SK-N-SH cells, we employed capture RNA sequencing (cRNA-seq) and achieved a two order of magnitude enrichment of Pcdh RNA transcripts (Figure S1). Remarkably, this enrichment revealed a high level of antisense RNA transcription of the Pcdh $\alpha$  alternate exons, which contain dual CBSs in SK-N-SH cells (Figures 1D and S1B). By contrast, antisense RNA transcription was not detected within the two c-type exons,  $\alpha$ c1 and  $\alpha$ c2, which do not contain CBSs within their exons (Figure 1D). Antisense RNA was not observed in the Pcdh  $\beta$  or  $\gamma$  variable exons in SK-N-SH cells, which, like  $\alpha$ c1 and  $\alpha$ c2, do not contain exonic CBSs (Figure S1B). We refer to the observed antisense RNA as as-lncRNA, as this high-molecular-weight RNA lacks open reading frames that encode protein. For clarity, we refer to the sense Pcdh coding RNA as s-cRNA (sense coding RNA).

### Convergent Promoters in Both the Pcdh $\alpha$ Alternative Exons and HS5-1 Enhancer

In order to characterize the nature of the antisense RNAs and to gain mechanistic insights into their function, we first localized their transcription start sites and the location of the promoter-paused RNA polymerase II (RNAPII) using Start-Seq (Nechaev et al., 2010). RNAs isolated from stalled RNAPII at promoters are approximately 15–45 nt long and contain a 5' 7mG-cap (Figure 2A). Sequencing of these short RNAs revealed the position of paused RNAPII, thus acting as a proxy for the location of RNAPII-engaged promoters and the transcriptional start site at a nucleotide-base resolution (Figure S2A). As expected, we observed promoter-proximal RNAPII at the pCBS-proximal promoter of the active Pcdh  $\alpha$ 4,  $\alpha$ 8,  $\alpha$ 12, and  $\alpha$ c1 exons and at the promoter of  $\alpha$ c2 in SK-N-SH cells (Figure 2B). To our surprise, however, we also observed promoter-proximal RNAPII just upstream of the eCBS for  $\alpha$ 4,  $\alpha$ 8, and  $\alpha$ 12 in the antisense orientation (Figure 2B). Thus, sequences near the two CBSs of active Pcdh $\alpha$  genes act as convergent promoters, where antisense and sense RNA converge and partially overlap (Figure 2C; Pcdh $\alpha$ 4 is shown). This is in contrast to the singular pCBS site in Pcdh $\alpha$ c1, which acts as a more canonical divergent promoter, where transcription of the antisense and sense RNA occurs in opposite directions and does not overlap (Figure 2C). Remarkably, Start-Seq analysis also identified a similar convergent promoter architecture of the two CBSs in the HS5-1 enhancer (Figures 2B and 2C). The position of transcriptional start site (TSS) for Pcdh  $\alpha$ 4,  $\alpha$ 8, and  $\alpha$ 12 are shown in Figure 2D.

Mapping the location of the Pcdh $\alpha$  as-lncRNA promoters with respect to the as-lncRNAs revealed that these nuclear RNA



**Figure 2. Convergent *Pcdhα* Alternative Exon Promoters and the HS5-1 Enhancer**

(A) Schematic diagram of Start-Seq.

(B) Paused RNAPII (Start-Seq) relative to total RNA (cRNA-seq), CTCF, H3K4me3, and H3K27ac (ChIP-seq) in SK-N-SH cells.

(C) Promoter architectures for *Pcdhα4* (convergent), *Pcdhα1* (divergent), and the HS5-1 enhancer (convergent).

(D) TSS of the as-lncRNA and s-cRNA from *Pcdh α4*,  $\alpha 8$ , and  $\alpha 12$ .

(E) RNA splicing patterns of polyadenylated as-lncRNAs (cRNA-seq) initiated from the *Pcdhα4* and *Pcdhα12* as indicated by the splice junctions in reads mapping to the as-lncRNAs, relative to CTCF and H3K4me3. Red triangles, antisense 5'ss described in (F).

(F) Position, sequence, and conservation of the antisense 5'ss located upstream of the pCBS (blue). CTCF is in violet. Bar graph, distribution of the distance of the 5'ss from the pCBS.

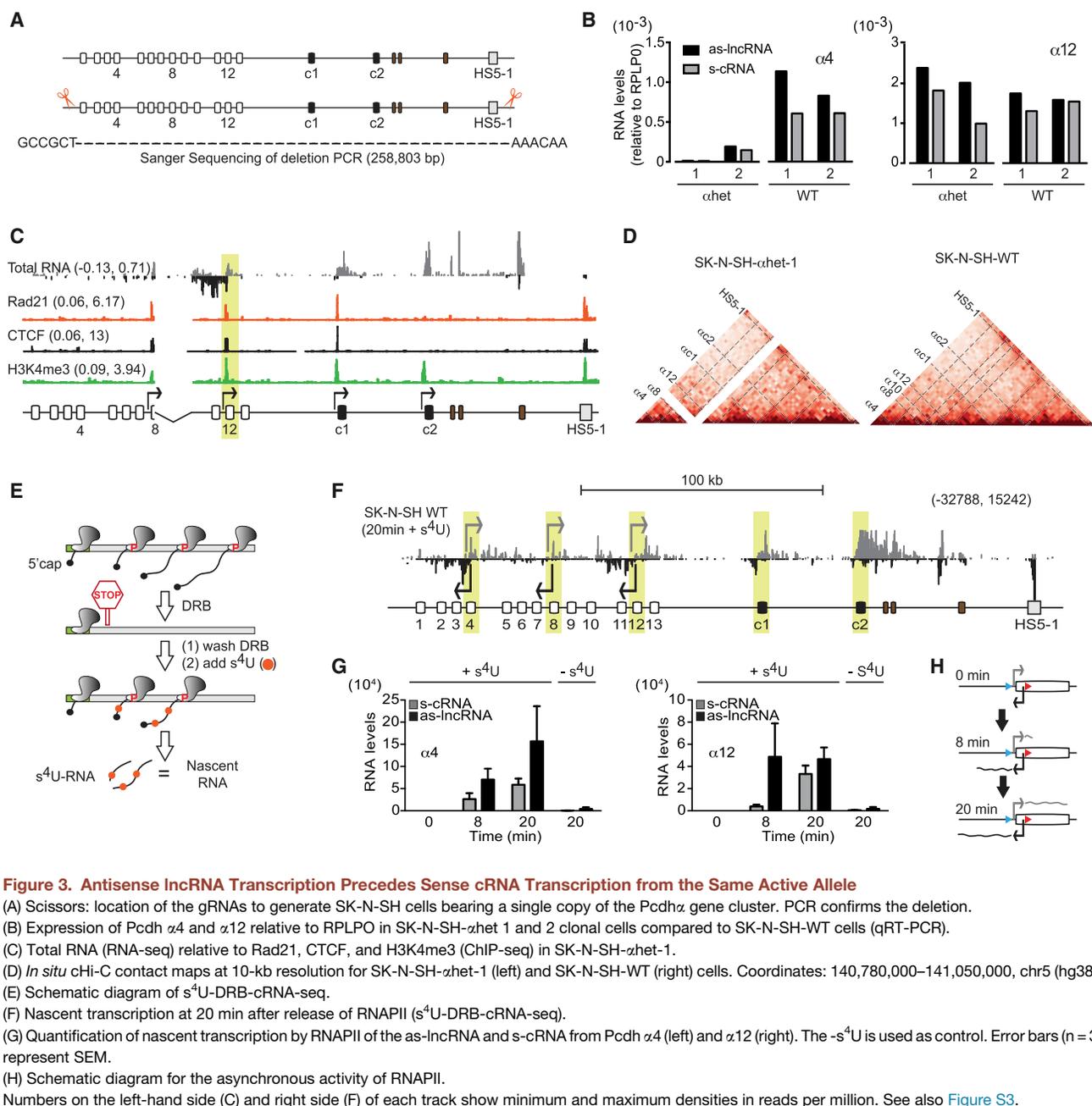
For (B) and (E), numbers on the left-hand side of each track show minimum and maximum densities in reads per million. x axis, linear sequence of the genomic organization of the *Pcdhα* cluster. Arrows in (E) show position of transcription start sites by Start-Seq. See also Figure S2.

precursors can be as long as 20 kb in length and are spliced and polyadenylated with half-lives of the same order of magnitude as their respective s-cRNAs (Figure S1F). As an example, the as-lncRNA that initiates at the eCBS-proximal promoter of *Pcdhα4* in SK-N-SH cells is transcribed through the pCBS-proximal promoter of *Pcdhα4* and extends in the antisense direction all the way to the intergenic sequence between the *Pcdh α1* and  $\alpha 2$  exons (more than 20 kb; Figure 2E). By contrast, the antisense RNA that initiates at the eCBS promoter of *Pcdhα12* extends to the *Pcdhα11* exon (Figure 2E). In addition, we discovered the presence of a highly conserved 5' splice site (5'ss), encoded in the antisense direction about 7 bp upstream of the pCBS core

motif (Figure 2F). Usage of that 5'ss results in the most abundant polyadenylated as-lncRNA spliced isoform (Figure 2E). Remarkably, this site is absent from the pCBS of *Pcdhα1*, as well as from the pCBS sites of the *Pcdh β* and  $\gamma$  clusters. These observations suggest that RNA splicing of this promoter-embedded 5' splice site may be coupled to the activation of the pCBS promoter (see Discussion).

#### Antisense lncRNA and Sense-Coding RNA Are Transcribed from the Same Active Allele

The cRNA-seq data obtained from SK-N-SH cells revealed a direct correlation between sense and antisense RNA transcription



**Figure 3. Antisense lncRNA Transcription Precedes Sense cRNA Transcription from the Same Active Allele**

(A) Scissors: location of the gRNAs to generate SK-N-SH cells bearing a single copy of the *Pcdh $\alpha$*  gene cluster. PCR confirms the deletion.  
 (B) Expression of *Pcdh $\alpha 4$*  and  $\alpha 12$  relative to RPLP0 in SK-N-SH- $\alpha$ het 1 and 2 clonal cells compared to SK-N-SH-WT cells (qRT-PCR).  
 (C) Total RNA (RNA-seq) relative to Rad21, CTCF, and H3K4me3 (ChIP-seq) in SK-N-SH- $\alpha$ het-1.  
 (D) *In situ* cHi-C contact maps at 10-kb resolution for SK-N-SH- $\alpha$ het-1 (left) and SK-N-SH-WT (right) cells. Coordinates: 140,780,000–141,050,000, chr5 (hg38).  
 (E) Schematic diagram of  $s^4$ U-DRB-cRNA-seq.  
 (F) Nascent transcription at 20 min after release of RNAPII ( $s^4$ U-DRB-cRNA-seq).  
 (G) Quantification of nascent transcription by RNAPII of the as-lncRNA and s-cRNA from *Pcdh $\alpha 4$*  (left) and  $\alpha 12$  (right). The - $s^4$ U is used as control. Error bars (n = 3) represent SEM.  
 (H) Schematic diagram for the asynchronous activity of RNAPII.  
 Numbers on the left-hand side (C) and right side (F) of each track show minimum and maximum densities in reads per million. See also Figure S3.

and transcriptionally active *Pcdh $\alpha$*  alternate exons. Because transcription of the *Pcdh $\alpha$*  alternate exons occurs independently on the two allelic chromosomes (Esumi et al., 2005), we sought to determine whether the as-lncRNA and the s-cRNA were transcribed from the same *Pcdh $\alpha$*  locus allele. To accomplish this, we used CRISPR-Cas9 gene editing to generate SK-N-SH cells heterozygous for the *Pcdh $\alpha$*  gene cluster, SK-N-SH- $\alpha$ het (Figure 3A). We isolated two clones (SK-N-SH- $\alpha$ het 1 and 2) expressing primarily  $\alpha 12$ ,  $\alpha c1$ , and  $\alpha c2$  from the remaining copy of the *Pcdh $\alpha$*  gene cluster (Figures 3B and 3C). Both clones showed expression of the as-lncRNA and s-cRNA from *Pcdh $\alpha 12$*  (Figures 3B and 3C), confirming that sense and antisense transcription

originate from the same allele. Furthermore, chromatin immunoprecipitation sequencing studies (ChIP-seq) for CTCF and Rad21, a subunit of the cohesin complex, as well as capture *in situ* high-throughput chromosome conformation capture studies (cHi-C) performed in  $\alpha$ het-1 also demonstrated that the *Pcdh $\alpha$*  alternate exons, from which sense and antisense RNAs are transcribed, are bound by CTCF and cohesin and engaged in enhancer interactions between alternate promoters and the HS5-1 enhancer DNA looping (Figures 3C and 3D). We note that the  $\alpha$ het-1 and  $\alpha$ het-2 clones share a 16.7-kb deletion that truncates the *Pcdh $\alpha 8$*  exon and removes the *Pcdh $\alpha 9$*  and  $\alpha 10$  exons (Figures 3C and 3D). This deletion was previously reported as a

common feature of individuals from multiple populations of European and East Asian descent with no discernable phenotypic consequence (Noonan et al., 2003).

