Antisense lncRNA transcription drives stochastic Protocadherin α promoter choice

Daniele Canzio^{1,2,*}, Chiamaka L. Nwakeze^{1,2,*}, Adan Horta^{1,2}, Sandy M. Rajkumar^{1,2}, Eliot L. Coffey³, Erin E. Duffy⁴, Rachel Duffié^{1,2}, Matthew D. Simon⁴, Stavros Lomvardas^{1,2}, Tom Maniatis^{1,2,5,#}

1. Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, 10032
2. Mortimer B. Zuckerman Mind Brain and Behavior Institute, Columbia University, New York, NY, 10027
3. Whitehead Institute for Biomedical Research, Cambridge, 02142
4. Department of Molecular Biophysics and Biochemistry, Yale University, 06516
5. New York Genome Center, New York, NY 10013

* These authors contributed equally to the work

To whom correspondence should be addressed: tm2472@cumc.columbia.edu

The authors declare no competing financial interests.

SUMMARY

Stochastic and combinatorial activation of clustered Protocadherin (Pcdh) α , β , and γ gene promoters generates a cell-surface identity code in individual neurons that functions in neural circuit assembly. Here we show that Pcdh α promoter choice requires transcription of a long noncoding RNA (lncRNA) initiated from newly identified promoters located in the protein coding sequence of each Pcdh α exon. Antisense transcription of the lncRNA through the sense promoter results in its activation and in DNA demethylation of the binding sites for the CCCTC-binding protein, CTCF, located in close proximity to both sense and antisense promoters. Increased CTCF binding promotes the assembly of long-range DNA contacts between the activated promoter and a neuron-specific enhancer, thus locking in the epigenetic state of the stochastically chosen Pcdh α promoter. Examination of this hierarchical molecular mechanism in differentiating olfactory sensory neurons, suggests that antisense Pcdh α transcription is a key prerequisite for stochastic Pcdh α promoter choice *in vivo*.

INTRODUCTION

During brain development, individual neurons differentiate into distinct functional cell types, they 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 vs non-self) (Grueber and Sagasti, 2010; Lefebvre et al., 2015; Zipursky and Grueber, 2013). This process is required to avoid clumping or synaptic engagement between self-neurites, and at the same time, establish functional synaptic connections with other neurons of the same or different neuronal cell type (Grueber and Sagasti, 2010; Lefebvre et al., 2015; Zipursky and Grueber, 2013). The process by which neurites from the same neuron recognize and repel each other is known as self-avoidance and requires a unique combination of cell-surface homophilic recognition molecules that function as a molecular identity code or barcode (Zipursky and Grueber, 2013; Zipursky and Sanes, 2010). Previous studies have identified the nature and function of such barcodes in invertebrates and vertebrates (Mountoufaris et al., 2017; Zipursky and Grueber, 2013; Zipursky and Sanes, 2010). In flies, the *Drosophila* down syndrome cell adhesion molecule (*Dscam1*) provides unique cell surface barcodes for neuronal self-avoidance (Hattori et al., 2008). In this case, stochastic alternative RNA splicing of a *Dscam1* precursor messenger RNA (pre-mRNA) in individual neurons can generate over 19,000 distinct extracellular protein isoforms (Wojtowicz et al., 2004). In an extraordinary example of convergent evolution, the same cell-surface mechanism involving specific homophilic interactions followed by repulsion is used for self-avoidance in vertebrates. However, in this case, protein diversity is provided by the clustered Protocadherins (*Pcdhs*) rather than by *Dscam1* homolog proteins (Chen et al., 2013; Mountoufaris et al., 2017; Yagi, 2013; Zipursky and Sanes, 2010). Moreover, rather

than stochastic alternative RNA splicing of a single *Dscam1* pre-mRNA in flies, the extraordinary functional diversity of clustered *Pcdhs* is a consequence, at least in part, of the unique genomic arrangement of the *Pcdh* genes in three closely linked clusters (designated as α , β , and γ) and a mechanism of stochastic and combinatorial promoter choice which remains poorly understood (Esumi et al., 2005; Tasic et al., 2002; Wang et al., 2002; Wu and Maniatis, 1999a; Wu et al., 2001).

The three *Pcdh* gene clusters, together, span nearly 1 million base pairs (bp) of DNA, and are organized into variable and constant regions, reminiscent of the organization of immunoglobulin and T-cell receptor gene clusters (Wu and Maniatis, 1999b). The variable regions in the *Pcdh* α and γ cluster are further distinguished into alternate and c-types. The organization of the human *Pcdh* gene cluster, which is conserved throughout vertebrate evolution, is illustrated in Figure 1A. Neuron-specific expression of individual *Pcdh* α genes requires long-range DNA looping between random *Pcdh* α 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 1B). Conserved transcriptional promoter sequences are located immediately proximal to every *Pcdh* α exon (Tasic et al., 2002) while the HS5-1 enhancer is located downstream of the constant exons, between the *Pcdh* α and the β clusters (Ribich et al., 2006) (Figure 1A). These stochastic promoter/enhancer 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 1B and 1C). CTCF is an 11-zinc finger domain protein that, together with the Cohesin complex, plays a central role as an insulator of chromatin domains, and mediates genome-wide promoter/enhancer interactions (Carretero et al., 2010; Ghirlando and

Felsenfeld, 2016; Ong and Corces, 2014). All *Pcdhα* alternate exons contain two CTCF binding sites (CBS), one in the promoter (pCBS) and the other in the protein coding sequence in the downstream exon (eCBS) (Guo et al., 2012; Monahan et al., 2012) (Figure 1B). The two binding sites are separated by approximately 1000 base pairs, and similarly spaced CBS sites are located in the HS5-1 enhancer (L-CBS and R-CBS) (Guo et al., 2012; Monahan et al., 2012) (Figure 1B). Interestingly, the CTCF binding sites in *Pcdhα* promoters and the HS5-1 enhancer are in opposite relative orientations, and inversion of the HS5-1 enhancer results in a significant decrease in *Pcdhα* gene cluster expression, demonstrating the functional importance of this arrangement (Guo et al., 2015). This opposite relative orientation of promoter and enhancer CTCF binding sites appears to be a general feature of eukaryotic chromosomes genome-wide, suggesting that the relative orientation of CTCF binding sites plays an important role in determining the specificity of enhancer/promoter interactions (Guo et al., 2015; Rao et al., 2014).

An additional insight into the formation of a *Pcdhα* promoter/enhancer complex is provided by the observation that there is an inverse relationship between *Pcdhα* gene expression and DNA methylation of the GpC-rich CTCF binding sites in the *Pcdhα* promoters (Tasic et al., 2002; Toyoda et al., 2014). Specifically, the CTCF/Cohesin complex associates exclusively with active promoters, which are characterized by hypomethylation of the CTCF binding sites (CBS) (Guo et al., 2012). By contrast, CBS sites within inactive promoters are hypermethylated and appear refractive to CTCF/Cohesin binding (Guo et al., 2012). Yet, despite the fact that DNA methylation of the CTCF binding sites is likely to play an important role in the mechanism of stochastic *Pcdhα* promoter choice by the CTCF/Cohesin complex, the time during neuronal differentiation at which *Pcdhα* promoter choice is not known, nor is the ground state of *Pcdhα* promoter DNA methylation known. For instance, is the ground

state of the *Pcdhα* gene cluster methylated and repressed, and its activation involves stochastic demethylation? Or, is the ground state unmethylated and promoter choice depends on stochastic activation followed by methylation of the promoters that are not chosen?

Here, we use a combination of cell-culture and *in vivo* model systems, to provide evidence that the ground state of the DNA of *Pcdhα* promoters is methylated and transcriptionally silent, and transcription of an antisense lncRNA is required to demethylate, derepress and activate *Pcdhα* mRNA transcription through a CTCF/Cohesin-dependent long-range DNA looping between the promoter and the *Pcdhα* neuronal-specific HS5-1 enhancer.

RESULTS

Transcription of sense and antisense RNA from clustered *Pcdhα* alternate exons

The formation of a promoter/enhancer-CTCF/Cohesin complex plays a critical role in the mechanism of stochastic promoter activation of *Pcdhα* alternate exons (Guo et al., 2015; Kehayova et al., 2011; Monahan et al., 2012; Ribich et al., 2006). However, the mechanism by which random *Pcdhα* promoters are activated is not understood and mechanistic studies of promoter choice are not possible *in vivo*, as every neuron expresses a distinct repertoire of *Pcdhα* alternate exons. Therefore, we made use of the well-characterized human neuroblastoma cell line SK-N-SH, which stably expresses a distinct repertoire of *Pcdhα* 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 is indistinguishable from that observed in single neurons *in vivo* (Esumi et al., 2005; Mountoufaris et al., 2017). SK-N-SH cells thus provide a multicellular “avatar” for studying single cell expression of *Pcdh* genes and they provide internal controls for exons that are transcriptionally silent.

Another challenge to the study of Pcdh α promoter choice is the low level of expression of Pcdh genes. To optimize the analysis of Pcdh RNA precursors (pre-mRNA) and mature (mRNA) RNAs in SK-N-SH cells, we used capture RNA-Sequencing (cRNA-Seq) which allowed an enrichment for Pcdh RNA transcripts up to two orders of magnitude (Figure S1A and S1B). Remarkably, cRNA-Seq performed in SK-N-SH cells revealed a previously undetected high level of antisense RNA transcription originating within the transcriptionally active Pcdh α alternate exons (Pcdh α 4, 8 and 12) (Figure 1D and S1B). By contrast, antisense RNA transcription was not detected within the two c-type exons, α c1 and α c2, which do not contain CBSs within their exons (Figure 1D). Similarly, antisense RNA was not observed in the Pcdh β or γ variable exons in SK-N-SH cells, which also do not contain exonic CBS sites (Figure S1B). We refer to the antisense RNA as antisense long noncoding RNA, as-lncRNA, as it appears to lack protein-coding sequences, based on analyses of the open reading frames. For clarity, we refer to the sense Pcdh coding RNA as, s-cRNA (sense coding RNA).