In contrast to SK-N-SH cells, a mixed population of primary neurons, each expressing a distinct repertoire of *Pcdh $\alpha$*  alternative exons, should collectively express as-lncRNAs from all the *Pcdh $\alpha$*  alternate exons, but not from *Pcdh $\alpha$*   $\alpha$ c1 and  $\alpha$ c2 or from the  $\beta$  or  $\gamma$  exons. As predicted, analysis of RNA from human primary neurons and from mouse mature olfactory sensory neurons (mOSNs) revealed lncRNA expression exclusively from all the *Pcdh $\alpha$*  alternate exons (Figures S2B and S2C). As in SK-N-SH cells, the as-lncRNAs expressed in human and mouse primary neurons are spliced and polyadenylated (Figures 2E, S2B, and S2C). However, in contrast to SK-N-SH cells, the levels of the as-lncRNAs in both human and mouse primary neurons appeared lower. We speculate that this difference could be a consequence of the mitotic (SK-N-SH) and the post-mitotic (primary neurons) state of the two cell types. We also note that an antisense lncRNA from the *Pcdh $\alpha$* 12 exon, similar to the one described and characterized above, was reported in human brain samples, but its significance was not understood (Lipovich et al., 2006).

### The Asymmetric Nature of *Pcdh $\alpha$* Convergent Promoters Results in Asynchronous Sense and Antisense RNA Transcription

Antisense convergent transcription is a widespread phenomenon in the mammalian genome. Yet its function, as well as the mechanism by which actively transcribing RNA polymerases translocate along a common stretch of DNA in opposite directions, remains unclear (see Discussion). To assess the activity of RNAPII at the pCBS-proximal and eCBS-proximal promoters, we analyzed transcription in SK-N-SH cells using s<sup>4</sup>UDRB-seq (Fuchs et al., 2014). This method combines synchronization of RNAPII at promoters by 5,6-dichloro-1- $\beta$ -D-ribofuranosylbenzimidazole (DRB) with incorporation of the nucleoside 4-thiouridine (s<sup>4</sup>U) during RNA synthesis (Figure 3E). Consistent with the Start-Seq data, we observed convergent elongating RNAPII from both pCBS- and eCBS-proximal promoters of  $\alpha$ 4,  $\alpha$ 8, and  $\alpha$ 12 and divergent RNAPII from the pCBS-proximal promoter of *Pcdh $\alpha$* c1 (Figure 3F). We also observed convergent elongating RNAPII at the HS5-1 enhancer, consistent with the presence of convergent promoters as described above (Figure 3F). These data reveal a remarkable symmetry between the location of CTCF/cohesin binding sites and sense and antisense transcription from the *Pcdh $\alpha$*  alternate promoters and the HS5-1 enhancer. However, in contrast to the sense and antisense RNA transcribed from *Pcdh $\alpha$*  alternate exons, both enhancer RNAs are not polyadenylated and therefore appear to rapidly turn over (Figures 1D, 1F, and S2B).

Interestingly, quantification of nascent transcription of the antisense and sense RNAs assayed by s<sup>4</sup>UDRB-seq revealed that, although RNAPII molecules at the *Pcdh $\alpha$*  active exons transcribe in a convergent manner, their activities appear asynchronous. That is, the as-lncRNA is transcribed earlier than the s-cRNA (Figures 3G and 3H). This asynchronous RNAPII activity reveals an intrinsic asymmetry in the activities of the two promoters, an observation consistent with the fact that the two promoters differ in their ability to bind to distinct classes of tran-

scription factors (TFs), as well as the fact that the two CBSs, proximal to the sense and antisense promoters, differ in sequence and in their affinity for CTCF (Figure S2D).

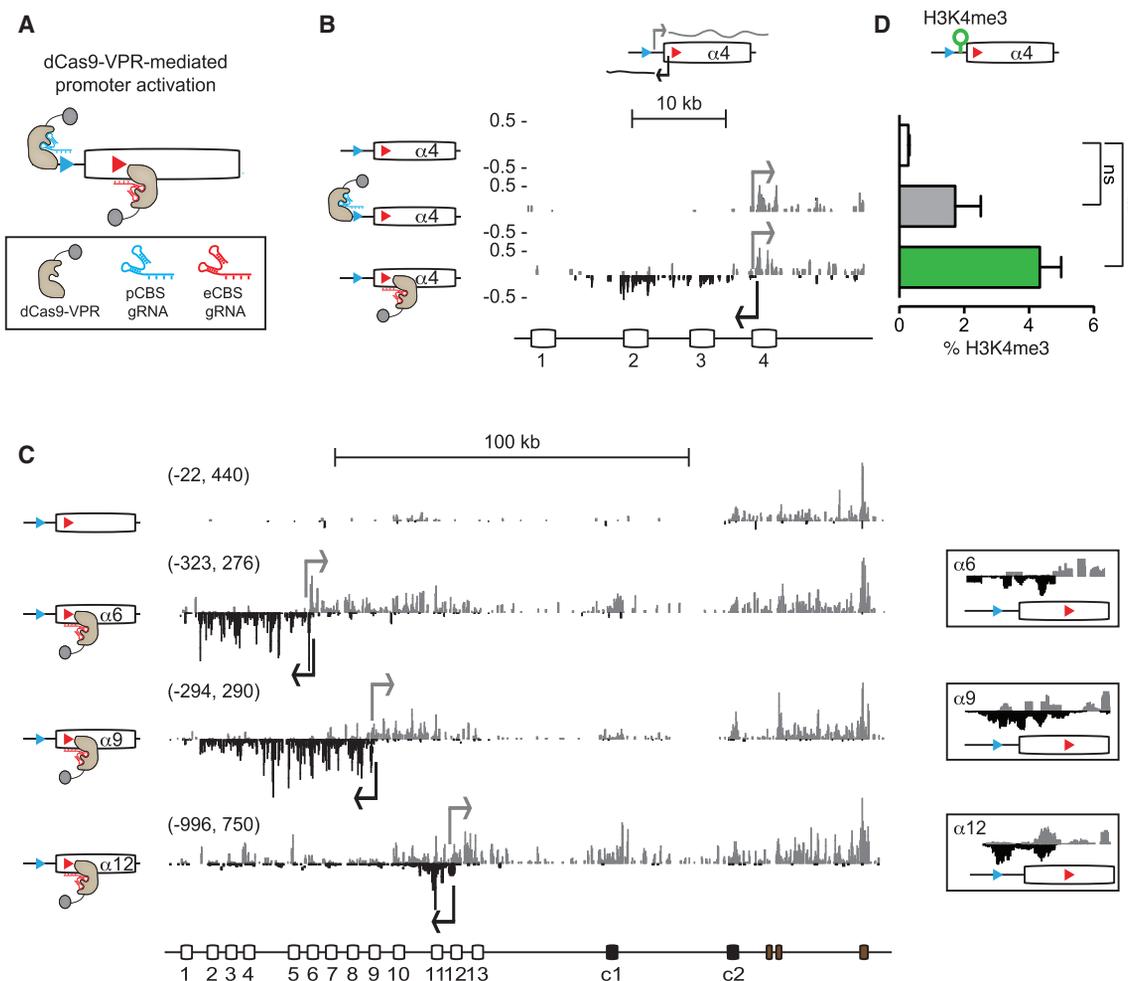
### Transcription of Antisense lncRNAs Triggers Activation of *Pcdh $\alpha$* Sense Promoters

To understand the functional significance of the asynchronous activity of the *Pcdh $\alpha$*  sense and antisense promoters, we designed a gain-of-function assay to uncouple transcription of the as-lncRNA from transcription of the sense-coding *Pcdh $\alpha$*  mRNA in the context of the endogenous *Pcdh $\alpha$*  gene cluster. Specifically, we made use of a catalytic-inactive CRISPR-dCas9 protein fused to a tripartite transcriptional activator (dCas9-VPR; Chavez et al., 2015) to selectively activate the pCBS-proximal or eCBS-proximal promoters of silent *Pcdh $\alpha$*  genes (Figure 4A). We chose HEK293T cells, as most *Pcdh $\alpha$*  genes are transcriptionally silent in this cell line, with the exception of *Pcdh $\alpha$*   $\alpha$ 10 and  $\alpha$ c2. This property of HEK293T cells, together with the modularity of the CRISPR-dCas9 system, made it possible to selectively design guide RNAs for the transcriptional activation of *Pcdh $\alpha$*   $\alpha$ 4,  $\alpha$ 6,  $\alpha$ 9, and  $\alpha$ 12 (Figure S3). As expected, dCas9-VPR activation of the *Pcdh $\alpha$* 4 sense promoter resulted in robust synthesis of the *Pcdh $\alpha$* 4 s-cRNA (Figure 4B). Unexpectedly, activation of the *Pcdh $\alpha$* 4 antisense promoter not only led to high levels of antisense RNA transcription, but high levels of sense RNA transcription were also observed (Figure 4B). This pattern of sense and antisense RNA transcription did not depend upon the number of dCas9-VPR activators (1 versus 4) or on their position relative to the CBSs (Figure S4A). Most importantly, this pattern of transcription mirrored that of active exons observed in SK-N-SH cells (Figure 1D). As for the *Pcdh $\alpha$* 4, we also observed the same relationship between as-lncRNA and sense transcription for the *Pcdh $\alpha$*   $\alpha$ 6,  $\alpha$ 9, and  $\alpha$ 12 exons (Figures 4C and S4B).

These observations suggest that transcription of antisense RNA from the eCBS-proximal promoter activates the upstream cognate pCBS-proximal promoter to generate sense-coding RNA. To test this possibility, we measured the levels of histone H3 lysine 4 trimethylation (H3K4me3), a histone post-translational modification that marks transcriptionally active promoters and is detected on DNA between the two CTCF-bound CBSs in active *Pcdh $\alpha$*  genes (Figure 1D). As shown in Figure 4D, chromatin immunoprecipitation followed by qPCR (ChIP-qPCR) resulted in an increase in H3K4me3 upon transcriptional activation of the antisense promoter by dCas9-VPR.