The cRNA-Seq data obtained from SK-N-SH cells revealed a direct correlation between sense and antisense RNA transcription and transcriptionally active Pcdh α alternate exons (Figure 1D). Because transcription of the Pcdh α 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 allele. To address this, we used CRISPR-Cas9 gene editing and generate SK-N-SH cells heterozygous for the Pcdh α gene cluster (Figure S2, SK-N-SH- α het). We isolated two clones (SK-N-SH- α het 1 and 2) expressing α 12, α c1 and α c2 from the remaining copy of the Pcdh α gene cluster (Figure S3A). Both clones showed expression of the as-lncRNA and s-cRNA from Pcdh α 12 (Figure S3A and S3B), confirming that sense and antisense transcription originate from the same

allele. For one of the two clones isolated, α het-1, we also performed chromatin immunoprecipitation studies (ChIP-Seq) for CTCF and Rad21, a subunit of the Cohesin complex, as well as *in situ* capture chromosome conformation studies (cHi-C) to examine long-range DNA interactions between the active Pcdh α 12 and the HS5-1 enhancer (Figure S3A and S3C). These studies demonstrated that the Pcdh α alternate exons from which sense and antisense RNAs are transcribed are bound by the CTCF/Cohesin complex and engaged in a promoter-enhancer complex (Figure S3A). It is notable that the α het-1 and α het-2 clones share a 16.7 kilobase deletion that truncates the Pcdh α 8 exon and removes the Pcdh α 9 and α 10 exons (Figure S3C). Remarkably, this deletion was previously reported as a common feature of individuals from multiple populations of European and East Asian descent with no apparent phenotypic consequence (Noonan, 2003). This deletion appears to affect both sense and antisense RNA expression from the Pcdh α 8 exon (data not shown).

These data show that transcriptionally active Pcdh α alternative exons express both sense and antisense RNAs. Thus, in contrast to the SK-N-SH cells, a mixed population of primary neurons, each expressing a distinct repertoire of Pcdh α alternative exons, should express as-lncRNAs from the Pcdh α 1 to 13 exons, but not from Pcdh α c1 and α c2, or from the β or γ exons. As predicted, analysis of RNA from human primary neurons revealed lncRNA expression exclusively from the Pcdh α 1-13 exons, and from the HS5-1 enhancer (Figure S4A). Similarly, analysis of mouse olfactory sensory neurons (OSNs) also revealed lncRNA expression by all the Pcdh α alternate exons (Figure S4B). Thus, all Pcdh α alternative exons in human cell lines and human and mouse primary neurons, explored in this study, express as-lncRNAs. Furthermore, our RNA-Seq data revealed that the Pcdh α as-lncRNAs are polyadenylated (Figure S4).

Convergent promoters in the *Pcdh α* alternative exons and HS5-1 enhancer

In order to characterize the nature of the antisense RNAs and to explore their function, we localized the promoters driving their transcription by identifying the location of promoter-paused RNAPII using Start-Seq (Nechaev et al., 2010) (Figure 2A). RNA isolated from stalled RNAPII at promoters are approximately 15 to 45 nucleotides long, and contain a 5' 7meG-cap (Figure 2A). Isolation and 3' end sequencing of these short RNAs reveals the position of paused RNAPII, and thus acts as a proxy for the location of RNAPII-engaged promoters (Figure 2A). As expected, we observed RNAPII pausing at the pCBS-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, *Pcdh* $\alpha 4$ and $\alpha c1$ are shown). However, we also observed RNAPII pausing just upstream of the eCBS for $\alpha 4$, $\alpha 8$, and $\alpha 12$ in the antisense orientation (Figure 2B, *Pcdh $\alpha 4$* is shown). Thus, the two CBSs in the active *Pcdh α* genes act as convergent promoters (Figure 2C). This is in contrast to the sole pCBS site in *Pcdh $\alpha c1$* , which acts as a divergent promoter (Figure 2C). Start-Seq analysis also identified a similar convergent promoter architecture of the two CBSs in the HS5-1 enhancer (Figure 2B and 2C). Mapping the location of the as-lncRNA promoters with respect to the as-lncRNAs revealed that the nuclear RNA precursors can be as long as 20,000 base pairs (Figure 1D). As an example, the as-lncRNA that initiates at the eCBS-promoter of *Pcdh $\alpha 4$* in SK-N-SH cells is transcribed through the pCBS-promoter of *Pcdh $\alpha 4$* and extends in the antisense direction all the way to the *Pcdh $\alpha 1$* exon (more than 20,000 nucleotides upstream) (Figure 1D).

Antisense convergent transcription is widespread in the mammalian genome. Yet, the activity of convergent transcribing polymerases remains unclear (see Discussion for more details). To assess RNAPII activity at the pCBS- and eCBS-promoters, we analyzed transcription in SK-N-SH cells using s⁴UDRB-Seq (Fuchs et al., 2014; Singh and Padgett,

2009) (Figure 2D). This method combines synchronization of RNAPII at promoters with incorporation of the nucleoside 4-thiouridine (s^4U) during RNA synthesis. SK-N-SH cells were treated with 5,6-Dichloro-1- β -D-ribofuranosylbenzimidazole (DRB) to block RNAPII carboxy-terminal domain (CTD) phosphorylation, a step required to release paused RNAPII from promoters in their transition from initiation to productive elongation. DRB inhibition is reversible, and upon removal from the cell culture media, a wave of newly transcriptional elongating RNAPII leads to the incorporation of s^4U into newly synthesized RNAs. s^4U is rapidly incorporated into living cells without the need of cell lysis or nuclear isolation. Given the thiol-specific reactivity of s^4U , s^4U -labeled RNA can be covalently and reversibly captured and sequenced. Consistent with the Start-Seq data, we observed convergent, elongating RNAPII from both pCBS- and eCBS-promoters of $\alpha 4$, $\alpha 8$ and $\alpha 12$, and divergent RNAPII from the pCBS-proximal promoter of $\alpha c1$ (Figure 2E and Figure S5). We also observed convergent elongating RNAPII at the HS5-1 enhancer, also consistent with the presence of convergent promoters (Figure 2E and Figure S5). Thus, there is a remarkable symmetry between the location of CTCF/Cohesin binding sites and sense and antisense transcription from the *Pcdh α* promoters and the HS5-1 enhancer. However, contrary to the sense and antisense RNA transcribed from a *Pcdh α* alternate exon, both enhancer RNAs are not polyadenylated and therefore appear to turn over faster (Figure S4 and S5).

Additional evidence for the presence of convergent promoters associated with both the pCBS and eCBS sequences was provided by the analysis of published ENCODE DNaseI hypersensitivity and ChIP-Seq data, which revealed the binding of distinct classes of transcription factors (TF) to the pCBS and eCBS sites of transcriptionally active alternate exons in SK-N-SH cells (Figure S6). Specifically, TFs belonging to the ETS family bind to the pCBS-promoter, while TFs belonging to the bHLH family bind to the eCBS-promoter (Figure

S6). Interestingly, both these classes of TFs are implicated in regulating genes involved in neuronal development and differentiation, such as different members of cell-adhesion proteins (Hollenhorst et al., 2011).

Transcription of antisense lncRNAs triggers activation of sense promoters

Understanding the architecture of the sense and antisense promoters of the *Pcdhα* genes and the HS5-1 enhancer provided the opportunity to address mechanistic questions regarding *Pcdhα* stochastic promoter choice. Specifically, we aimed to test the functional relationship between transcriptional activation of the sense and antisense promoters, and transcription of their respective RNAs. To do so, we designed a gain-of-function approach that allowed us to uncouple transcription of the antisense long noncoding RNA from transcription of the sense coding *Pcdhα* mRNA in the context of the endogenous *Pcdhα* 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- or eCBS-promoters of silent *Pcdhα* genes (Figure 3A). We chose HEK293T cells, as in this cell line most *Pcdhα* genes, with the exception of *Pcdh α10* and *α2*, are silent. This made it possible to selectively activate *Pcdh α4*, *α6*, *α9* and *α12* (Figure S7). As expected, dCas9-VPR activation of the *Pcdhα4* sense promoter resulted in robust synthesis of the *Pcdhα4* s-cRNA (Figure 3B and 3C). By contrast, activation of the *Pcdhα4* antisense promoter resulted not only in high levels of transcription of antisense RNA, but high levels of sense RNA was also observed (Figure 3B and 3C). This pattern of sense and antisense RNA transcription observed upon eCBS-promoter activation mirrored that observed in primary neurons and SK-N-SH cells at steady-state (Figure 1D and Figure S4). Similar to *Pcdhα4*

gene, activation of the eCBS-promoters of the Pcdh α 6, α 9, and α 12 also resulted in transcription of both sense and antisense RNAs (Figure 3D).

These observations suggest the possibility that transcription of the antisense RNA by the eCBS-promoter activates its cognate pCBS-promoter to allow for transcription of the sense coding RNA. To test this model, we measured the levels of histone H3 lysine 4 trimethylation (H3K4me3) at the Pcdh α 4 sense promoter. This histone post-translational modification is a marker for transcriptionally active promoters and is enriched between two CBS sites of Pcdh α active exons in SK-N-SH cells (Figure 1D). We therefore carried out chromatin immunoprecipitation studies followed by quantitative PCR (ChIP-qPCR) and observed an increase in H3K4me3 at the Pcdh α 4 sense promoter upon transcription the antisense RNA by dCas9-VPR (Figure 3E).