### Antisense lncRNA Transcription Promotes CTCF Binding and Long-Range Enhancer/Promoter DNA Interactions

The expression of *Pcdh $\alpha$*  sense RNA transcripts requires the binding of CTCF and cohesin to the pCBSs and eCBSs and long-range DNA looping between active promoters and the HS5-1 enhancer (Guo et al., 2012, 2015). In ChIP-seq experiments, we observed that both CBSs of *Pcdh $\alpha$*   $\alpha$ 4,  $\alpha$ 6,  $\alpha$ 9, and  $\alpha$ 12 in the HEK293T parental cell line used in this study are not bound to CTCF or to the cohesin subunit, Rad21 (Figure S3B). We therefore asked whether antisense transcription promotes the binding of CTCF to its binding sites in the activated exon. Consistent with the mechanistic coupling of promoter activation



#### Figure 4. Transcription of the Antisense lncRNA Triggers Activation of Sense Promoters

(A) Schematics of dCas9-VPR-mediated activation of pCBS- and eCBS-proximal promoters.

(B) Activation of sense and antisense *Pcdhα4* promoters by dCas9-VPR (RNA-seq).

(C) Activation of antisense *Pcdhα6*,  $\alpha9$ , and  $\alpha12$  promoters by dCas9-VPR (cRNA-seq). Side boxes show zoom-in view of the convergent transcription.

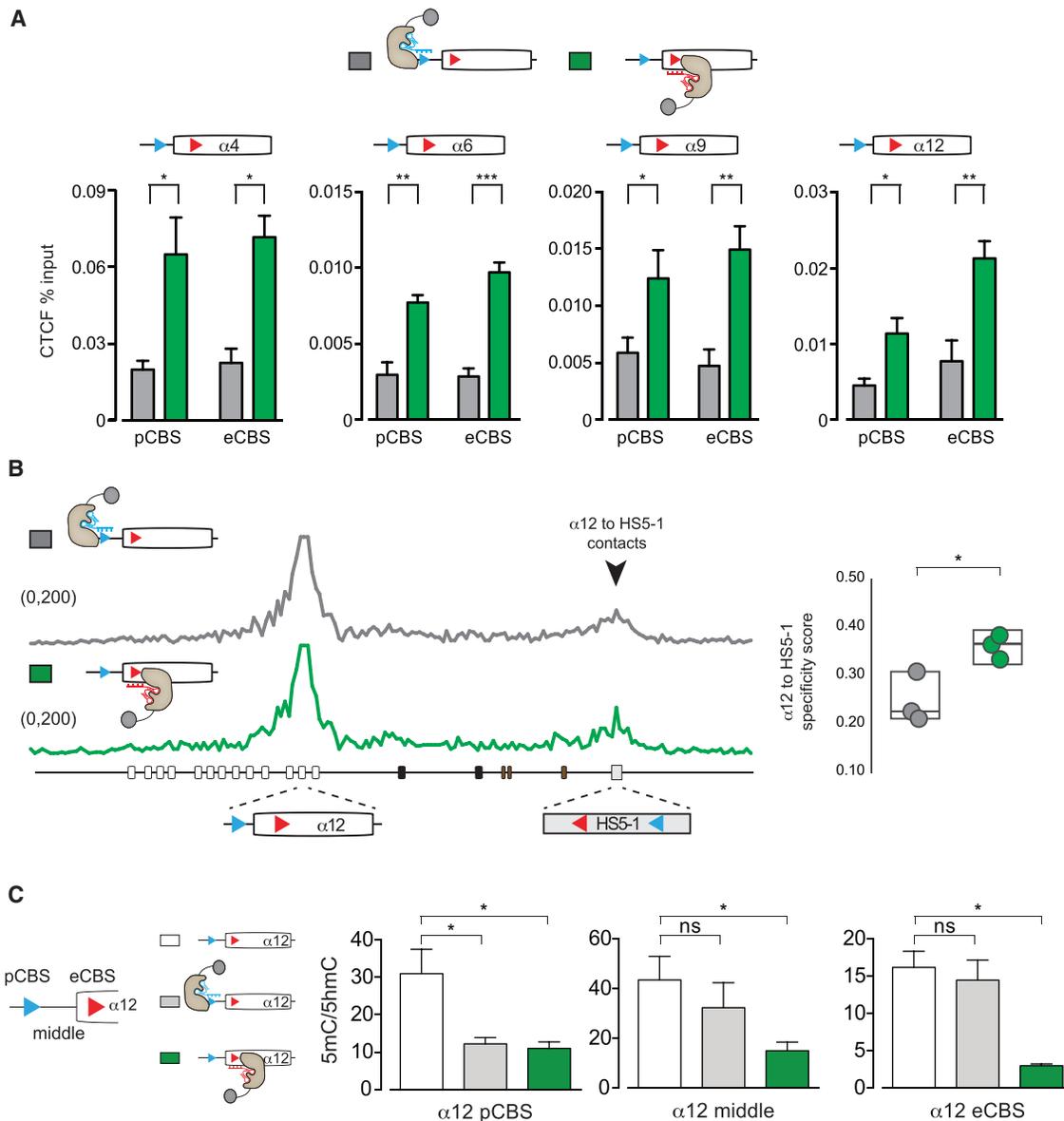
(D) Enrichment of H3K4me3 at the *Pcdhα4* promoter (ChIP-qPCR). Error bars (n = 3) represent SEM and statistical significance calculated with Student's unpaired t test.

For (B) and (C), the x axis shows the linear sequence of the genomic organization of the *Pcdhα* cluster. Arrows, initiation of transcription. Numbers on the left-hand side of each track show the minimum and maximum densities in reads per million. See also Figure S4.

and CTCF/cohesin binding (Guo et al., 2012; Monahan et al., 2012), we observed a statistically significant enrichment of CTCF occupancy at both the pCBSs and eCBSs upon dCas9-VPR activation of their antisense promoters relative to the activation of their sense promoters (Figure 5A). We note that the levels of CTCF binding at the activated *Pcdhα* promoters measured by ChIP-qPCR was lower than the one measured for a constitutive promoter, such as GAPDH, but significantly higher than an intergenic DNA site (Figure S4C). We speculate that this lower CTCF enrichment is a consequence of the high degree of cell heterogeneity as a result of transient transfections of the dCas9-VPR constructs.

The binding of CTCF to the promoters and exons of dCas9-VPR-activated genes suggests the possibility that antisense transcription from the activated exon leads to CTCF/cohesin-

dependent long-range DNA looping between the active promoter and the HS5-1 enhancer. To address this hypothesis, we focused on the *Pcdhα12* exon and performed three biologically independent *in situ* ChIP experiments with HEK293T cells transfected with dCas9-VPR to activate either the *Pcdhα12* pCBS-proximal or eCBS-proximal promoter (Figure S4D). To best discern newly formed long-range DNA contacts between the HS5-1 enhancer and the *Pcdhα12* promoter, we calculated a specificity score indicating the signal-to-noise ratio of enhancer/promoter interaction in a 15-kb window at 5-kb resolution (Figure S4E). This analysis revealed a modest but statistically significant increase in specific DNA contacts between the *Pcdhα12* promoter and the HS5-1 enhancer upon activation of the antisense *Pcdhα12* promoter compared to the sense promoter (Figure 5B). Importantly, dCas9 without the transcriptional



**Figure 5. Antisense lncRNA Transcription Promotes CTCF Binding and HS5-1 Enhancer/Promoter DNA Interactions by DNA Demethylation of the CBSs**

(A) CTCF occupancy at the pCBSs and the eCBSs of  $Pcdh\alpha 4$ ,  $\alpha 6$ ,  $\alpha 9$ , and  $\alpha 12$  upon activation of the pCBS- (gray) and the eCBS-proximal (green) promoter by dCas9-VPR (ChIP-qPCR).

(B) (Left) Virtual 4C with  $Pcdh\alpha 12$  promoter as a viewpoint for HEK293T cells activated with dCas9-VPR targeting the pCBS- (gray) or the eCBS-proximal promoter (green) of  $Pcdh\alpha 12$ . Black arrow, specific interaction between the  $Pcdh\alpha 12$  promoter and the HS5-1 enhancer. (Right) Specificity scores for the interaction of the  $Pcdh\alpha 12$  exon to the HS5-1 enhancer are shown.

(C) Relative levels of 5mC and 5hmC at the pCBS, eCBS, and middle sequences of  $Pcdh\alpha 12$  in HEK293T cells (white) and HEK293T cells with  $Pcdh\alpha 12$  sense (gray) or antisense (green) promoters activated by dCas9-VPR.

Error bars ( $n = 3$ ) represent SEM and statistical significance calculated with Student's unpaired t test. See also Figure S5.

activator domain did not result in the formation of  $Pcdh\alpha 12$ /HS51 contacts (Figure S4F).

#### Antisense lncRNA Transcription Promotes DNA Demethylation of $Pcdh\alpha$ Promoters

The data presented thus far support a model in which antisense lncRNA transcription mediates the recruitment of CTCF to active

$Pcdh\alpha$  alternate promoters. Given the observation that DNA methylation of the CBSs blocks CTCF binding (Bell and Felsenfeld, 2000) and that both pCBS and eCBS sequences contain CpG dinucleotides, we reasoned that DNA demethylation could be the mechanism by which CTCF/cohesin binds to the pCBS and eCBS following as-lncRNA transcription. To gain insight into the potential role of DNA methylation in the modularity of

CTCF binding to both pCBSs and eCBSs, we obtained nucleotide resolution of the methylation of the CpG dinucleotides within the CBSs by examining published ENCODE whole-genome bisulfite sequencing (WGBS) data from SK-N-SH cells (Figure S5). Consistent with genome-wide studies (Wang et al., 2012), these data reveal how methylation at position C2 and C12 in the core CTCF motif can affect CTCF binding at both CBSs (Figures S5C and S5D) and further suggest how additional methylation sites flanking C2 and C12 could also contribute to the regulation of CTCF binding to the pCBS and eCBS (Figure S5E). Quantification of the ENCODE WGBS data also revealed that the DNA sequence between the two *Pcdh $\alpha$*  CBSs (“middle”) is hypomethylated in active exons (Figures S5F and S5G), consistent with previous reports on the relationship between methylation, CTCF binding, and promoter activity (Guo et al., 2012; Kawaguchi et al., 2008; Tasic et al., 2002).

In mammals, 5-methylcytosine (5mC)-modified CpG sequences are converted to unmodified cytosine (C) by the activity of TET deoxygenase enzymes, which mediate the oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (Wu and Zhang, 2017). Thymine DNA glycosylase (TDG) then converts 5caC to C by a base excision repair mechanism (Wu and Zhang, 2017). 5hmC is a stable oxidation intermediate, and its detection is a proxy for a pathway to active demethylation catalyzed by the TET proteins. Therefore, to directly test the possibility that transcription of the as-lncRNA leads to demethylation of CpG elements, we measured the levels of 5mC and 5hmC for the *Pcdh $\alpha$ 12* in HEK293T cells by methylated DNA immunoprecipitation (MeDIP) upon dCas9-VPR-mediated activation of its respective sense and antisense promoters. Consistent with our hypothesis, activation of the *Pcdh $\alpha$ 12* eCBS promoter resulted in a decrease of the ratio between 5mC and 5hmC levels at the pCBS, the eCBS, and at the DNA sequence between the two CBSs (Figure 5C). By contrast, activation of the *Pcdh $\alpha$ 12* pCBS-proximal promoter resulted in a statistically significant decrease of the ratio between 5mC and 5hmC levels only for the pCBS (Figure 5C). To detect base-pair resolution of the changes occurring at the eCBS, we performed bisulfite reactions followed by Sanger DNA sequencing and observed a higher level of demethylation of all three CpG sites in the eCBS when antisense RNA is transcribed relative to when only sense transcription is initiated (Figure S5H).

### Demethylation of *Pcdh $\alpha$* Promoters Correlates with Activation *In Vivo*

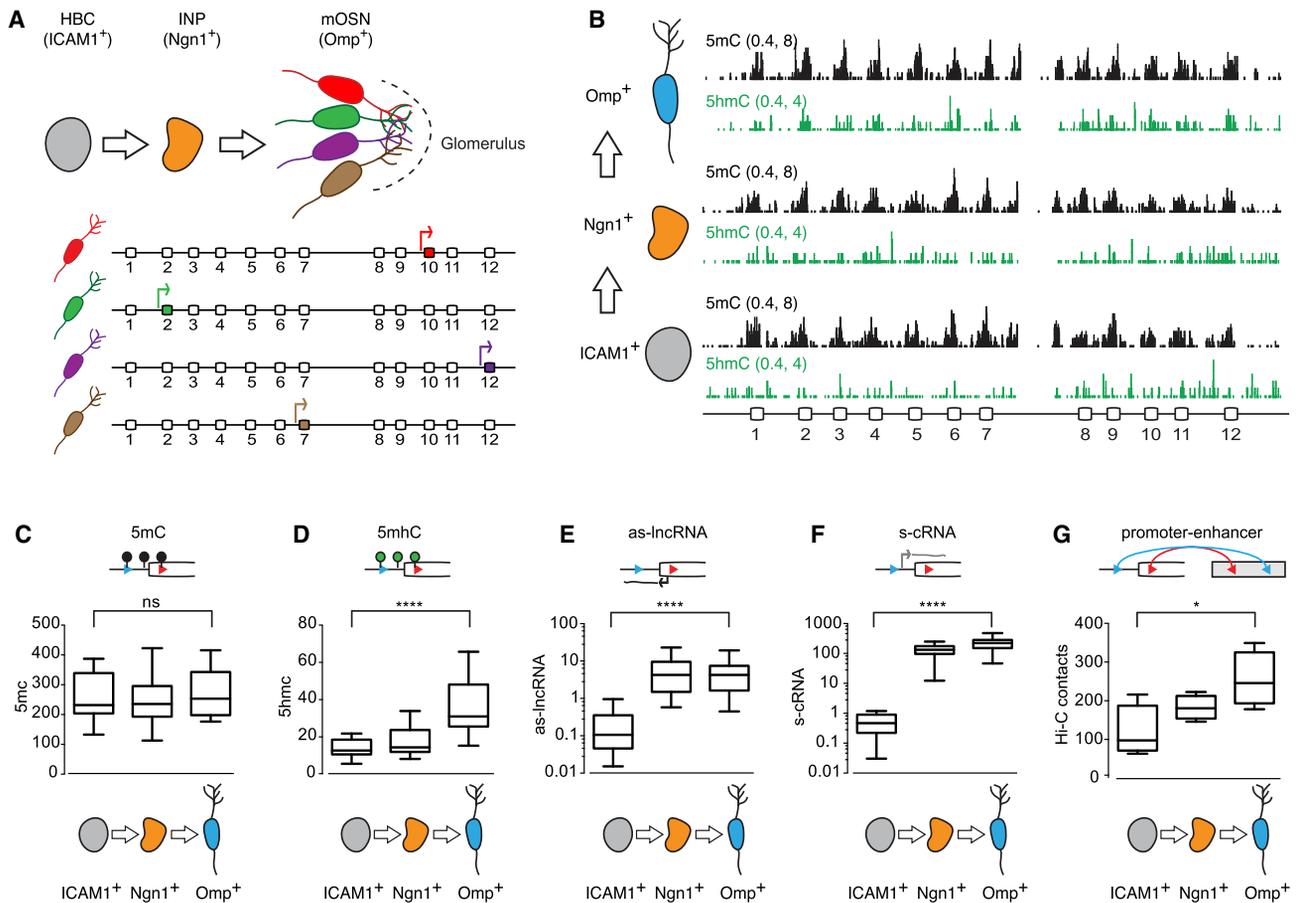
The data presented above suggest that the ground state of *Pcdh $\alpha$*  promoter DNA is methylated, and DNA demethylation, targeted by transcription of an antisense lncRNA, controls the binding of CTCF and *Pcdh $\alpha$*  sense promoter activation. To test this model *in vivo*, we made use of the mouse main olfactory sensory epithelium (mOE) as an *in vivo* developmental system to study the relationship between promoter DNA methylation and *Pcdh $\alpha$*  gene expression, as the *Pcdh* gene cluster is stochastically and combinatorially expressed in OSNs, and that *Pcdh $\alpha$*  genes play a fundamental role in OSN wiring (Hasegawa et al., 2008; Mountoufaris et al., 2017; Figure 6A). We re-analyzed recently published data from a study of the levels of 5mC and

5hmC in the three cell types that represent discrete neurodevelopmental stages in the mOE: horizontal basal cells (ICAM1<sup>+</sup>); immediate neural precursors (Ngn1<sup>+</sup>); and mature olfactory sensory neurons (Omp<sup>+</sup>; Figure 6A; Colquitt et al., 2013). Horizontal basal cells are quiescent multipotent cells that produce all of the cell types present in the mOE; immediate neural precursors are post-mitotic cell precursors of olfactory sensory neurons, and olfactory sensory neurons are terminally differentiated primary sensory neurons. Consistent with our model, we found that the *Pcdh $\alpha$*  alternate exons and their promoters are enriched in 5mC in iCAM1<sup>+</sup> cells, indicating that the pre-neuronal ground state of all *Pcdh $\alpha$*  alternate promoter DNA is methylated and repressed (Figures 6B and 6C). However, with the development of olfactory sensory neurons (ICAM1<sup>+</sup> → Ngn1<sup>+</sup> → Omp<sup>+</sup>), we observed an increase of 5hmC in the *Pcdh $\alpha$*  alternate promoters and exons (Figures 6B and 6D). To determine whether conversion of 5mC to 5hmC is accompanied by activation of *Pcdh $\alpha$*  promoters, we performed RNA-seq experiments in ICAM1<sup>+</sup>, Ngn1<sup>+</sup>, and Omp<sup>+</sup> cells. Consistent with our hypothesis, conversion of 5mC to 5hmC correlates with the expression of both antisense long noncoding and sense-coding *Pcdh $\alpha$*  RNAs (Figures 6E, 6F, and S6A). Finally, we determined whether *Pcdh $\alpha$*  expression is accompanied by the formation of long-range DNA contacts between the *Pcdh $\alpha$*  promoters and the HS5-1 enhancer *in vivo* and performed *in situ* Hi-C experiments in ICAM1<sup>+</sup>, Ngn1<sup>+</sup>, and Omp<sup>+</sup> cells (Figure S6B). We observed a strong increase in alternate promoters/HS5-1 enhancer interaction during neuronal differentiation of the mOE (Figure 6G). These data, collectively, provide *in vivo* support of our observations made in human cell lines.

### Stochastic DNA Demethylation Ensures Random *Pcdh $\alpha$* Promoter Choice by the CTCF/CoheS Complex via DNA Loop Extrusion

Analysis of the Hi-C data from Ngn1<sup>+</sup> and Omp<sup>+</sup> cells revealed architectural “stripes” along the *Pcdh $\alpha$*  gene cluster (Figures S6B and 7A), a feature that has been associated with the activity of the cohesin complex in the assembly of enhancer/promoter complexes during DNA loop extrusion (Vian et al., 2018). A prediction of the DNA loop extrusion model for the assembly of a *Pcdh $\alpha$*  enhancer/promoter complex is that uncoupling CTCF binding to *Pcdh $\alpha$*  promoters from DNA looping to the HS5-1 enhancer by the cohesin complex should result in an overall loss of expression of all *Pcdh $\alpha$*  exons. To test this possibility, we conditionally deleted the cohesin subunit, Rad21, in mouse olfactory sensory neurons (Figure S7A) using OMPiresCre. With this driver, Rad21 is deleted in post-mitotic, fully differentiated OSNs in which *Pcdh $\alpha$*  promoter choice has already occurred (Figures 6C–6G and S7B). However, upon deletion of Rad21, a loss of long-range DNA contacts between the *Pcdh $\alpha$*  promoters and the HS5-1 enhancer was observed (Figures 7A and 7B). More importantly, loss of DNA contacts correlated with a significant loss of expression of all *Pcdh $\alpha$*  exons as determined by RNA-seq (Figure 7C). Thus, continuous cohesin activity appears to be required for the maintenance of DNA looping in the *Pcdh $\alpha$*  cluster, even in the absence of cell division.

These data suggest that stochastic antisense transcription ensures random demethylation of *Pcdh $\alpha$*  promoters to ensure an



### Figure 6. DNA Demethylation at *Pcdhα* Promoters Correlates with Their Activation *In Vivo*

(A) (Top) Schematics of the maturation of the mouse main olfactory epithelium (OE) and the assembly of a functional neural circuit (glomerulus). (Bottom) Schematics of stochastic *Pcdhα* promoter choice in individual mOSNs are shown.

(B) 5mC (black) and 5hmC (green) profiles of the *Pcdhα* alternate promoters and exons in ICAM1<sup>+</sup>, Ngn1<sup>+</sup>, and Omp<sup>+</sup> cells. x axis, linear sequence of the mouse *Pcdhα* cluster. Numbers on the left-hand side of each track show minimum and maximum read densities in reads per million.

(C–F) Average of cumulative reads per million (RPM) for the *Pcdhα* alternate promoters/exons for 5mC (C), 5hmC (D), as-lncRNAs (E), and s-cRNAs (F).

(G) Average of cumulative *in situ* Hi-C contacts for the *Pcdhα* alternate promoters/exons.

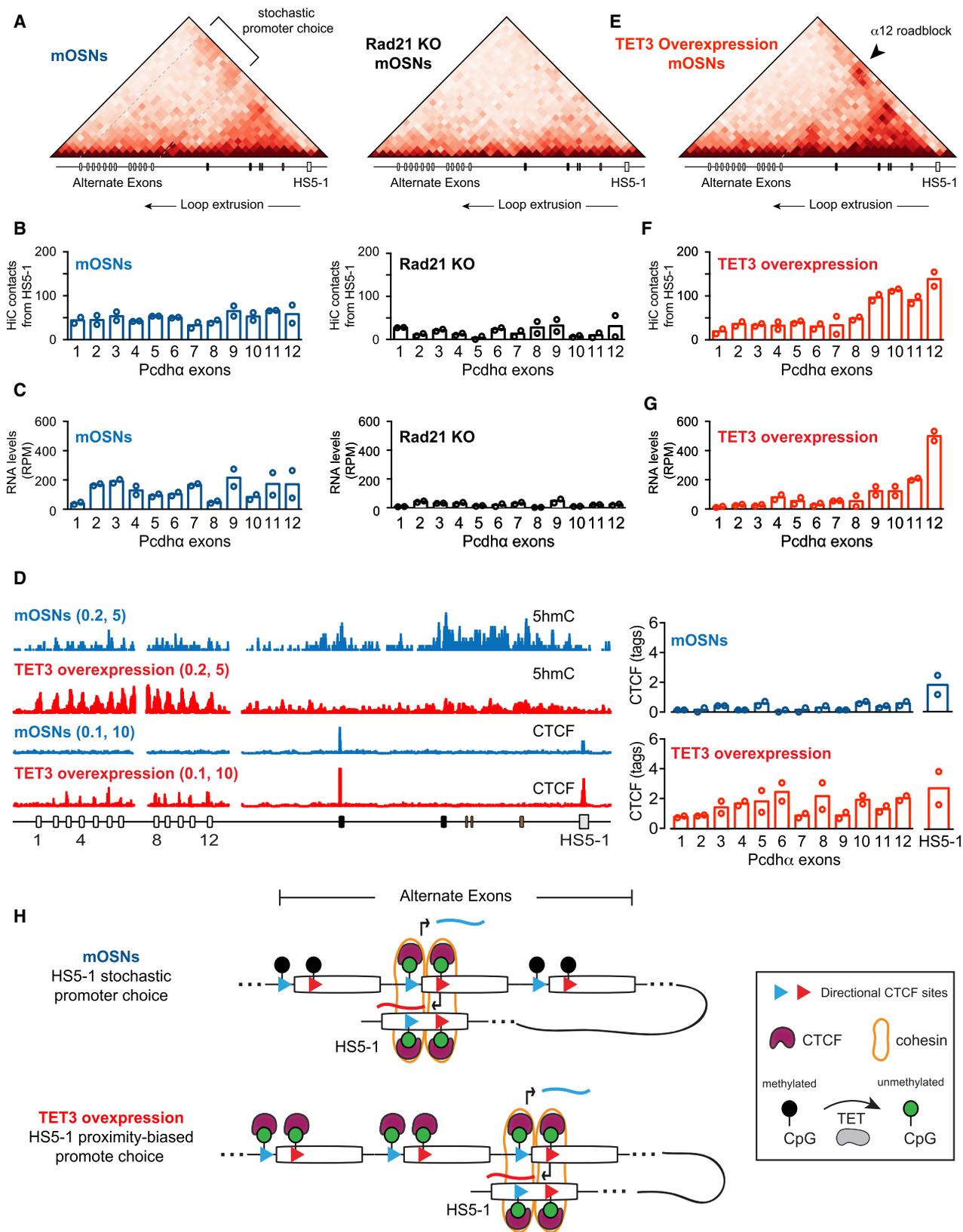
For (C)–(G), data in box and whiskers. Error bars, minimum and maximal values and statistical significance calculated with one-way ANOVA. See also Figure S6.