Taken together, these data are consistent with a model in which transcription of the antisense RNA by the eCBS-promoter activates its cognate pCBS-promoter, thus generating convergent sense and antisense transcripts (Figure 3F). This level of exquisite specificity is remarkable, considering that the as-lncRNA transcribes through multiple upstream pCBS-promoters (Figure 1D and 3D), yet the only pCBS-promoter that is activated is the one proximal to the site of initiation of the antisense RNA. We speculate that this proximal specificity is a consequence of functional coupling between transcription and RNA processing mediated by the carboxy-terminal (CTD) of the RNAPII and the cap-binding complex (Maniatis and Reed, 2002).

Antisense transcription leads to CTCF binding and promoter/enhancer long-range DNA interactions

The expression of *Pcdh* α sense transcripts correlates with the binding of CTCF and Cohesin proteins to the pCBS and eCBS sites and to long-range DNA looping between active promoters and the HS5-1 enhancer (Guo et al., 2012; 2015). Using ChIP-Seq, we observed that both CBSs of the inactive *Pcdh* α 4, α 6, α 9, and α 12 in the HEK293T used in this study are not occupied by CTCF (Figure 4A). We therefore asked whether antisense transcription promotes the binding of the CTCF protein to its sites on the activated exon. To accomplish this, we again turned to the dCas9-VPR-based gain-of-function strategy described above and illustrated in Figure 3A. Consistent with the mechanistic coupling of promoter activation and CTCF/Cohesin binding (Guo et al., 2012; Monahan et al., 2012), we observed a statistically significant enrichment of CTCF occupancy at both pCBS and eCBS sites upon dCas9-VPR activation of their antisense promoters relative to the activation of their sense promoters (Figure 4B).

The binding of CTCF suggested 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 possibility, we focused on the *Pcdh* α 12 exon. We performed triplicate *in situ* cHi-C experiments on HEK293T cells transfected with dCas9-VPR to activate either the *Pcdh* α 12 pCBS- or eCBS-promoter and obtained a combined contact matrices at 2kb resolution (Figure 5A). Analysis of *in situ* cHi-C data from *Pcdh* α 12 pCBS-activated HEK293T cells showed three long-range DNA contacts corresponding to the interactions between *Pcdh* α 10, α 11 and α 13 and the HS5-1 enhancer (Figure 5B, left). All three of these exons are indeed stably bound by CTCF in our HEK293T cells (Figure 4A). Most importantly, we did not observe a long-range DNA interaction between

the *Pcdh* α 12 and the HS5-1 enhancer (Figure 5B, left). By contrast, analysis of *in situ* ChIP data from *Pcdh* α 12 eCBS-activated HEK293T cells revealed the formation of newly well-defined long-range DNA interactions between the activated *Pcdh* α 12 promoter/exon and the HS5-1 enhancer, and a concomitant reduction of contacts between the *Pcdh* α 10, α 11 and α 13 and the HS5-1 enhancer (Figure 5B, middle). Consistent with the formation of a *Pcdh* α 12/HS51 complex, we observed an increase of CTCF binding at the HS5-1 enhancer in HEK293T cells activated with dCas9-VPR targeting the eCBS- compared to the pCBS-promoter (Figure S8). Most importantly, the same type of promoter/enhancer landscape observed here is characteristic of that observed for the SK-N-SH cells (Figure S3C and Figure 5B, right). As described above, in these cells, *Pcdh* α 12 expression is epigenetically stable and both its pCBS and eCBS sites are bound by the CTCF protein and the Cohesin complex (Figure 1D) and in contact with their corresponding L- and R-CBSs on the HS5-1 enhancer (Figure S3C).

Together, these data suggest that binding of the CTCF protein and formation of long-range DNA looping between active *Pcdh* α promoters and the HS5-1 enhancer are coupled and regulated by a loop extrusion mechanism. We propose that, as previously described for other promoter-enhancer complexes in the mammalian genome (Fudenberg et al., 2016), long-range DNA looping between the active promoters and the HS5-1 enhancer occurs as a consequence of stalled Cohesin complexes by extrusion barriers. According to this model, CTCF bound to DNA acts as such an extrusion barrier, and therefore stalls actively translocating Cohesin molecules at the activated *Pcdh* α promoters. By contrast, the inactive *Pcdh* α promoter, which is not bound by CTCF, is not brought into contact with the HS5-1 enhancer.

Antisense lncRNA transcription leads to the conversion of 5mC to 5hmC at Pcdh α promoters

The data presented thus far suggest a model in which the stochastic choice of Pcdh α alternate promoters requires coupling between transcription of antisense lncRNAs and the assembly of a promoter/enhancer complex by CTCF/Cohesin. However, it is not clear how transcription of the antisense lncRNA leads to the recruitment of the CTCF/Cohesin complex, the assembly of a functional promoter/enhancer complex, and stable transcriptional activation of a Pcdh α sense coding RNA. As mentioned above, a key difference between Pcdh α active and inactive sense promoters is that the former display low levels of CpG methylation by the pCBS and eCBS sites, while the later are highly methylated (Guo et al., 2012; Kawaguchi et al., 2008; Tasic et al., 2002). Given this correlation, and the observation that DNA methylation of the CTCF binding sites blocks CTCF binding (Bell and Felsenfeld, 2000), a potential mechanism of CTCF/Cohesin binding to exons following as-lncRNA transcription is demethylation of the methylated CpG sequences.

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). To test the possibility that transcription of the as-lncRNA leads to demethylation of CpG islands, we measured the levels of 5mC and 5hmC for the Pcdh α 4 and the Pcdh α 12 promoter/exons in HEK293T cells by Methylated DNA Immunoprecipitation (MedIP) upon dCas9-VPR mediated activation of their respective sense (pCBS) and antisense (eCBS) promoters. Consistent with our hypothesis, activation of the Pcdh α 4 and α 12 eCBS-promoters resulted in a decrease of

5mC/5hmC levels at the pCBS and the eCBS sites (Figure 6A and 6B). By contrast, activation of the *Pcdhα4* and the *Pcdhα12* pCBS-promoter resulted in only a statistically significant decrease of 5mC/5hmC levels by the pCBS but not the eCBS (Figure 6A and 6B). These data are consistent with the hypothesis that transcription of the antisense lncRNA promotes CpG DNA conversion from 5mC to 5hmC of both pCBS and eCBS sites, and suggest that demethylation of both sites is a required step for the stable binding of CTCF to its DNA binding sites.

Conversion of 5mC to 5hmC at *Pcdhα* promoters correlates with their activation *in vivo*

The data presented above are consistent with a model in which the ground state of a *Pcdhα* promoter DNA is methylated, and conversion of 5mC to 5hmC, targeted by transcription of an antisense lncRNA, controls *Pcdhα* 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 *Pcdhα* gene expression. Previous studies have shown that the *Pcdh* gene cluster is stochastically and combinatorially expressed in OSNs, and that the *Pcdhα* genes play a fundamental role in OSN wiring (Hasegawa et al., 2008; 2016; Mountoufaris et al., 2017). We re-analyzed recent published work carried out to characterize the levels of 5mC and 5hmC in the three cell types that represent discrete neurodevelopmental stages in the mOE: horizontal basal cells (*iCam1*⁺), immediate neural precursors (*Ngn1*⁺) and mature olfactory sensory neurons (*Omp*⁺) (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 cells precursors to olfactory sensory neurons. Olfactory sensory neurons are

terminally differentiated primary sensory neurons. Consistent with our model, we found that clustered *Pcdhα* alternate exons and their promoters are enriched in 5mC in *iCam1*⁺ cells indicating that the ground state of all *Pcdhα* alternate promoters is methylated and repressed (Figure 7A and 7B). However, with the development of olfactory sensory neurons (*iCam1*⁺ → *Ngn1*⁺ → *Omp*⁺), we observed an increase of 5hmC in the *Pcdhα* alternate promoters and exons (Figure 7A and 7B). To test whether conversion of 5mC to 5hmC is accompanied by activation of *Pcdhα* promoter and their expression, we performed RNA-Seq experiments in *iCam1*⁺, *Ngn1*⁺ and *Omp*⁺ cells. Consistent with our hypothesis, conversion of 5mC to 5hmC correlated with the expression of both antisense long noncoding and sense coding *Pcdhα* RNAs (Figure 7C). Finally we tested whether *Pcdhα* expression is accompanied by the formation of long-range DNA contacts between the *Pcdhα* promoters and the HS5-1 enhancer *in vivo*, and performed *in situ* Hi-C experiments in *iCam1*⁺, *Ngn1*⁺ and *Omp*⁺ cells. Consistent with our model, we observed a remarkable increase in alternate promoters/HS5-1 enhancer interactions during neuronal differentiation of the mOE (Figure 7D and Figure S9).