HS5-1-distant-independent assembly of a CTCF/cohesin-mediated enhancer/promoter complex by DNA loop extrusion. A prediction of this model is that uncoupling DNA demethylation from antisense lncRNA transcription would result in a non-random looping of *Pcdhα* promoters to the HS5-1 enhancer. We tested this possibility by overexpressing Tet3 in OSNs (Figure S7A). Tet3 is the most highly expressed Tet protein in OSNs and has been shown to associate with the *Pcdhα* promoters in differentiated neuronal precursor cells (Li et al., 2016). Overexpression of Tet3 resulted in strong demethylation of *Pcdhα* promoters, as indicated by a large increase in 5hmC levels (Figures 7D and S7C) and by an increase of CTCF binding to CBSs genome-wide (Figure S7D), and to all *Pcdhα* exons, irrespective of the transcription state of their cognate as-lncRNAs (Figures 7D and S7E). To address the function of uncoupling as-lncRNA transcription from stochastic DNA demethylation, we performed Hi-C and RNA-seq on mOSNs overexpressing Tet3. Remarkably, despite the fact that all *Pcdhα* exons are bound by CTCF

and that the expression of the as-lncRNAs is maintained (Figures 7D and S7E), overexpression of Tet3 resulted in a strong bias in *Pcdhα* promoter/HS5-1 enhancer contacts, specifically biased toward the *Pcdhα*12 promoter (Figures 7E and 7F) and a concomitant bias in *Pcdhα*12 expression relative to all other *Pcdhα* exons, as determined by RNA-seq (Figure 7G). Thus, CTCF bound to the CBSs of *Pcdhα*12 created a “roadblock” for cohesin, preventing the HS5-1 enhancer from engaging upstream *Pcdhα* promoters.

### DISCUSSION

Stochastic, combinatorial expression of individual *Pcdh* protein isoforms in Purkinje (Esumi et al., 2005) and olfactory sensory neurons (Mountoufaris et al., 2017) generates distinct combinations of Protocadherin isoforms that function as a cell-surface identity code for individual neurons (Mountoufaris et al., 2018). This conclusion has been confirmed more broadly through



(legend on next page)

single-cell RNA sequencing studies in a variety of neuronal cell types (Tasic et al., 2018). Here, we identify a mechanism by which *Pcdh $\alpha$*  alternate exon promoters are stochastically activated in individual neurons and propose a model that may apply more broadly to regulate enhancer/promoter interactions and gene expression in vertebrates.

### Insights into the Mechanism of Stochastic *Pcdh $\alpha$* Promoter Choice

We provide evidence that stochastic activation of individual *Pcdh $\alpha$*  alternate promoters requires mechanistic coupling between transcription of a large multiply spliced, polyadenylated antisense lncRNA and DNA demethylation of the *Pcdh $\alpha$*  promoters and CBSs (Figure 7H). Transcription of this lncRNA, initiated at the eCBS-proximal promoter, leads to the demethylation, de-repression, and activation of *Pcdh $\alpha$*  proximal sense strand promoters. This step occurs coordinately with CTCF binding to its CBSs located proximal to both promoters and the formation of CTCF/cohesin-dependent long-range DNA looping between the demethylated promoter and the HS5-1 enhancer. These observations are consistent with a promoter scanning mechanism, in which the HS5-1 enhancer, bound by CTCF and cohesin, translocates to the most enhancer-proximal demethylated and CTCF-bound promoter by DNA loop extrusion, leading to the stochastic production of a specific *Pcdh $\alpha$*  mRNA (Figure 7H). Remarkably, the as-lncRNA initiated at a *Pcdh $\alpha$*  eCBS-proximal promoter transcribes through its cognate pCBS-proximal promoter and extends through upstream sense promoters. However, the only sense promoter that is activated in this process is the sense promoter immediately proximal to the antisense promoter. We speculate that this proximal specificity is a consequence of functional coupling between transcription and RNA processing mediated by the C-terminal domain (CTD) of the RNAPII, the cap-binding complex, and the spliceosome (Maniatis and Reed, 2002). In support of this hypothesis, we identified a highly conserved and active 5' splice site immediately upstream of each pCBS site in the *Pcdh $\alpha$*  alternate exons (Figures 2E and 2F). Thus, the spliceosome may be recruited to the vicinity of the sense promoter by transcriptional readthrough. Although functional coupling between Tet-mediated DNA demethylation, CTCF, and the spliceosome has been reported elsewhere (Marina and Oberdoerffer, 2016), additional studies will be required to test this hypothesis in the context of *Pcdh $\alpha$*  promoters.

A fundamental question raised by our model is how antisense promoters are stochastically activated in individual neurons dur-

ing development. Given the observation that the ground state of the *Pcdh $\alpha$*  gene cluster is inactive and marked by 5mC in horizontal basal cells in the mouse olfactory epithelium, we speculate that activation of eCBS-proximal promoters in the *Pcdh $\alpha$*  gene cluster is regulated by transcription factors capable of binding methylated DNA. Finally, we note that the mechanism of stochastic promoter choice in the *Pcdh $\alpha$*  gene cluster must differ from the mechanism of promoter choice in the *Pcdh $\beta$*  and *Pcdh $\gamma$*  gene clusters. Specifically, the *Pcdh $\beta$*  and *Pcdh $\gamma$*  gene clusters do not have CBSs within the alternate exons, and we have not detected as-lncRNAs in either the *Pcdh $\beta$*  or *Pcdh $\gamma$*  gene clusters. Additional studies will be required to understand the mechanisms that underlie stochastic promoter choice in these *Pcdh* gene clusters.

### The Molecular Logic of Convergent Promoters

Convergent transcription, such as the example described for the *Pcdh $\alpha$*  alternate exons, can produce long and stable antisense noncoding RNAs that overlap with the sense-coding RNA (Brown et al., 2018). Interestingly, genes that are activated by antisense convergent RNA are characterized by an overall low level of expression of sense and antisense RNAs and a unique chromatin signature that facilitates their transcription (Brown et al., 2018). We speculate that, at least in the case described here, low levels of RNA expression, together with differences in the chromatin environment in the two convergent promoters, permits the two convergent RNAPII to productively translocate along DNA without significant interference.

The example of convergent transcription described here also suggests a model in which noncoding antisense RNA transcription couples RNAPII activity to a DNA deoxygenase TET enzyme activity. We note that there are precedents for a transcription-dependent mechanism of transcriptional activation coupled to DNA demethylation. Specifically, transcription of the tumor suppressor gene, *TCF21*, is activated by the antisense lncRNA, *TARID*, whose transcription is initiated at an intronic promoter sequence located within the *TCF21* gene (Arab et al., 2014). Transcription of *TARID* leads to the formation of promoter-associated R loops. These DNA-RNA hybrids are recognized by the growth arrest and DNA damage protein 45A, *GADD45A*, which, in turn, mediates the recruitment of TET1 to drive TET-mediated DNA demethylation and activation of the *TCF21* sense strand promoter (Arab et al., 2019). It is therefore possible that the same, or similar, mechanism is used for stochastic choice of *Pcdh $\alpha$*  promoters.

### Figure 7. Stochastic DNA Demethylation Ensures Random *Pcdh $\alpha$* Promoter Choice via DNA Loop Extrusion

(A) Hi-C contacts maps at 10-kb resolution for the *Pcdh $\alpha$*  cluster in wild-type (left) and *Rad21* KO (right) mOSNs; max: 100 reads per billion Hi-C contacts.  
 (B and C) Average Hi-C contacts of the HS5-1 enhancer with the individual *Pcdh $\alpha$*  promoters (B) and average RPM values of s-cRNA for individual *Pcdh $\alpha$*  exons (C) in mOSNs (blue) and mOSNs *Rad21* KO (black).  
 (D) (Left) 5hmC (MeDIP-seq) and CTCF (ChIP-seq) profiles in mOSNs (blue) and mOSNs upon Tet3 overexpression (red). (Right) Quantification of CTCF binding is shown.  
 (E) Hi-C contact maps at 10-kb resolution for the *Pcdh $\alpha$*  cluster in mOSNs overexpressing Tet3; max: 100 reads per billion Hi-C contacts.  
 (F and G) Average Hi-C contacts of the HS5-1 enhancer with the individual *Pcdh $\alpha$*  promoters (F) and average RPM values of s-cRNA for individual *Pcdh $\alpha$*  exons (G) mOSNs overexpressing Tet3.  
 (H) Model for how coupling of as-lncRNA transcription and DNA demethylation ensures stochastic and HS5-1 distance-independent choice of a *Pcdh $\alpha$*  promoter. Overexpression of Tet3 results in non-random and HS5-1 distance-biased *Pcdh $\alpha$*  promoter choice.  
 See also Figure S7.

## A General Mechanism for Stochastic Promoter Activation

We used the differentiating mouse olfactory epithelium as an *in vivo* model system for studying stochastic Pcdh $\alpha$  gene activation. We were therefore struck by the similarities in regulatory logic between Pcdh $\alpha$  and olfactory receptor (OR) promoter choice. In both cases, the ground state of the stochastically chosen promoters is repressed and inaccessible to transcriptional activator proteins. In the case of the Pcdh $\alpha$  gene cluster, this repression is mediated predominantly by DNA methylation (Tasic et al., 2002; Toyoda et al., 2014), and OR genes are repressed by the assembly of constitutive heterochromatin (Magklara et al., 2011). In both of these cases, however, repressive DNA or histone modifications are replaced by activating marks, concomitantly with selective binding of transcription factors that promote DNA looping between promoters and distant enhancers. As all the Pcdh $\alpha$  genes are clustered on a single chromosome, stochastic Pcdh $\alpha$  promoter choice is accomplished in *cis* via DNA looping to the enhancer. This mechanism of promoter choice differs from OR promoter choice, which has been shown to require the formation of a multi-chromosomal, multi-enhancer hub that activates only one out of 2,800 OR alleles distributed throughout the genome (Markenscoff-Papadimitriou et al., 2014; Monahan et al., 2019). Most likely, reliance on *cis* versus *trans* interactions also explains why Pcdh $\alpha$  and OR genes require distinct mechanisms to achieve transcriptional stochasticity. In the case of Pcdh $\alpha$  genes, CTCF and cohesin are critical for stochastic enhancer/promoter interactions. The proposed loop extrusion mechanism allows the HS5-1 enhancer to scan the gene cluster locally for the most proximal promoter bound by CTCF. In contrast, OR enhancers cannot deploy loop extrusion mechanisms to activate OR transcription because this process cannot accommodate *trans* chromosomal interactions, which may explain the absence of CTCF and cohesin binding sites in OR enhancers and promoters (Monahan et al., 2019). Consequently, as Pcdh $\alpha$  choice relies on stable CTCF promoter binding, DNA demethylation provides an effective mechanism for stochastic promoter activation. An important consequence of this mechanism is that, because antisense transcription and DNA demethylation are coupled and appear to occur in a stochastic fashion, DNA loop extrusion will not create a bias toward the selection of the Pcdh $\alpha$  promoter most proximal to the enhancer (Pcdh $\alpha$ 13 and Pcdh $\alpha$ 12 in human and mouse, respectively). Rather, DNA loop extrusion identifies the promoter bound to CTCF, providing an elegant mechanism to overcome selection biases driven by genomic proximity. In fact, we have shown that such a bias occurs if as-lncRNA transcription and DNA demethylation are uncoupled. Finally, our experiments highlight another important property of the loop extrusion-mediated enhancer/promoter complex mechanism: the dynamic nature of enhancer promoter interactions that requires continuous cohesin expression even in post-mitotic cells. This observation is reminiscent of the cell-division-independent role of cohesin in the expression of the T cell receptor  $\alpha$  locus (Seitan et al., 2011). Continual maintenance of promoter enhancer interactions is further highlighted by the striking observation that demethylation of all the Pcdh $\alpha$  promoters, after one is chosen, results in bias toward the HS5-1-proximal alternate promoters. These observations suggest that, if Pcdh $\alpha$

promoter choice is stable for the life of OSNs, then a mechanism must be in place to prevent demethylation of the non-chosen promoters.

It remains to be seen whether the proposed mechanism of stochastic Pcdh $\alpha$  choice is applicable to other clustered gene families where stochastic gene expression occurs. An interesting example of promoter stochasticity is the process of V(D)J recombination, whereby cohesin-mediated loop extrusion appears to bias RAG-mediated recombination of the variable Vh exons that are most proximal to the iE $\mu$  enhancer (Jain et al., 2018). However, even in this system, there is a set of Vh exons that recombine in a distance-independent fashion, which could be accomplished by similar molecular mechanisms as the ones described here, ensuring optimal diversity in the generation of immunoglobulins.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines and Cell culture
  - Generation of a CRISPR-inducible SK-N-SH cell line (SK-N-SH-iCas9)
  - Generation of SK-N-SH heterozygous for the Pcdh $\alpha$  cluster (SK-N-SH- $\alpha$ het)
  - Animals
- METHOD DETAILS
  - Fluorescence activated cell sorting of HBCs, INPs and mOSNs
  - Transfections of plasmids into HEK293T cells
  - RNA isolation and sequencing
  - Design of the myBaits Capture Library
  - RNAPII pausing
  - RNAPII elongation
  - RNA half-life
  - Chromatin Immunoprecipitation (ChIP-Seq and ChIP-qPCR)
  - In situ Chromatin Capture Conformation (Hi-C)
  - Methylated DNA Immunoprecipitation (MeDIP)
  - Bisulfite DNA Reactions
  - Immunofluorescence
  - Bioinformatic Analysis of Sequencing Data
  - CRISPR gRNA design
  - In vitro transcription of gRNAs
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cell.2019.03.008>.

## ACKNOWLEDGMENTS

We thank Drs. Richard Axel, Charles Zuker, Max Gottesman, David Hirsh, Germano Cecere, George Mountoufaris, and Enrico Cannavo and members of the

Maniatis, Lomvardas, and Simon labs for critical discussions and suggestions on the manuscript. D.C. and T.M. would like to thank Dr. Karen Adelman for advice with the Start-Seq experiment, Dr. Ye Zhang and Dr. Ben Barres for the generous gift of human brain neurons, Dr. Victor Lobanenkov and Dr. Elena Pugacheva for the CTCF monoclonal antibody, and Ira Schieren for assistance with flow cytometry. This work was supported by a Helen Hay Whitney Postdoctoral Fellowship (D.C. and R.D.), an NIH: National Institutes of Health, United States, F32 Postdoctoral Fellowship GM108474 (K.M.), an NIH F31 Predoctoral Fellowship DC016785 (A.H.), an NIH Path to Independence Award K99/R00 K99GM121815 (D.C.), NIH New Innovator Award DP2 HD083992-01 (M.D.S.), NIH grants 1R01MH108579 and 5R01NS088476 (T.M.), and NIH grant R01DC013560 and HHMI: Howard Hughes Medical Institute: USA Faculty Scholar (S.L.).

#### AUTHOR CONTRIBUTIONS

D.C. and T.M. identified, developed, and addressed the core questions regarding promoter choice. D.C. performed the bulk of the experiments with help from S.M.R. and E.L.C. C.L.N. helped D.C. in the establishment of the dCas9-VPR activation assays. A.H. and S.L. helped to develop chromosome conformation studies, and A.H. performed the experiments and analyzed the data. E.E.D. and M.D.S. helped develop the RNAPII elongation studies, and E.E.D. and D.C. performed the experiments. R.D. performed RNA sequencing experiments in the developing mouse olfactory epithelium. K.M. generated the Rad21 conditional knockout mouse line and performed Hi-C in the Rad21KO mOSNs. S.O. trained D.C. in the bioinformatics analysis of the data. D.C. and T.M. wrote the paper with the help of all the authors.

#### DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 17, 2018  
 Revised: November 8, 2018  
 Accepted: March 4, 2019  
 Published: April 4, 2019

#### REFERENCES

- Arab, K., Park, Y.J., Lindroth, A.M., Schäfer, A., Oakes, C., Weichenhan, D., Lukanova, A., Lundin, E., Risch, A., Meister, M., et al. (2014). Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A. *Mol. Cell* *55*, 604–614.
- Arab, K., Karaulanov, E., Musheev, M., Trnka, P., Schäfer, A., Grummt, I., and Niehrs, C. (2019). GADD45A binds R-loops and recruits TET1 to CpG island promoters. *Nat. Genet.* *51*, 217–223.
- Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* *405*, 482–485.
- Brown, T., Howe, F.S., Murray, S.C., Wouters, M., Lorenz, P., Seward, E., Rata, S., Angel, A., and Mellor, J. (2018). Antisense transcription-dependent chromatin signature modulates sense transcript dynamics. *Mol. Syst. Biol.* *14*, e8007.
- Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., P R Iyer, E., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J., et al. (2015). Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* *12*, 326–328.
- Colquitt, B.M., Allen, W.E., Barnea, G., and Lomvardas, S. (2013). Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cell identity. *Proc. Natl. Acad. Sci. USA* *110*, 14682–14687.
- Duffy, E.E., Rutenberg-Schoenberg, M., Stark, C.D., Kitchen, R.R., Gerstein, M.B., and Simon, M.D. (2015). Tracking distinct RNA populations using efficient and reversible covalent chemistry. *Mol. Cell* *59*, 858–866.
- 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.
- Fuchs, G., Voickek, Y., Benjamin, S., Gilad, S., Amit, I., and Oren, M. (2014). 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biol.* *15*, R69.
- Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of chromosomal domains by loop extrusion. *Cell Rep.* *15*, 2038–2049.
- Ghirlando, R., and Felsenfeld, G. (2016). CTCF: making the right connections. *Genes Dev.* *30*, 881–891.
- 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. USA* *109*, 21081–21086.
- 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.
- Hasegawa, S., Hamada, S., Kumode, Y., Esumi, S., Katori, S., Fukuda, E., Uchiyama, Y., Hirabayashi, T., Mombaerts, P., and Yagi, T. (2008). The protocadherin-alpha family is involved in axonal coalescence of olfactory sensory neurons into glomeruli of the olfactory bulb in mouse. *Mol. Cell. Neurosci.* *38*, 66–79.
- 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.
- Hollenhorst, P.C., McIntosh, L.P., and Graves, B.J. (2011). Genomic and biochemical insights into the specificity of ETS transcription factors. *Annu. Rev. Biochem.* *80*, 437–471.
- Jain, S., Ba, Z., Zhang, Y., Dai, H.-Q., and Alt, F.W. (2018). CTCF-binding elements mediate accessibility of RAG substrates during chromatin scanning. *Cell* *174*, 102–116.e14.
- Kawaguchi, M., Toyama, T., Kaneko, R., Hirayama, T., Kawamura, Y., and Yagi, T. (2008). Relationship between DNA methylation states and transcription of individual isoforms encoded by the protocadherin-alpha gene cluster. *J. Biol. Chem.* *283*, 12064–12075.
- Kehayova, P., Monahan, K., Chen, W., and Maniatis, T. (2011). Regulatory elements required for the activation and repression of the protocadherin-alpha gene cluster. *Proc. Natl. Acad. Sci. USA* *108*, 17195–17200.
- Lefebvre, J.L., Sanes, J.R., and Kay, J.N. (2015). Development of dendritic form and function. *Annu. Rev. Cell Dev. Biol.* *31*, 741–777.
- Li, X., Yue, X., Pastor, W.A., Lin, L., Georges, R., Chavez, L., Evans, S.M., and Rao, A. (2016). Tet proteins influence the balance between neuroectodermal and mesodermal fate choice by inhibiting Wnt signaling. *Proc. Natl. Acad. Sci. USA* *113*, E8267–E8276.
- Lipovich, L., Vanisri, R.R., Kong, S.L., Lin, C.-Y., and Liu, E.T. (2006). Primate-specific endogenous cis-antisense transcription in the human 5q31 protocadherin gene cluster. *J. Mol. Evol.* *62*, 73–88.
- Magklara, A., Yen, A., Colquitt, B.M., Clowney, E.J., Allen, W., Markenscoff-Papadimitriou, E., Evans, Z.A., Kheradpour, P., Mountoufaris, G., Carey, C., et al. (2011). An epigenetic signature for monoallelic olfactory receptor expression. *Cell* *145*, 555–570.
- Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. *Nature* *416*, 499–506.
- Marina, R.J., and Oberdoerffer, S. (2016). Epigenomics meets splicing through the TETs and CTCF. *Cell Cycle* *15*, 1397–1399.
- Markenscoff-Papadimitriou, E., Allen, W.E., Colquitt, B.M., Goh, T., Murphy, K.K., Monahan, K., Mosley, C.P., Ahituv, N., and Lomvardas, S. (2014). Enhancer interaction networks as a means for singular olfactory receptor expression. *Cell* *159*, 543–557.
- 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. USA* *109*, 9125–9130.