DISCUSSION

Stochastic, combinatorial expression of individual *Pcdh* protein isoforms in purkinje (Esumi et al., 2005) and olfactory sensory neurons (Mountoufaris et al., 2017), as revealed by single-cell RNA expression studies, generates distinct combinations of Protocadherin isoforms that function as a cell-surface identity code for individual neurons. This code is required for self-avoidance of sister neurites, a process critical for normal neuronal connectivity and neurite patterning (Lefebvre et al., 2012; Mountoufaris et al., 2017). The extraordinary functional diversity of *Pcdhs* required for self-avoidance is a consequence of a mechanism of stochastic promoter activation of alternate exon promoters, which occurs

independently in each of the three gene clusters, and on both allelic chromosomes. This diversity is further expanded through nearly random dimerization of distinct Pcdh isoforms, the unique anti-parallel interactions of homophilic Pcdh complexes at the cell surface, and a stringent homophilic specificity (Rubinstein et al., 2017). In contrast to purkinje and olfactory neurons, however, serotonergic neurons require only a single Pcdh identity provided by the expression of the c-type, Pcdh α c2 protein, which functions in axonal tiling (Chen et al., 2017; Katori et al., 2017). Thus, the establishment of differential transcriptional programs regulating the expression of the alternate and c-type Pcdh gene clusters isoforms enables distinct Pcdh functions in different neuronal cell types. Uncovering such transcriptional programs will reveal fundamental principles of neural circuit assembly and the role of Pcdh-mediated neural cell-surface diversity in the mammalian brain. Here we identify a mechanism by which Pcdh α alternate exon promoters are stochastically activated in individual neurons, and discuss the possibility that this mechanism may function more broadly in promoter choice and gene expression in vertebrates.

Insights into the mechanism of stochastic promoter choice

Here, we provide evidence that stochastic activation of individual Pcdh α alternate promoters requires mechanistic coupling between transcription of an antisense lncRNA and DNA looping between individual Pcdh α promoters and the HS5-1 enhancer (Figure 7E). This mechanism requires the binding of the CTCF/Cohesin complex to both the promoter and enhancer. Specifically, each Pcdh α alternate exon is characterized by convergent promoters in close proximity to the two CTCF/Cohesin binding sites (pCBS and eCBS). The eCBS-promoter initiates transcription of a long noncoding RNA that extends through the pCBS-promoter and into upstream intergenic sequences, leading to transcriptional activation of the

pCBS-promoter. This process is accompanied by the conversion of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) of the pCBS and eCBS sequences. While 5mC is usually associated with transcriptionally inactive genes, 5hmC is associated with gene activation (Colquitt et al., 2013; Ficz et al., 2011; Kriaucionis and Heintz, 2009). This epigenetic switch from 5mC to 5hmC allows for the binding of the CTCF protein to its dual binding sites. CTCF, together with the Cohesin complex, mediates long-range DNA interactions between the active promoter and the *Pcdh α* cluster-specific enhancer, HS5-1. Formation of a promoter/enhancer complex commits the *Pcdh α* promoter choice and thus facilitates transcription of a *Pcdh α* mRNA.

A fundamental question raised by these observations is how antisense promoters are stochastically activated in individual neurons during development. Given the observation that the ground state of the *Pcdh α* gene cluster is inactive and marked by 5mC (Figure 7A and 7B), we speculate that activation of eCBS-promoters in the *Pcdh α* gene cluster is regulated by the activity of transcription factors capable of binding methylated DNA. In contrast to the *Pcdh α* gene cluster, the alternate exons in the *Pcdh β* and *Pcdh γ* clusters bear a single CBS site in their promoters (pCBS), but lack a CTCF/Cohesin binding site and the antisense promoter in the downstream exon and therefore do not transcribe antisense lncRNAs. Nevertheless, *Pcdh β* and *Pcdh γ* promoter choice is stochastic (Esumi et al., 2005; Mountoufaris et al., 2017) and transcriptional enhancer elements, similar to the HS5-1 enhancer, located distal to the *Pcdh γ* gene cluster have been proposed to be required for their transcription (Yokota et al., 2011). Thus, the mechanism of random promoter choice in these gene clusters remains unknown, and is likely to be cell type-specific. Indeed, by contrast to the *Pcdh α* gene cluster, which is expressed exclusively in the nervous system, the *Pcdh β* and *Pcdh γ* gene clusters are expressed more broadly (<https://www.gtexportal.org/home/>).

The molecular logic of convergent promoters

Bi-directional RNA transcription is a common feature of mammalian promoters and enhancers (Wu and Sharp, 2013; Core et al., 2014; Henriques, et al, 2018). The transcripts can be divergent, thus non-overlapping, or, as is the case of the *Pcdh α* gene cluster, convergent, which produces overlapping complementary RNAs. Divergent transcription at promoters usually produces upstream non-coding RNAs, transcribed toward the 5' end of the gene, that are on average 50-2000 nucleotides long and relatively unstable (Wu and Sharp, 2013). Divergent transcription is widespread in mammals and is characteristic of active promoters that lack a distinct TATA box motif and are enriched of CpG islands (Wu and Sharp, 2013). By contrast, convergent transcription can produce long and stable antisense noncoding RNAs that often overlap with the sense coding RNA (Brown et al., 2018; Mayer et al., 2015). These antisense RNAs can function to either activate or repress transcription of the coding RNA from the sense promoter, in a process known as transcription interference (Bonasio and Shiekhataar, 2014). Interestingly, contrary to divergent promoters, active genes marked by convergent transcription are characterized by a unique chromatin signature (Brown et al., 2018; Mayer et al., 2015; Murray and Mellor, 2016; Scruggs et al., 2015). In this case, antisense RNA transcription has been proposed to actively shape a unique chromatin environment necessary to promote transcription of the cognate sense RNA. The example of convergent transcription described here suggests a model in which noncoding antisense RNA transcription plays a fundamental role in stochastic promoter activation by establishing an epigenetically stable chromatin architecture at *Pcdh α* promoters, which is critical to the formation of proper promoter/enhancer long-range DNA looping. This mechanism involves coupling between RNAPII, a DNA deoxygenase TET enzyme and the insulator complex CTCF/Cohesin. We note that a transcription-dependent mechanism of transcriptional

activation is not unprecedented. Specifically, transcription of the tumor suppressor gene, TCF21, was shown to be activated by an antisense RNA whose transcription is initiated at an intronic promoter sequence located within the TCF21 gene (Arab et al., 2014). Like the mechanism proposed here, transcription through the TCF21 promoter leads to TET-mediated DNA demethylation and activation of the TCF21 sense strand promoter. Here, we show that this mechanism is used for stochastic choice of *Pcdh α* promoters, which has profound implications regarding neuronal circuit assembly during development.

De-repression followed by stochastic enhancer-promoter interactions as a general mechanism for transcriptional activation

Since we used the differentiating mouse olfactory epithelium as an *in vivo* model system for stochastic *Pcdh α* gene activation, we could not ignore the striking similarities in the regulatory logic between *Pcdh α* 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 α* gene cluster, this repression is mediated predominantly by DNA methylation (Tasic et al., 2002; Toyoda et al., 2014), while OR genes are repressed by the assembly of constitutive heterochromatin (Magklara et al., 2011; Monahan et al., 2017). In both of these cases, however, repressive DNA or histone modifications are replaced by activating marks, concomitantly with selective binding of transcription factors that recruit distant transcriptional enhancers to the transcriptional start site. Because all the *Pcdh α* genes are clustered in a single chromosome, stochastic *Pcdh α* choice can be accomplished with cis-acting enhancers only. This is in contrast to OR gene choice, which appears to require the formation of a multi-chromosomal multi-enhancer hub that activates only one out the 2800 OR alleles distributed throughout the genome (Horta et

al., 2018; Markenscoff-Papadimitriou et al., 2014; Monahan et al., 2018). Most likely, reliance on *cis* versus *trans* interactions also explains why the two systems require different molecular players to achieve transcriptional stochasticity: in the case of *Pcdhα* genes, CTCF and Cohesin are ideal to promote stochastic enhancer/promoter interactions since the loop extrusion mechanism will essentially allow the HS5-1 enhancer to scan the genome 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* interactions, explaining the absence of CTCF and Cohesin from OR enhancers and promoters (Monahan et al., 2018). Consequently, as *Pcdhα* choice relies on stable CTCF promoter binding, DNA methylation/hydroxymethylation provides an effective mechanism for stochastic promoter activation. An important consequence of this mechanism is that, since antisense transcription and hydroxymethylation appear to occur in a stochastic fashion, and independently of the HS5-1 enhancer, loop extrusion will not create a bias toward the selection of the *Pcdhα* promoter most proximal to the enhancer (*Pcdhα13* and *Pcdhα12* in human and mouse, respectively). Instead, loop extrusion will identify the promoter that is bound by CTCF, providing an elegant mechanism to overcome selection biases driven by genomic proximity. Therefore, it would not be surprising if other clustered genomic systems use the same principle to assure that distant DNA elements are utilized as frequently as proximal ones, with prime example the process of V(D)J recombination, whereby proximity bias would be detrimental to the generation of the most diverse immunoglobulins and T cell receptors possible.

AUTHORS 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 substantial help from C.L.N and help from R.M.S. and E.L.C. A.H. and S.L. helped to develop the chromosome conformation studies, and A.H. performed the experiment with the help of D.C. and C.L.N 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. helped D.C. performing RNA sequencing experiments in the mouse olfactory epithelium. D.C. and T.M. wrote the paper with the help of all the authors.

ACKNOWLEDGEMENTS

We thank Ira Schieren for assistance with flow cytometry. We thank Dr. Axel, Dr. Zuker, Dr. Max Gottesman, David Hirsh, Dr. Germano Cecere, Dr. Kevin Monahan, Dr. Mountoufaris and members of the Maniatis, Lomvardas and Simon labs for helpful discussions and suggestions on the manuscript. D.C. would like to thank Sean O’Keeffe for his training in bioinformatics analysis of the data and Dr. Karen Adelman for advice with the Start-Seq experiments. D.C. and T.M. would like to thank Dr. Ye Zhang and Dr. Ben Barres for the generous gift of human brain neurons and Dr. Victor Lobanenko and Dr. Elena Pugacheva for the generous gift of the CTCF monoclonal antibody. This work was supported by the Helen Hay Whitney Postdoctoral Fellowship and NIH Path to Independence Award K99/R00 K99GM121815 (D.C.), and the NIH grants 1R01MH108579 and 5R01NS088476 (T.M.).