- Monahan, K., Horta, A., and Lomvardas, S. (2019). LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. *Nature* 565, 448–453.
- Mountoufaris, G., Chen, W.V., Hirabayashi, Y., O’Keeffe, S., Chevee, M., Nwakeze, C.L., Polleux, F., and Maniatis, T. (2017). Multiclusted Pcdh diversity is required for mouse olfactory neural circuit assembly. *Science* 356, 411–414.
- 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.
- Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science* 327, 335–338.
- Noonan, J.P., Li, J., Nguyen, L., Caoile, C., Dickson, M., Grimwood, J., Schmutz, J., Feldman, M.W., and Myers, R.M. (2003). Extensive linkage disequilibrium, a common 16.7-kilobase deletion, and evidence of balancing selection in the human protocadherin alpha cluster. *Am. J. Hum. Genet.* 72, 621–635.
- Ong, C.-T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* 15, 234–246.
- Pugacheva, E.M., Rivero-Hinojosa, S., Espinoza, C.A., Méndez-Catalá, C.F., Kang, S., Suzuki, T., Kosaka-Suzuki, N., Robinson, S., Nagarajan, V., Ye, Z., Boukaba, A., Rasko, J.E., Strunnikov, A.V., Loukinov, D., Ren, B., and Lobanenkov, V.V. (2015). Comparative analyses of CTCF and BORIS occupancies uncover two distinct classes of CTCF binding genomic regions. *Genome Biol.* 16, 161.
- Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680.
- Ribich, S., Tasic, B., and Maniatis, T. (2006). Identification of long-range regulatory elements in the protocadherin-alpha gene cluster. *Proc. Natl. Acad. Sci. USA* 103, 19719–19724.
- Seitan, V.C., Hao, B., Tachibana-Konwalski, K., Lavagnoli, T., Mira-Bontenbal, H., Brown, K.E., Teng, G., Carroll, T., Terry, A., Horan, K., et al. (2011). A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature* 476, 467–471.
- 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.
- Tasic, B., Yao, Z., Graybuck, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., et al. (2018). Shared and distinct transcriptomic cell types across neocortical areas. *Nature* 563, 72–78.
- Toyoda, S., Kawaguchi, M., Kobayashi, T., Tarusawa, E., Toyama, T., Okano, M., Oda, M., Nakauchi, H., Yoshimura, Y., Sanbo, M., et al. (2014). Developmental epigenetic modification regulates stochastic expression of clustered protocadherin genes, generating single neuron diversity. *Neuron* 82, 94–108.
- Vian, L., Pękowska, A., Rao, S.S.P., Kieffer-Kwon, K.R., Jung, S., Baranello, L., Huang, S.C., El Khattabi, L., Dose, M., Pruett, N., et al. (2018). The energetics and physiological impact of cohesin extrusion. *Cell* 173, 1165–1178.e20.
- 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.
- Wang, H., Maurano, M.T., Qu, H., Varley, K.E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., et al. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* 22, 1680–1688.
- Wu, Q., and Maniatis, T. (1999). A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* 97, 779–790.
- Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* 18, 517–534.
- Zhu, Z., González, F., and Huangfu, D. (2014). The iCRISPR platform for rapid genome editing in human pluripotent stem cells. *Methods Enzymol.* 546, 215–250.
- Zipursky, S.L., and Grueber, W.B. (2013). The molecular basis of self-avoidance. *Annu. Rev. Neurosci.* 36, 547–568.
- Zipursky, S.L., and Sanes, J.R. (2010). Chemoaffinity revisited: dscams, protocadherins, and neural circuit assembly. *Cell* 143, 343–353.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
CTCF monoclonal	<a href="#">Pugacheva et al., 2015</a>	N/A
Rabbit polyclonal anti Rad21	Sigma-Aldrich	ab922
Rabbit polyclonal anti H3K4me3	ThermoFisher	PA5-27029
Rabbit polyclonal anti H3K27ac	Abcam	ab4729
Mouse monoclonal anti Flag	Sigma	F1804
<b>Biological Samples</b>		
Fetal brain neurons	Dr. Ben Barres	N/A
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
5,6-Dichloro-1- $\beta$ -D-ribofuranosylbenzimidazole (DRB)	Sigma	D1916-50MG
4-Thiouridine, 98+%	Alfa Aesar	J60679
<b>Critical Commercial Assays</b>		
SMARTer® Stranded Total RNA-Seq Kit - Pico Input Mammalian	Clontech	635005
NEBNext® Ultra II DNA Library Prep Kit for Illumina	NEB	E7645S
<b>Deposited Data</b>		
Raw and analyzed data	This paper	GSE115862
<b>Experimental Models: Cell Lines</b>		
SK-N-SH cells	ATCC	HTB-11
<i>HEK293T cells</i>	ATCC	CRL-1573
<b>Experimental Models: Organisms/Strains</b>		
Mouse: R6/2: B6CBA-Tg(HDexon1)62Gpb/3J	The Jackson Laboratory	JAX: 006494
keratin5-creER;rt-gfp mice	Lomvardas Lab	N/A
ngn1-GFP	Lomvardas Lab	N/A
omp-IRES-GFP	Lomvardas Lab	N/A
tetotet3-IRES-GFP;omptta	Lomvardas Lab	N/A
tetoGFP;omptta	Lomvardas Lab	N/A
rad21fl/fl;ompre	Lomvardas Lab	N/A
<b>Recombinant DNA</b>		
U6-gRNA-CMV-dCas9-VPR	This paper	N/A
U6-gRNA-CMV-dCas9-VPR-GFP	This paper	N/A
U6-gRNA-CMV-dCas9	This paper	N/A

### CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and request for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Tom Maniatis ([tm2472@cumc.columbia.edu](mailto:tm2472@cumc.columbia.edu)).

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

#### Cell lines and Cell culture

SK-N-SH cells (XX female) were purchased from ATCC and cultured in RPMI-1640 supplemented with 10% (vol/vol) FBS, 1X GlutaMax, 1mM sodium pyruvate, 1X non-essential amino acids, and 1% penicillin-streptomycin. HEK293T cells (XX female) were purchased from ATCC and cultured in DMEM supplemented with 10% (vol/vol) FBS, 1X GlutaMax, 1mM sodium pyruvate, 1X non-essential amino acids, and 1% penicillin-streptomycin. Cells were maintained at 37°C in a 5% (vol/vol) CO<sub>2</sub> incubator.

### Generation of a CRISPR-inducible SK-N-SH cell line (SK-N-SH-iCas9)

CRISPR-inducible SK-N-SH cells were generated as previously described for Human pluripotent stem cells (hPSCs) (Zhu et al., 2014) with the following differences: (1) the Puro-Cas9 donor plasmid was substituted with a GFP-Cas9 donor plasmid and (2) the Neo-M2rtTA donor plasmid was substituted with a mCherry-M2rtTA donor plasmid. Dual color cells were sorted by flow cytometry and genotyped by PCR and further karyotyped.

### Generation of SK-N-SH heterozygous for the *Pcdh $\alpha$* cluster (SK-N-SH- $\alpha$ het)

SK-N-SH-iCas9 cells were plated at 50% density in a 6-well dish, dox-induced (at a concentration of 2 mg/mL) for 48 hours (refresh Media with 1X RPMI with Dox for every day of induction). On days 3 and 5, the cells were transfected with 1  $\mu$ g (total) of sgRNAs. On day 6, the GFP/mCherry positive and DAPI negative were single cells sorted on plates pre-coated with MEF feeder cells. The cells were allowed to grow for a month until visible colonies were observed, replica plated and genotyped by PCR. We isolated two clones (1 and 2) and named this cell line as SK-N-SH- $\alpha$ het. Deletion of one copy of the *Pcdh $\alpha$*  cluster in the SK-N-SH- $\alpha$ het1 clone was further confirmed by Sanger DNA sequencing and further karyotyped.

### Animals

Mice were treated in compliance with the rules and regulations of IACUC under protocol number AC-AAO3902. All experiments were performed on primary FACS-sorted cells from dissected main olfactory epithelium from animals (both male and female) of age between 4 to 12 weeks. HBC cells were sorted from keratin5-creER;rt-gfp mice, INP cells were sorted from the brightest GFP populations of *ngn1*-GFP mice, OSNs were sorted from *omp*-IRES-GFP mice. Rad21 conditional knockout mOSNs was achieved by crossing Rad21 conditional allele mice (Seitan et al., 2011) to OMP-ires-Cre mice (*Omp*<sup>tm1<sup>(cre)</sup>Jae</sup>). Recombined cells were purified by including a Cre-inducible tdTomato allele (*Rosa26-tdtomato*, *Gt(Rosa)26Sor*<sup>tm14(CAG-tdTomato)Hze/J</sup>) in the cross and selecting tdTomato positive cells by FACS. Overexpression of Tet3 in mOSNs was achieved by crossing tetotet3-IRES-GFP to *omptta* mice to obtain tetotet3-IRES-GFP;*omptta* mice. Control mice were achieved by crossing tetogFP to *omptta* mice to obtain tetogFP;*omptta* mice. GFP positive cells were sorted by FACS for both tetotet3-IRES-GFP;*omptta* and tetogFP;*omptta* mice. In the text and the figures, we refer to the Rad21 conditional knockout in mOSNs as Rad21 KO and the Tet3 overexpression in mOSNs as Tet3 overexpression.

## METHOD DETAILS

### Fluorescence activated cell sorting of HBCs, INPs and mOSNs

Cells were dissociated into a single-cell suspension by incubating freshly dissected main olfactory epithelium with papain for 40 minutes at 37°C according to the Worthington Papain Dissociation System. Following dissociation and filtering for three times through a 35  $\mu$ m cell strainer, cells were resuspended in 1X PBS with 5% FBS. For *in situ* Hi-C and ChIP-Seq experiments, upon dissociation, cells were fixed with 1% formaldehyde for 10 minutes at room temperature. Formaldehyde was quenched by adding glycine to a final concentration of 0.125 M for 5 minutes at room temperature. Cells were then washed with 1X cold PBS and resuspended in 1X PBS with 5% FBS. Fluorescent cells were then sorted on a BD Aria II or Influx cell sorter.

### Transfections of plasmids into HEK293T cells

One day prior to lipid-mediated transfection, HEK293T cells were seeded in a 6-well plate at a density of about 2 million cells per well. For plasmid DNA transfections, 3  $\mu$ g of total DNA was added to 125  $\mu$ L of Opti-MEM containing 5  $\mu$ L of P300 reagent, followed by an addition 125  $\mu$ L of Opti-MEM containing 7.5  $\mu$ L of Lipofectamine 3000 per well. The two solutions were mixed and incubated at room temperature for 5 minutes and the solution was added dropwise to cells. Plates were then incubated at 37°C for 48 or 72 hours in a 5% CO<sub>2</sub> incubator. After incubation, cells were harvested in 1 mL of TRIzol.

### RNA isolation and sequencing

RNA was isolated using TRIzol. Cell lysate was extracted with bromo-chloropropane and RNA was precipitated with 100% isopropanol supplemented with 10  $\mu$ g of glycoblue for 10 min at room temperature and then pelleted at 16,000 x g for 30 min at 4°C. The RNA pellet was washed once with 75% ethanol and then resuspended in RNase-free water to a maximal concentration of 200ng/ $\mu$ L. Genomic DNA contaminants were removed by Turbo DNase. Removal of Turbo DNase was performed by phenol:chloroform extraction and RNA was precipitated as described above and resuspended in RNase-free water and stored at -80°C.

Sequencing libraries for total RNA and polyadenylated RNA from SK-N-SH cells and human neurons were made using the NEBNext Ultra II Directional RNA Library Prep Kit. Sequencing libraries for total RNA from HEK293T cells and the SK-N-SH- $\alpha$ het clones were made using the SMARTer Stranded Total RNA-Seq Pico input mammalian RNA kit. The quality of all the libraries was assessed by bioanalyzer and quantified using a combination of bioanalyzer and qubit. Libraries were sequenced on a NEXT-Seq 500/550.

### Design of the myBaits Capture Library

To overcome the low level of *Pcdh* expression in both primary neurons and SK-N-SH cells, we made use of an RNA-based enrichment strategy to capture pre-processed and mature RNA species. We refer to this approach as Capture RNA-Sequencing (cRNA-Seq) (see also [Figure S1](#) for a schematic of the myBaits enrichment procedure).

myBaits targeted capture kits were designed and purchased from MYcroarray (Arbor Biosciences, <https://arborbiosci.com>). A total of 16,357 biotinylated RNA probes covering about 90.42% of the *Pcdh*  $\alpha$  (chr5: 140159476-140429082, hg19) and  $\gamma$  (chr5: 140705658-140911381, hg19) clusters were synthesized. We also designed baits for the *CBX5* locus (chr12:54624724-54673956, hg19) to serve as a positive control for our enrichment protocol. Baits were designed satisfying at least one of the following conditions:

- No blast hit with a  $T_m$  above 60°C
- No more than 2 hits at 62.5-65°C or 10 hits in the same interval and at least one neighbor candidate being rejected
- No more than 2 hits at 65-67.5°C and 10 hits at 62.5-65°C and two neighbor candidates on at least one side being rejected
- No more than a single hit at or above 70°C and no more than 1 hit at 65-67.5°C and 2 hits at 62.5-65°C and two neighbor candidates on at least one side being rejected

Sequencing libraries from RNA-Seq or HiC-Seq were multiplexed at the desired ratio and captured using the myBaits Capture Library protocol for 18 hours at 65°C. Captured libraries were eluted in RNase-free water and further amplified. The quality of captured libraries was assessed by bioanalyzer and quantified using a combination of bioanalyzer and qubit. Libraries were sequenced on a NEXT-Seq 500/550.

### RNAPII pausing

Start-Seq experiments were previously described ([Nechaev et al., 2010](#)) with the following changes: (1) about 10 million SK-N-SH cells were used for each replicate experiment, (2) the 2  $\mu$ L of RNA 5' Pyrophosphohydrolase, RppH, (NEB M0356S, 5 U/ $\mu$ L) was used in conjunction with ThermoPol Buffer (NEB B9004) to remove the 5' cap to the short-RNAs for 1 hr at 37°C,

(3) RNA-Seq libraries were prepared with the NEXTflex small RNA kit v3. Start-RNA libraries were sequenced using single-end 75-nt cycles on an Illumina NextSeq 500/550 instrument.

The location of promoter-proximal RNAPII and the transcriptional start sites (TSS) were determined by analysis of the full-length reads.

### RNAPII elongation

SK-N-SH cells were treated with 100  $\mu$ M of 5,6-Dichloro-1- $\beta$ -D-ribofuranosylbenzimidazole (DRB) or DMSO for 6 hours to block phosphorylation of the carboxy-terminal domain (CTD) of RNAPII, which is required to release paused RNAPII from promoters in the transition from initiation to productive elongation. DRB inhibition is reversible, and upon removal from the cell culture media, a wave of newly transcriptionally elongating RNAPII leads to the incorporation of 4-thiouridine ( $s^4$ U) into newly synthesized RNAs.  $s^4$ U is rapidly incorporated into living cells without the need of cell lysis or nuclear isolation. Given the thiol-specific reactivity of  $s^4$ U,  $s^4$ U-labeled nascent RNA can be covalently and reversibly captured and sequenced.  $s^4$ UDRB experiments were performed as previously described ([Fuchs et al., 2014](#)) with the following changes: 1 mM  $s^4$ U was added to media 20 min before cells were harvested. After 6h, DRB and  $s^4$ U-containing media was removed and replaced with  $s^4$ U-containing media, and cells were harvested with TRIzol after 0, 8, or 20 min after DRB removal. Cells were flash frozen and stored at  $-80^\circ\text{C}$ . A no DRB and a no  $s^4$ U controls were also performed.

Total RNA was purified and  $s^4$ U-RNA was enriched using MTS-biotin chemistry ([Duffy et al., 2015](#)). Briefly, cells were lysed in TRIzol, extracted once with chloroform and the nucleic acids were precipitated with isopropanol. DNA was removed with Turbo DNase. DNase protein was removed by phenol:chloroform:isoamylalcohol extraction, and the RNA was isolated using isopropanol precipitation. RNA was sheared to  $\sim$ 200 bp by adding shearing buffer (150 mM Tris-HCl pH 8.3, 225 mM KCl, 9 mM  $\text{MgCl}_2$ ) and heating to 94  $^\circ\text{C}$  for 4 min, followed by quenching on ice with EDTA. Sheared RNA was purified using a modified protocol with the RNeasy Mini Kit (QIAGEN). To biotinylate the  $s^4$ U-RNA, 150  $\mu$ g sheared RNA was incubated with 60  $\mu$ g MTS-biotin in biotinylation buffer (150  $\mu$ L total volume) for 30 min. Excess biotin was removed via chloroform extraction using Phase-Lock Gel Tubes. RNA was precipitated with a 1:10 volume of 3 M NaOAc and an equal volume of isopropanol and centrifuged at 20,000  $\times$  g for 20 min. The pellet was washed with an equal volume of 75% ethanol. Purified RNA was dissolved in 200  $\mu$ L RNase-free water. Biotinylated RNA was separated from non-labeled RNA using glycogen-blocked Dynabeads Streptavidin C1 Beads (Invitrogen). Beads (200  $\mu$ L) were added to each sample and incubated for 15 min at room temperature, then washed three times with high salt wash buffer (1 mL each, 100 mM Tris-HCl (pH 7.4), 10 mM EDTA, 1 M NaCl, and 0.1% Tween-20). In order to improve the stringency of the washes, an additional three washes with buffer TE (10 mM Tris pH 7.4, 1 mM EDTA) at 55  $^\circ\text{C}$  were performed.  $s^4$ U-RNA was eluted from Dynabeads with 200  $\mu$ L freshly prepared elution buffer (10 mM DTT, 100 mM NaCl, 10 mM Tris pH 7.4, 1 mM EDTA) and incubated for 15 min. Enriched RNA was purified by ethanol precipitation and re-biotinylated as above. Excess biotin was removed via chloroform extraction using Phase-Lock Gel Tubes and RNA was purified by RNeasy Mini Kit.  $s^4$ U-RNA was enriched on streptavidin beads as above and beads were washed three times with high salt wash buffer.  $s^4$ U-RNA was eluted as above and spiked with 200 pg *Schizosaccharomyces pombe* total RNA. 10 ng total RNA from input and enriched RNA samples was used for library preparation with the SMARTer Stranded

Total RNA-seq Kit Pico Input Mammalian (Clontech) according to the manufacturer's instructions. Input and enriched samples were multiplexed with Illumina barcodes and sequenced using paired-end 2 × 75-nt cycles on an Illumina NextSeq 500/550 instrument.

### RNA half-life

SK-N-SH cells were treated with 100 μM DRB for 0, 15, 30, 60, 120, 240, 480, 960 minutes to inhibit transcription. Total RNA was purified as described above and levels of antisense lncRNA and sense cRNA were measured by qPCR.

### Chromatin Immunoprecipitation (ChIP-Seq and ChIP-qPCR)

The following antibodies were used for chromatin immunoprecipitation studies: CTCF (donated by Victor Lobanenkov), Rad21 (Abcam ab992), Histone H3 Lysine 4 tri-methyl (ThermoFisher PA5-27029), Histone H3 Lysine 27 acetylation (Abcam ab4729), FLAG (Sigma F1804). With the exception of ChIP-Seq experiments for CTCF performed in mOSNs where ~1 million sorted cells were used per IP, about 5 million cells were used. Cells were crosslinked with 1% formaldehyde for 10 minutes at room temperature. Formaldehyde was quenched by adding glycine to a final concentration of 0.125 M for 5 minutes at room temperature. Cells were then washed with 1X cold PBS with protein inhibitors twice and pelleted. Cell pellets were stored at -80°C till use. Cells were lysed in lysis buffer (50 mM Tris pH 7.5, 140 mM NaCl, 0.1% SDS, 0.1% sodium deoxycholate, 1% Triton X-100) for 10 minutes. Nuclei were span for 10 minutes at 1000 g and resuspended in the sonication buffer (10 mM Tris pH 7.5, 0.5% SDS) as 5<sup>6</sup> nuclei per 300 μL sonication buffer. Chromatin was sheared by Bioruptor for 30 cycles at cycling condition 30/30 (ON/OFF time in seconds). Following a spin at 13,000 g for 10 minutes to remove debris, the sheared chromatin was diluted such as the final binding buffer concentration was 15 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS) and incubated for 2 hours with dynabeads G pre-equilibrated in the binding buffer for pre-clearing of the chromatin. Post-cleared chromatin was then incubated with the specific antibody overnight (1 μg of antibody was used per 5<sup>6</sup> nuclei). The next day, dynabeads G were added to the chromatin-antibody mix for 2 hours. A total of four washes with 1X wash buffer (100 mM Tris pH 7.5, 500 mM LiCl, 1% NP-40, 1% sodium deoxycholate) and one wash with TE buffer (10 mM Tris pH 7.5, 1 mM EDTA) were performed. The elution was performed at 65°C for 1 hour in the elution buffer (1% SDS, 250 mM NaCl, 2 mM DTT). All steps, with the exception of the elution, were performed at 4°C. All buffers, with the exception of the TE and elution buffer contained 1X protease inhibitors. The eluted chromatin was reverse-crosslinked overnight at 65°C and the DNA was purified with the Zymo DNA kit.

Libraries for ChIP-Seq were prepared using the NEBNext Ultra II DNA Library Prep Kit. The quality of the libraries was assessed by bioanalyzer and quantified using a combination of bioanalyzer and qubit. Libraries were sequenced on a NEXT-Seq 500/550.

### In situ Chromatin Capture Conformation (Hi-C)

HEK293T cells transfected with dCas9-VPR-GFP plasmids were fixed with 1% formaldehyde and GFP-positive cells were FACS-sorted. About 500,000 cells (SK-N-SH or HEK293T) were lysed and intact nuclei were processed through an *in situ* Hi-C protocol as previously described with a few modifications (Rao et al., 2014). Briefly, cells were lysed with 50 mM Tris pH 7.5 0.5% Igepal, 0.25% Sodium-deoxycholate, 0.1% SDS, 150 mM NaCl, and protease inhibitors. Pelleted intact nuclei were then resuspended in 0.5% SDS and incubated for 20 minutes at 65°C for nuclear permeabilization. After quenching with 1.1% Triton-X for 10 minutes at 37°C, nuclei were digested with 6 U/μl of DpnII in 1x DpnII buffer overnight at 37°C. Following initial digestion, a second DpnII digestion was performed at 37°C for 2 hours. DpnII was heat-inactivated at 65°C for 20 minutes. For the 1.5hr fill-in at 37°C, biotinylated dGTP was used instead of dATP to increase ligation efficiency. Ligation was performed at 25°C for 4 hours. Nuclei were then pelleted and sonicated in 10 mM Tris pH 7.5, 1 mM EDTA, 0.25% SDS on a Covaris S220 for 16 minutes with 2% duty cycle, 105 intensity, 200 cycles per burst, 1.8-1.85 W, and max temperature of 6°C. DNA was reverse cross-linked overnight at 65°C with proteinase K and RNase A.

Reverse cross-linked DNA was purified with 2x Ampure beads following the standard protocol. Biotinylated fragments were enriched using Dynabeads MyOne Streptavidin T1 beads. The biotinylated DNA fragments were prepared for next-generation sequencing on the beads by using the Nugen Ovation Ultralow kit protocol with some modifications. Following end repair, magnetic beads were washed twice at 55°C with 0.05% Tween, 1 M NaCl in Tris/EDTA pH 7.5. Residual detergent was removed by washing the beads twice in 10 mM Tris pH 7.5. End repair buffers were replenished to original concentrations, but the enzyme and enhancer was omitted before adaptor ligation. Following adaptor ligation, beads underwent five washes with 0.05% Tween, 1 M NaCl in Tris/EDTA pH 7.5 at 55°C and two washes with 10mM Tris pH 7.5. DNA was amplified by 10 cycles of PCR, irrespective of starting material. Beads were reclaimed and amplified unbiotinylated DNA fragments were purified with 0.8x Ampure beads. Quality and concentration of libraries were assessed by Agilent Bioanalyzer and Qubit. *In situ* Hi-C libraries from SK-N-SH and HEK293T cells were size-selected and enriched as described above using the myBaits Capture Library protocol described above and sequenced paired-end on NextSeq 500 (2x75bp).

### Methylated DNA Immunoprecipitation (MeDIP)

The following antibodies were used: 5-Methylcytosine (5-mC) antibody (Active Motif 39649) and 5-Hydroxymethylcytosine (5-hmC) antibody (Active Motif 39791).