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.
- Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* 405, 482–485.
- Bonasio, R., and Shiekhattar, R. (2014). Regulation of Transcription by Long Noncoding RNAs. *Annu. Rev. Genet.* 48, 433–455.
- 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.
- Carretero, M., Remeseiro, S., and Losada, A. (2010). Cohesin ties up the genome. *Current Opinion in Cell Biology* 22, 781–787.
- 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 Meth* 12, 326–328.
- Chen, W.V., Maniatis, T., Maniatis, T., and Chen, W.V. (2013). Clustered protocadherins. *Development (Cambridge, England)* 140, 3297–3302.
- Chen, W.V., Nwakeze, C.L., Denny, C.A., O'Keeffe, S., Rieger, M.A., Mountoufaris, G., Kirner, A., Dougherty, J.D., Hen, R., Wu, Q., et al. (2017). *Pcdhac2* is required for axonal tiling and assembly of serotonergic circuitries in mice. *Science* 356, 406–411.
- 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. U.S.A.* 110, 14682–14687.
- Duffy, E.E., and Simon, M.D. (2009). *Enriching s4U-RNA Using Methane Thiosulfonate (MTS) Chemistry* (Hoboken, NJ, USA: John Wiley & Sons, Inc.).
- 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.
- Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Systems* 3, 95–98.
- 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.

- Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398–402.
- Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology* **32**, 279–284.
- Fuchs, G., Voichkek, 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.
- Fuchs, G., Voichkek, Y., Rabani, M., Benjamin, S., Gilad, S., Amit, I., and Oren, M. (2015). Simultaneous measurement of genome-wide transcription elongation speeds and rates of RNA polymerase II transition into active elongation with 4sUDRB-seq. *Nat Protoc* **10**, 605–618.
- Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports* **15**, 2038–2049.
- Ghirlando, R., and Felsenfeld, G. (2016). CTCF: making the right connections. *Genes & Development* **30**, 881–891.
- González, F., Zhu, Z., Shi, Z.-D., Lelli, K., Verma, N., Li, Q.V., and Huangfu, D. (2014). An iCRISPR Platform for Rapid, Multiplexable, and Inducible Genome Editing in Human Pluripotent Stem Cells. *Stem Cell* **1–31**.
- Grueber, W.B., and Sagasti, A. (2010). Self-avoidance and tiling: Mechanisms of dendrite and axon spacing. *Cold Spring Harbor Perspectives in Biology* **2**, a001750.
- Guo, Y., Maniatis, T., Monahan, K., Myers, R.M., Monahan, K., Wu, H., Gertz, J., Varley, K.E., Li, W., Myers, R.M., et al. (2012). CTCF/cohesin-mediated DNA looping is required for protocadherin promoter choice. *Proceedings of the National Academy of Sciences* **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. *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.
- Hasegawa, S., Kumagai, M., Hagihara, M., Nishimaru, H., Hirano, K., Kaneko, R., Okayama, A., Hirayama, T., Sanbo, M., Hirabayashi, M., et al. (2016). Distinct and Cooperative Functions for the Protocadherin- α , - β and - γ Clusters in Neuronal Survival and Axon Targeting. *Front. Mol. Neurosci.* **9**, 529.
- Hattori, D., Millard, S.S., Wojtowicz, W.M., and Zipursky, S.L. (2008). Dscam-mediated cell recognition regulates neural circuit formation. *Annu. Rev. Cell Dev. Biol.* **24**, 597–620.

- 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 Reports* 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.
- Horta, A., Monahan, K., Bashkirova, L., and Lomvardas, S. (2018). Cell type-specific interchromosomal interactions as a mechanism for transcriptional diversity. *bioRxiv* 287532.
- Katori, S., Noguchi-Katori, Y., Okayama, A., Kawamura, Y., Luo, W., Sakimura, K., Hirabayashi, T., Iwasato, T., and Yagi, T. (2017). Protocadherin- α C2 is required for diffuse projections of serotonergic axons. *Scientific Reports* 7, 15908.
- 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. *Journal of Biological Chemistry* 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. *Proceedings of the National Academy of Sciences* 108, 17195–17200.
- Knight, P.A., and Ruiz, D. (2013). A fast algorithm for matrix balancing. *IMA J Numer Anal* 33, 1029–1047.
- Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930.
- Lefebvre, J.L., Kostadinov, D., Chen, W.V., Maniatis, T., and Sanes, J.R. (2012). Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. *Nature* 488, 517–521.
- 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.
- Magklara, A., Yen, A., Colquitt, B.M., Clowney, E.J., Magklara, A., 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.
- 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.
- Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S. (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *161*, 541–554.

- Monahan, K., Horta, A., Mumbai-Wafula, A., Li, L., Zhao, Y., Love, P., and Lomvardas, S. (2018). Ldb1 mediates trans enhancement in mammals. *bioRxiv* 287524.
- 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- α gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 9125–9130.
- Monahan, K., Schieren, I., Cheung, J., Mumbey-Wafula, A., Monuki, E.S., and Lomvardas, S. (2017). Cooperative interactions enable singular olfactory receptor expression in mouse olfactory neurons. *Elife* **6**, 1083.
- Mountoufaris, G., Chen, W.V., Hirabayashi, Y., O'Keeffe, S., Chevee, M., Nwakeze, C.L., Polleux, F., and Maniatis, T. (2017). Multiclust Pcdh diversity is required for mouse olfactory neural circuit assembly. *Science* **356**, 411–414.
- Murray, S.C., and Mellor, J. (2016). Using both strands: The fundamental nature of antisense transcription. *BioArchitecture* **6**, 12–21.
- Nechaev, S., Fargo, D.C., Santos, dos, 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.
- Ong, C.-T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* **15**, 234–246.
- 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. *Science* **343**, 1603–1610.
- Ribich, S., Tasic, B., and Maniatis, T. (2006). Identification of long-range regulatory elements in the protocadherin- α gene cluster. *Proceedings of the National Academy of Sciences* **103**, 19719–19724.
- Rubinstein, R., Goodman, K.M., Maniatis, T., Shapiro, L., and Honig, B. (2017). Structural origins of clustered protocadherin-mediated neuronal barcoding. *Semin. Cell Dev. Biol.* **69**, 140–150.
- Scruggs, B.S., Gilchrist, D.A., Nechaev, S., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2015). Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. *Mol. Cell* **58**, 1101–1112.
- Shykind, B.M., Rohani, S.C., O'Donnell, S., Nemes, A., Mendelsohn, M., Sun, Y., Axel, R., and Barnea, G. (2004). Gene switching and the stability of odorant receptor gene choice. *Cell* **117**, 801–815.
- Singh, J., and Padgett, R.A. (2009). Rates of in situ transcription and splicing in large human genes. *Nat Struct Mol Biol* **16**, 1128–1133.
- 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.
- Toyoda, S., Okano, M., Tarusawa, E., Kawaguchi, M., Hirabayashi, M., Kobayashi, T., Toyama, T., Oda, M., Nakauchi, H., Yoshimura, Y., et al. (2014). Developmental epigenetic modification regulates stochastic expression of clustered protocadherin genes, generating single neuron diversity. *Neuron* **82**, 94–108.
- 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 & Development* **16**, 1890–1905.
- Wojtowicz, W.M., Zipursky, S.L., Clemens, J.C., Flanagan, J.J., and Millard, S.S. (2004). Alternative Splicing of Drosophila Dscam Generates Axon Guidance Receptors that Exhibit Isoform-Specific Homophilic Binding. *118*, 619–633.
- Wu, Q., and Maniatis, T. (1999a). A striking organization of a large family of human neural cadherin-like cell adhesion genes. *97*, 779–790.
- Wu, Q., and Maniatis, T. (1999b). A striking organization of a large family of human neural cadherin-like cell adhesion genes. *97*, 779–790.
- Wu, Q., Maniatis, T., Noonan, J.P., Zhang, T., Cheng, J.F., Kim, Y., Grimwood, J., Schmutz, J., Dickson, M., Zhang, M.Q., et al. (2001). Comparative DNA Sequence Analysis of Mouse and Human Protocadherin Gene Clusters. *Genome Research* **11**, 389–404.
- Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat Rev Genet* **18**, 517–534.
- Wu, X., and Sharp, P.A. (2013). Divergent transcription: a driving force for new gene origination? *155*, 990–996.
- Yagi, T. (2013). Genetic basis of neuronal individuality in the mammalian brain. *J. Neurogenet.* **27**, 97–105.
- 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.
- Zhu, Z., González, F., and Huangfu, D. (2014). The iCRISPR platform for rapid genome editing in human pluripotent stem cells. *Meth. 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. *143*, 343–353.

MAIN FIGURE LEGENDS

Figure 1: Transcription of sense and antisense RNA from *Pcdhα* alternative exons

(A) Genomic organization of the human *Pcdh* gene cluster. (B) Top: an example of DNA looping between the HS5-1 enhancer and a *Pcdhα* promoter (*Pcdhα12* is shown). Bottom: Architecture of the CTCF binding sites (CBSs) in *Pcdhα* alternative exons and the HS5-1 enhancer. The CTCF/Cohesin complex mediates DNA looping between a chosen promoter and the HS5-1 enhancer. (C) Chromosome-independent promoter/enhancer long-range DNA looping. (D) Sense (grey) and antisense (black) transcription from the *Pcdhα* cluster in SK-N-SH cells. Distribution of CTCF, Rad21 (a Cohesin subunit) and H3K4me3 assayed by ChIP-Seq indicate transcriptionally active exons in SK-N-SH cells. *Pcdhαc2* is active but not bound by CTCF or Rad21. The active exons are highlighted in yellow. The x-axis represents the linear sequence of the genomic organization of the *Pcdhα* cluster. The numbers on the left-hand side of each track represent the minimum and maximum densities in read per million.

Figure 2: Convergent promoters in the *Pcdhα* alternative exons and HS5-1 enhancer

(A) Schematic diagram of Start-Seq. (B) Paused RNAPII relative to total RNA and CTCF binding. (C) Promoter architectures for *Pcdhα4* (convergent), *Pcdhαc1* (divergent) and the HS5-1 enhancer (convergent). (D) Schematic diagram of s⁴UDRB-cRNA-Seq. (E) Transcription by RNAPII relative to total RNA and H3K4me3 in SK-N-SH cells. The x-axis represents the linear sequence of the genomic organization of the *Pcdhα* cluster. The numbers on the left-hand side of each track represent the minimum and maximum densities in read per million.

Figure 3: Transcription of the antisense lncRNA triggers activation of sense promoters

(A) Schematics of dCas9-VPR-mediated activation of pCBS- and eCBS-promoters. (B) Distribution of dCas9-VPR targeted to the pCBS-promoter (blue) and the eCBS-promoter (red) of *Pcdhα4* in HEK293T cells assayed by ChIP-Seq. (C) Activation of the sense and antisense promoters in *Pcdhα4* by dCas9-VPR. (D) Transcription of sense and antisense RNA upon activation of the eCBS-promoters of *Pcdh α4, α6, α9, α12* by dCas9-VPR. The data for *Pcdh α6, α9, α12* and no activation are cRNA-Seq. The data for *Pcdh α4* are RNA-Seq. (E) Enrichment of H3K4me3 at the *Pcdhα4* promoter measured by ChIP-qPCR. (F) Schematic diagram for how transcription of the as-lncRNA from an eCBS-promoter leads to activation of pCBS sense promoter and transcription of the s-cRNA. For (B-D), the x-axis represents the linear sequence of the genomic organization of the *Pcdhα* cluster. For (B) and (D), the numbers on the left-hand side of each track represent the minimum and maximum densities in read per million. Those numbers are indicated as the y-axis. For (E), errors (n=3) represent s.e.m. and statistical significance was calculated with a Student unpaired *t*-test.

Figure 4: Antisense lncRNA transcription leads to CTCF binding

(A) Distribution of CTCF in HEK293T cells assayed by ChIP-Seq. The exons free of CTCF and activated in this study are highlighted in yellow. (B) Enrichment of CTCF occupancy at the pCBS and the eCBS sites of *Pcdh α4, α6, α9, α12* assayed by ChIP-qPCR. Errors (n>3) represent s.e.m. and statistical significance was calculated with a Student unpaired *t*-test.

Figure 5: Antisense lncRNA transcription leads to promoter/enhancer DNA interactions

(A) *In situ* cHi-C contact maps at 2kb resolution for the *Pcdhα* cluster in HEK293T cells activated with dCas9-VPR targeting the pCBS-promoter (Left) or the eCBS promoter (Right)

of Pcdh α 12. Coordinates: chr5:140,780,000-141,046,000, hg38. The position of the Pcdh α 12 exon and the HS5-1 enhancer are indicated by the dotted lines. The inset is a zoom-in of the region surrounding the Pcdh α 12 exon in complex with the HS5-1 enhancer. (B) Local peak analysis of *in situ* cHi-C long-range interactions between Pcdh α promoters and the HS5-1 enhancer. *In situ* cHi-C of HEK293T cells activated with dCas9-VPR targeting the pCBS promoter (Left) or the eCBS promoter (Middle) of Pcdh α 12. *In situ* cHi-C of SK-N-SH cells (Right).

Figure 6: Antisense lncRNA transcription leads to the conversion of 5mC to 5hmC

(A) Relative levels of 5mC and 5hmC at the pCBS of Pcdh α 4 and Pcdh α 12. (B) Relative levels of 5mC and 5hmC at the eCBS of Pcdh α 4 and Pcdh α 12. Levels of 5mC and 5hmC are assayed by MeDIP-qPCR. Errors (n=3) represent s.e.m. and statistical significance was calculated with a Student unpaired *t*-test.

Figure 7: Conversion of 5mC to 5hmC at Pcdh α promoters correlates with their activation *in vivo*

(A) 5mC (Black) and 5hmC (Green) profiles of the Pcdh α alternate promoters and exons in horizontal basal cells (iCam1⁺), immediate neural precursors (Ngn1⁺) and mature olfactory sensory neurons (Omp⁺) of the mouse main olfactory epithelium. The x-axis represents the linear sequence of the genomic organization of the mouse Pcdh α cluster. The numbers on the left-hand side of each track represent the minimum and maximum read densities in read per million. (B) Average of cumulative RPM values for the Pcdh α alternate promoters/exons for 5mC (top) and 5hmC (bottom). (C) Average of cumulative RPM values for the Pcdh α antisense lncRNA (as-lncRNA) and sense coding RNA (s-cRNA). (D) Average of cumulative

in situ Hi-C contacts for the Pcdh α alternate promoters/exons. (E) Model of which transcription of a lncRNA triggers activation of Pcdh α promoter and expression of its mRNA by mediating DNA demethylation of the pCBS and eCBS, a step required for the binding of the CTCF/Cohesin protein complex. CTCF/Cohesin binding promotes the assembly of a promoter/HS5-1 enhancer complex.

For (B-D), data are represented in Box and whiskers. Error bars represent minimum and maximal values and statistical significance was calculated with one-way ANOVA.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1: RNA-Sequencing and Capture RNA-Sequencing

(A) Schematic diagram of Capture RNA-Sequencing (cRNA-Seq). The white, pink and blue bars indicate RNA from the Pcdh α , β and γ clusters, respectively. The brown bars indicates RNA from the rest of the genome. (B) RNA-Seq and cRNA-Seq from total RNA isolated from SK-N-SH cells. Red bar: myBaits for the Pcdh α and γ clusters.

Figure S2: Generation of a SK-N-SH cell heterozygous for the Pcdh α gene cluster

(A) Pcdh α cluster. Red arrows indicate the location of the gRNAs used to delete a copy of the Pcdh α cluster. (B) Array Comparative Genomic Hybridization (aCGH) analysis confirms the deletion of a copy of a Pcdh α cluster in the SK-N-SH- α het1 cells.

Figure S3: Antisense and sense RNA are transcribed from the same Pcdh α allele

(A) RNA-Seq, ChIP-Seq and *in situ* cHi-C from the SK-N-SH- α het-1. (B) Expression of Pcdh α 4 and 12 in SK-N-SH- α het 1 and 2 clonal cells compared to SK-N-SH-WT cells. (C) *In situ* cHi-C contact maps at 10kb resolution SK-N-SH- α het-1 (Left) and SK-N-SH-WT (Right) cells. Coordinates: chr5:140,780,000-141,050,000, hg38.

Figure S4: Expression of the *Pcdhα* cluster *in vivo*

(A) Polyadenylated (PolyA) RNA and Total RNA from human primary neurons. (B) Polyadenylated (PolyA) RNA and Total RNA from mouse olfactory sensory neurons.

Figure S5: 4sUDRB-Seq studies in SK-N-SH cells

Different time points and conditions of s⁴UDRB-Seq relative to total RNA and H3K4me3 in SK-N-SH cells.

Figure S6: Transcription Factors associated with the *Pcdhα* convergent promoters

DNaseI hypersensitivity and ChIP-Seq data for distinct transcription factors associated with the active exons in SK-N-SH cells relative to CTCF and total RNA.

Figure S7: Location of the gRNAs used to activate *Pcdhα* promoters by dCas9-VPR

Location relative to the pCBS and the eCBS sites of the gRNAs used to activate *Pcdh α4* (A), *α6* (B), *α9* (C) and *α12* (D)

Figure S8: Increase binding of CTCF at the HS5-1 enhancer upon activation of the *Pcdhα12* antisense promoter

Enrichment of CTCF occupancy at the L-CBS and the R-CBS sites of the HS5-1 enhancer assayed by ChIP-qPCR. Errors (n>3) represent s.e.m. and statistical significance was calculated with a Student unpaired *t*-test.

Figure S9: Chromosome architecture of the *Pcdh* gene cluster during the development of the olfactory epithelium

In situ cHi-C contact maps at 10kb resolution for horizontal basal cells (iCam1⁺), immediate neural precursors (Ngn1⁺) and mature olfactory sensory neurons (Omp⁺).

STAR METHODS

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).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines and Cell culture

SK-N-SH cells 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, 1% penicillin-streptomycin. HEK293T cells were purchased from ATCC and cultured in DMEM supplemented with 10% (vol/vol) FBS, 1X GlutaMax, 1mM sodium pyruvate, 1X non-essential amino acids, 1% penicillin-streptomycin. Cells were maintained at 37°C in a 5% (vol/vol) CO₂ 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) (González et al., 2014; 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 α cluster (SK-N-SH- α 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 μ g (total) of sgRNAs. On day 6, the GFP/mCherry positive and DAPI negative were single-cell 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- α het. Deletion of one copy of the Pcdh α cluster in the SK-N-

SH- α het1 clone was further confirmed by a CLG microarray test from Cell Line Genetics (www.clgenetics.com).

Animals

Mice were treated in compliance with the rules and regulations of IACUC under protocol number AC-AAAO3902. All experiments were performed on primary FACS-sorted cells from dissected main olfactory epithelium. 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 (Shykind et al., 2004).

METHODS DETAILS

Fluorescence activated cell sorting of HBCs, INPs and OSNs

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 μ m cell strainer, cells were resuspended in 1X PBS with 5% FBS. For *in situ* Hi-C experiments, after 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 1.8^6 cells per well. For plasmid DNA transfections, 3 μ g of total DNA was added to 125 μ L of Opti-MEM containing 5 μ L of P300 reagent, followed by an addition 125 μ L of Opti-MEM containing 7.5 μ L of Lipofectamine 3000 per well. The two solutions were mixed and incubated at room temperature for 5 minutes and the solution added dropwise to cells. Plates were then incubated at 37°C for 48 or 72 hours in a 5% CO₂ 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 μ 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/ μ 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- α 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 at the Columbia Genome Center.

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 purchase from MYcroarray (Arbor Biosciences, <http://www.arborbiosci.com>). A total of 16,357 biotinylated RNA probes covering about 90.42% of the Pcdh α (chr5: 140159476-140429082 and γ (chr5:140705658-140911381) clusters were synthesized. Baits were design 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 at the Columbia Genome Center.

RNAPII pausing

Start-Seq experiments were previously described (Nechaev et al., 2010) with the following changes: (1) about 10^6 SK-N-SH cells were used for each replicate experiment and (2) RNA-Seq libraries were prepared with the NEXTflex small RNA kit v3. The 3' end of the Start-RNA libraries were sequenced to determine the location of RNAPII pausing.

RNAPII elongation

SK-N-SH cells were treated with 100 μ M DRB or DMSO for 6 hours. s^4 UDRB experiments were performed as previously described (Fuchs et al., 2014; 2015) 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°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 and Simon, 2009; Duffy et al., 2015). Briefly, cells were lysed in TRIzol, extracted with chloroform once and the nucleic acids precipitated with isopropanol. DNA was removed with Turbo DNase. DNase protein was removed by phenol:chloroform:isoamylalcohol extraction, and RNA isolated using isopropanol precipitation. RNA was sheared to ~200 bp by adding shearing buffer (150 mM Tris-HCl pH 8.3, 225 mM KCl, 9 mM $MgCl_2$) and heating to 94 °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 μ g sheared RNA was incubated with 60 μ g MTS-biotin in biotinylation buffer (150 μ 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 x g for 20 min. The pellet was washed with an equal volume of 75% ethanol. Purified RNA was dissolved in 200 μ L RNase-free water. Biotinylated RNA was

separated from non-labelled RNA using glycogen-blocked Dynabeads Streptavidin C1 Beads (Invitrogen). Beads (200 μ 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 °C were performed. s^4 U-RNA was eluted from Dynabeads with 200 μ 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 $\times$ 75-nt cycles on an Illumina NextSeq 500/550 instrument.

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). 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 spun for 10 minutes at 1000g and resuspended in the sonication buffer (10 mM Tris pH 7.5, 0.5% SDS) as 5^6 nuclei per 300 μ l sonication buffer. Chromatin was sheared by bioruptor for 30 cycles at cycle condition 30/30 (ON/OFF time in seconds). Following a spin at 13,000g 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 μ g of antibody was used per 5⁶ nuclei). The next day, dynabeads G were added to the chromatin-antibody mix for 2 hours. A total of 4 washes with 1X wash buffer (100 mM Tris pH 7.5, 500 mM LiCl, 1% NP-40, 1% sodium deoxycholate) and 1 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)**

1.5⁶ SK-N-SH or HEK293T cells 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/ μ l 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 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 5 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).

HEK293T cells were transfected with the appropriate set of dCas9 plasmids and incubated at 37°C for 72 hours in a 5% CO₂ incubator. Genomic DNA was extracted using the PureLink Genomic DNA Mini Kit (Invitrogen). A total of 2 µg of DNA was diluted into 300 µl TE sonication buffer (10 mM Tris pH 7.5, 1 mM EDTA). Genomic DNA was sheared by bioruptor for 18 cycles at cycle condition 30/90 (ON/OFF time in seconds). The sheared DNA was diluted to a final IP buffer of 15 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100 and incubated overnight with 1 µg of antibody. The next day, a mixture of dynabeads A and G were added to the DNA-antibody mix for 2 hours. A total of 3 washes with 1X IP buffer were performed. The elution was performed at 55°C for 3 hour with rigorous shaking in the elution buffer (1% SDS, 250 mM NaCl). All steps, with the exception of the elution, were performed at 4°C. The eluted DNA was purified with the Zymo DNA kit.

Bioinformatic Analysis of Sequencing Data

For RNA-Seq experiments, raw FASTQ files were aligned with either Tophat or STAR using hg19 or mm10 reference genomes. When libraries were made following the SMARTer Stranded Total RNA-Seq, the initial 4 base pairs of both paired reads were trimmed prior to alignment.

For ChIP-Seq experiments, raw FASTQ files were aligned using Bowtie2 using hg19 reference genome upon adapter sequences removal using CutAdapt. Uniquely aligning reads were selected using Samtools and reads with alignment quality below 30 (-q 30) were removed. The HOMER software package was used to generate signal tracks.

For *in situ* Hi-C experiments, raw FASTQ files were processed through use of the Juicer Tools Version 1.76 pipeline (Durand et al., 2016) with one modification. Reads were aligned to hg38 using BWA 0.7.17 mem algorithm and specifying the -5 option implemented specifically for *in situ* Hi-C data. For captured Hi-C libraries, contact matrices were normalized to 2kb resolution by first reporting counts as reads per billion Hi-C contacts, then by normalizing with the Knight Ruiz (KR) matrix balancing algorithm (Knight and Ruiz, 2013) focused on the alpha Pcdh cluster (chr5:140780000-141046000; hg38). For uncaptured libraries (mm10 Hi-C), matrices were KR normalized genome wide.

For generating a contact matrix, scales were set to a minimum of 0 reads and a maximum of 2*(mean normalized reads) in order to report a relative enrichment of contacts.

Local peak analysis was performed by normalizing contacts between HS5-1 (chr5: 141020000- 141066000) and alternative Pcdh exons (chr5:140850000-140898000) to reads per billion Hi-C contacts at 5kb resolution. Scales were set to 0 and the maximum value for each plot.

DNaseI and ChIP data for H3K4me3, CTCF, Rad21, ELF1, GABP, TCF12, MAX, YY1 in SK-N-SH cells were obtained from the ENCODE data matrix.

In situ Hi-C data for INP and OSN cells were obtained from (Horta et al., 2018).

CRISPR gRNA design

All guide RNA (gRNAs) were designed as truncated 18mer long sequences to increase their binding specificity as previously described (Fu et al., 2014) using the CRISPR design web tool (<http://crispr.mit.edu>). With the exception of the Pcdh α 9, where a total of 2 gRNAs were used for activate either the pCBS- or the eCBS-promoters, we used 4 gRNAs for the activation of the pCBS- and eCBS-promoters of Pcdh α 4, α 6, α 12.

***In vitro* transcription of gRNAs**

The gRNAs were transcribed using the MEGAshortscript T7 Transcription Kit by Life Technologies (AM1354M), purified by phenol-chloroform and transfected in the SK-N-SH-iCas9 cells by RNAimax lipofectamine reagent.

QUANTIFICATION AND STATISTICS

The statistical tests used in this study are indicated in the respective figure legends. In general, data with single independent experiments were analyzed by Student unpaired *t*-test to determine statistical significant effects ($p < 0.05$). Data with multiple independent experiments were analyzed by one-way ANOVA to determine statistical significant effects ($p < 0.05$).

DATA AND SOFTWARE AVAILABILITY

The data discussed in this work will be available upon request.

SUPPLEMENTAL DATA TABLES

Supplemental Data Table 1: Table of the primers used in this study

Supplemental Data Table 2: Table of the guide RNAs (gRNAs) used in this study

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CTCF monoclonal	Pugacheva et al, 2015	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
Biological Samples		
Fetal brain neurons	Dr. Ben Barres	N/A
Chemicals, Peptides, and Recombinant Proteins		
5,6-Dichloro-1- β -D-ribofuranosylbenzimidazole (DRB)	Sigma	D1916-50MG
4-Thiouridine, 98+%	Alfa Aesar	J60679
Critical Commercial Assays		
SMARTer® Stranded Total RNA-Seq Kit - Pico Input Mammalian	Clontech	635005
NEBNext® Ultra™ II DNA Library Prep Kit for Illumina	NEB	E7645S
Deposited Data		
Raw and analyzed data	This paper	GEO: X
Experimental Models: Cell Lines		
SK-N-SH cells	ATCC	HTB-11
HEK293T cells	ATCC	CRL-1573
Experimental Models: Organisms/Strains		
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
Recombinant DNA		
CMV-dCas9-VPR	This paper	N/A

Supplemental Table 1: Table of the primers used in this study

ID	Name	Organism	Sequence	Use	Assay
DC_h_45	RPLP0 forward	human	CCTCACTGAGATCAGGGACATGTTGC	RNA expression	qPCR
DC_h_47	RPLP0 reverse	human	TGGTGCCCTGGAGATTTAGTGG	RNA expression	qPCR
DC_h_362	Pcdha4 as-lncRNA forward	human	TGGTAAAGAGCGACATCATGTG	RNA expression	qPCR
DC_h_363	Pcdha4 as-lncRNA reverse	human	ATGTGGACGTTGATCTCTCAGC	RNA expression	qPCR
DC_h_19	Pcdha4 s-cRNA forward	human	CAGAGAAGATCAGCTGCAGACA	RNA expression	qPCR
dc_h_chip_6	Pcdha4 s-cRNA reverse	human	CCCAGGGAAATTTGACTCTTAG	RNA expression	qPCR
DC_h_19	Pcdha4 spliced forward	human	CAGAGAAGATCAGCTGCAGACA	RNA expression	qPCR
DC_h_24	Pcdha4 spliced reverse	human	GCTCTCAGGGAGGCAGAGTAA	RNA expression	qPCR
DC_h_551	Pcdha12 as-lncRNA forward	human	ATTGGAATGTGGTGCCGAGA	RNA expression	qPCR
DC_h_552	Pcdha12 as-lncRNA reverse	human	CCACCTAAGACGGACCTCATG	RNA expression	qPCR
DC_h_21	Pcdha12 s-cRNA forward	human	GCCCAAGCCTTCAGCTGTC	RNA expression	qPCR
DC_h_chip_18	Pcdha12 s-cRNA reverse	human	GAGAATGTAGTTATAACCTAAACGG	RNA expression	qPCR
DC_h_21	Pcdha12 spliced forward	human	GCCCAAGCCTTCAGCTGTC	RNA expression	qPCR
DC_h_24	Pcdha12 spliced reverse	human	GCTCTCAGGGAGGCAGAGTAA	RNA expression	qPCR
DC_h_chip_24	Pcdha4 pCBS forward	human	TACTTACGGTTTGGAGCCACAT	ChIP and MEDIP	qPCR
DC_h_chip_39	Pcdha4 pCBS reverse	human	GGAATTTTCCCTTCTTCTCTCT	chromatin IP and DIP	qPCR
DC_h_chip_105	Pcdha4 pCBS/eCBS middle forward	human	GGCGTCTGCTGCTCTTACTTC	chromatin IP and DIP	qPCR
DC_h_chip_106	Pcdha4 pCBS/eCBS middle reverse	human	CGAGTAGTGGAGCTGACCGT	chromatin IP and DIP	qPCR
DC_h_chip_107	Pcdha4 eCBS forward	human	GAACCTGTCCATCGCGGAATC	chromatin IP and DIP	qPCR
DC_h_chip_108	Pcdha4 eCBS reverse	human	AAGTATTCATTTGGGCTCAGTCTG	chromatin IP and DIP	qPCR
DC_h_chip_26	Pcdha6 pCBS forward	human	GATGTCGCTGTCTACCATGAAGT	chromatin IP	qPCR
DC_h_chip_41	Pcdha6 pCBS reverse	human	CAGTCAGGGAATGATGCAGTAA	chromatin IP	qPCR
DC_h_chip_123	Pcdha6 eCBS forward	human	CGCCCTTGTTCCCGGTAGAG	chromatin IP	qPCR
DC_h_chip_124	Pcdha6 eCBS reverse	human	CTAGCCCGAAGTATTCGCTAG	chromatin IP	qPCR
DC_h_chip_29	Pcdha9 pCBS forward	human	ACTTTGGGCCACGTGATGTC	chromatin IP	qPCR
DC_h_chip_44	Pcdha9 pCBS reverse	human	ATGGGAAAAGGGCTGTATTTGT	chromatin IP	qPCR
DC_h_chip_125	Pcdha9 eCBS forward	human	CATTAACGACAACCTCCAGTG	chromatin IP	qPCR
DC_h_chip_126	Pcdha9 eCBS reverse	human	CACGTCCAGGAAGAAATACTCA	chromatin IP	qPCR
DC_h_chip_127	Pcdha12 pCBS forward	human	GACCCAGGAAGTGGCTAAACC	chromatin IP and DIP	qPCR
DC_h_chip_130	Pcdha12 pCBS forward	human	GTGAATGATTGGAATGTGGG	chromatin IP and DIP	qPCR
DC_h_chip_113	Pcdha12 pCBS/eCBS middle forward	human	CGTCTGCTGCTCTCGCTTC	chromatin IP and DIP	qPCR
DC_h_chip_114	Pcdha12 pCBS/eCBS middle forward	human	CGTGTTTGGCCTCCTCGTAG	chromatin IP and DIP	qPCR
DC_h_chip_115	Pcdha12 eCBS forward	human	CGCCGGTGTTCAGAGAAAGG	chromatin IP and DIP	qPCR
DC_h_chip_116	Pcdha12 eCBS forward	human	GCATCAGAAGCGCCTCTAG	chromatin IP and DIP	qPCR
DC_h_chip_77	HS5-1 L-CBS forward	human	GGGAGTGTCTGAGGGCTGAA	chromatin IP	qPCR
DC_h_chip_78	HS5-1 L-CBS reverse	human	TTCCGTACCATATGGATTGCTT	chromatin IP	qPCR
DC_h_chip_79	HS5-1 L-CBS/R-CBS middle forward	human	CAAATGTAGTCCGTCCTCAGTG	chromatin IP	qPCR
DC_h_chip_80	HS5-1 L-CBS/R-CBS middle reverse	human	GTACAAAAGTCAGTGTGCGTTATG	chromatin IP	qPCR
DC_h_chip_119	HS5-1 R-CBS forward	human	GGCCATTTCCTTTGTGTGGTGA	chromatin IP	qPCR
DC_h_chip_120	HS5-1 R-CBS reverse	human	GCCCTACCGGAAGCTGTCAAT	chromatin IP	qPCR

Supplemental Table 2: Table of the guide RNAs (gRNAs) used in this study

ID	Name	Organism	Sequence	Assay
a4_pCBS_1	gRNA for a4 sense promoter	human	AGTTTACAGTAGAGTGTG	dCas9-VPR activation
a4_pCBS_2	gRNA for a4 sense promoter	human	AACCATATACACTCTT	dCas9-VPR activation
a4_pCBS_3	gRNA for a4 sense promoter	human	CACAAAATACATGAGAGA	dCas9-VPR activation
a4_pCBS_4	gRNA for a4 sense promoter	human	CCACACTCAATCAATCAG	dCas9-VPR activation
a4_eCBS_1	gRNA for a4 antisense promoter	human	CGGCCTGGATTCCGCGA	dCas9-VPR activation
a4_eCBS_2	gRNA for a4 antisense promoter	human	CAGAGAAAAGTATTCATT	dCas9-VPR activation
a4_eCBS_3	gRNA for a4 antisense promoter	human	AACGGTGCCAGTCAACT	dCas9-VPR activation
a4_eCBS_4	gRNA for a4 antisense promoter	human	ATCCGAGGCGCCCTCTAG	dCas9-VPR activation
a6_pCBS_1	gRNA for a6 sense promoter	human	CGCCATCATTCAACAGTC	dCas9-VPR activation
a6_pCBS_2	gRNA for a6 sense promoter	human	ACATCATGCGGCTCCAAA	dCas9-VPR activation
a6_pCBS_3	gRNA for a6 sense promoter	human	GTGCAAATTACAAACGTG	dCas9-VPR activation
a6_pCBS_4	gRNA for a6 sense promoter	human	TAACTATTCTACACTCT	dCas9-VPR activation
a6_eCBS_1	gRNA for a6 antisense promoter	human	AGGAATAAGTTGTGTGC	dCas9-VPR activation
a6_eCBS_2	gRNA for a6 antisense promoter	human	GAGTTTATAGGTTAAGA	dCas9-VPR activation
a6_eCBS_3	gRNA for a6 antisense promoter	human	TGAGCCAACATCTGCAT	dCas9-VPR activation
a6_eCBS_4	gRNA for a6 antisense promoter	human	TCTACCGGGAACAAGGG	dCas9-VPR activation
a9_pCBS_1	gRNA for a9 sense promoter	human	GTCATACAAAACCTTTAG	dCas9-VPR activation
a9_pCBS_2	gRNA for a9 sense promoter	human	GCAGTCAGTCTGCTAAGA	dCas9-VPR activation
a9_eCBS_1	gRNA for a9 antisense promoter	human	GATTCTTTTGTGTCGCT	dCas9-VPR activation
a9_eCBS_2	gRNA for a9 antisense promoter	human	GAGCAATAAATGAAGCTC	dCas9-VPR activation
a12_pCBS_1	gRNA for a12 sense promoter	human	ATGATAAAGTGTAGGTTT	dCas9-VPR activation
a12_pCBS_2	gRNA for a12 sense promoter	human	AGAACCCTACACTTTATCA	dCas9-VPR activation
a12_pCBS_3	gRNA for a12 sense promoter	human	ACGATGGAGTGCTTGTGA	dCas9-VPR activation
a12_pCBS_4	gRNA for a12 sense promoter	human	CGGATCATTTATAAATGC	dCas9-VPR activation
a12_eCBS_1	gRNA for a12 antisense promoter	human	CATTTAGACTTAACGCAT	dCas9-VPR activation
a12_eCBS_2	gRNA for a12 antisense promoter	human	AGGCGCAGATTTCAGAAAC	dCas9-VPR activation
a12_eCBS_3	gRNA for a12 antisense promoter	human	ATCAGAAGCGCCCTCTAG	dCas9-VPR activation
a12_eCBS_4	gRNA for a12 antisense promoter	human	CAGCAATAAATTGAGTTT	dCas9-VPR activation
adel_1	gRNA used to generate SK-N-SH- α het	human	CTCGGATCTAGAGGATAG	Genome editing
adel_2	gRNA used to generate SK-N-SH- α het	human	GTGGTAAGATTGTTTGCT	Genome editing

Figure 2

A

B

C

D

E

A

B

Figure 5

A

B

A

B

A

B

Figure S2

B

SK-N-SH-iCas9

SK-N-SH- α het1

Figure S3

A

B

A

B

C

D