The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain

Yan Jiang1,10, Yong-Hwee Eddie Loh2,10, Prashanth Rajarajan1, Teruyoshi Hirayama3, Will Liao4, Bibi S Kassim1, Behnam Javidfar1, Brigham J Hartley4,9, Lisa Klefas1, Royce B Park1, Benoit Labonte2, Seok-Man Ho1, Sandhya Chandrasekaran1, Catherine Do5,6, Brianna R Ramirez2, Cyril J Peter1, Julia T C W2, Brian M Safaie1, Hirofumi Morishita1, Panos Roussos1,7,8, Eric J Nestler2, Anne Schaefer2, Benjamin Tycko5,6, Kristen J Brennand1, Takeshi Yagi3, Li Shen2 & Schahram Akbarian1

We report locus-specific disintegration of megabase-scale chromosomal conformations in brain after neuronal ablation of Setdb1 (also known as Kmt1e; encodes a histone H3 lysine 9 methyltransferase), including a large topologically associated 1.2-Mb domain conserved in humans and mice that encompasses >70 genes at the clustered protocadherin locus (hereafter referred to as cPcdh). The cPcdh topologically associated domain (TADcPcdh) in neurons from mutant mice showed abnormal accumulation of the transcriptional regulator and three-dimensional (3D) genome organizer CTCF at cryptic binding sites, in conjunction with DNA cytosine hypomethylation, histone hyperacetylation and upregulated expression. Genes encoding stochastically expressed protocadherins were transcribed by increased numbers of cortical neurons, indicating relaxation of single-cell constraint. SETDB1-dependent loop formations bypassed 0.2–1 Mb of linear genome and radiated from the TAD domain conserved in humans and mice that encompasses >70 genes at the clustered protocadherin locus (hereafter referred to as cPcdh). The cPcdh topologically associated domain (TADcPcdh) in neurons from mutant mice showed abnormal accumulation of the transcriptional regulator and three-dimensional (3D) genome organizer CTCF at cryptic binding sites, in conjunction with DNA cytosine hypomethylation, histone hyperacetylation and upregulated expression. Genes encoding stochastically expressed protocadherins were transcribed by increased numbers of cortical neurons, indicating relaxation of single-cell constraint. SETDB1-dependent loop formations bypassed 0.2–1 Mb of linear genome and radiated from the TADcPcdh fringes toward cis-regulatory sequences within the cPcdh locus, counterbalanced shorter-range facilitative promoter–enhancer contacts and carried loop-bound polymorphisms that were associated with genetic risk for schizophrenia. We show that the SETDB1 repressor complex, which involves multiple KRAB zinc finger proteins, shields neuronal genomes from excess CTCF binding and is critically required for structural maintenance of TADcPcdh.

A substantial portion of chromosomal material is compartmentalized into ‘topologically associated domains’ (TADs), which typically encompass several hundreds of kilobases of linear genome folded upon itself with regulatory proteins, including cohesin and the multifunctional CCCTC-binding factor (CTCF), to constrain expression of TAD-associated genes1–4. TAD structures do exist in brain5,6, and they have been studied in the context of risk haplotypes that are linked to neuropsychiatric disease by genome-wide association studies7. However, regulatory mechanisms remain unexplored. Here we report that neuronal maintenance of a subset of very large TADs (superTADs) critically requires SET domain bifurcated 1 (SETDB1; also known as ESET or KMT1E), a histone H3 Lys9 methyltransferase8. Setdb1 (Kmt1e) is important for prenatal development and pup survival9, and it broadly regulates reticulocyte suppression and transcriptional silencing in stem cells10–12. Little is known about its essential functions in differentiated cells, including neurons. By cell-type-specific 3D genome mappings and CTCF, DNA methylation and histone-modification profiling in conjunction with targeted epigenomic editing and conditional mutagenesis, we identified a SETDB1-dependent ‘shield’ that protects genomes from excessive CTCF binding. Loss of this shield uncovers unique locus-specific epigenomic vulnerabilities that trigger higher-order chromatin collapse on a megabase scale.

RESULTS

Locus-specific TAD disintegration in SETDB1-mutant neurons
To explore higher-order chromatin structures in a Setdb1-deficient brain, we generated a mouse line in which the Cre recombinase was expressed under the control of the calcium–calmodulin-dependent protein kinase II alpha subunit (Camk2a) gene promoter (Camk2a–Cre; hereafter referred to as CK–Cre) in postnatal forebrain neurons to drive a deletion in exon 3 of the loxP-flanked Setdb12lox/2lox allele, yielding a frameshift and premature stop upstream of the critical Tudor, methyl-CpG binding (MBD) and catalytic SET domains in the encoded SETDB1 protein. Brains from CK–Cre;Setdb12lox/2lox mutant mice showed normal gross cytoarchitecture, as compared to

1Friedman Brain Institute and Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA. 2Fishberg Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA. 3Kokoro-Biology Group, Laboratories for Integrated Biology, Graduate School of Frontier Biosciences, Osaka University, Suita, Japan. 4New York Genome Center, New York, New York, USA. 5Institute for Cancer Genetics, Department of Pathology and Cell Biology, Columbia University, New York, New York, USA. 6Herbert Irving Comprehensive Cancer Center, Columbia University, New York, New York, USA. 7Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA. 8Mental Illness Research, Education and Clinical Center, James J. Peters Virginia Medical Center, New York, New York, USA. 9Present address: VL39 Inc., Cambridge, Massachusetts, USA. 10These authors contributed equally to this work. Correspondence should be addressed to S.A. (schahram.akbarian@mssm.edu).

Received 3 February; accepted 5 June; published online 3 July 2017; doi:10.1038/ng.3906
Figure 1 3D genomes in Setdb1-deficient cortical neurons. (a) Left, Schematic representation of conditional Setdb1 ablation with loxP sites surrounding exon 3. Recombination results in a frame shift and generation of a premature stop codon (TGA) upstream of the Tudor, methyl-CpG-binding (MBD) and catalytic SET domains. Right, immunoblot for SETDB1 expression (histone H3 was used as a loading control) (top) and representative RNA-seq profiles (bottom) from the cortex of adult (MBD) and catalytic SET domains. Right, immunoblot for SETDB1 expression (histone H3 was used as a loading control) (top) and representative RNA-seq profiles (bottom) from the cortex of adult

that observed in Setdb1<sup>2lox/2lox</sup> control animals with wild-type levels of SETDB1. However, brains in the adult mutant mice were consistently smaller than those in adult control mice, but there was no premature cell death or neuronal loss in them, as assessed by flow-cytometry-based nuclei counts and COMET nuclear DNA damage assays (Fig. 1a and Supplementary Fig. 1). We conducted in situ Hi-C<sup>i</sup>, or genome-scale DNA–DNA proximity mapping in formalin-fixed, restriction-digested and re-ligated neuronal (NeuN<sup>+</sup>) nuclei that were collected by fluorescence-activated cell sorting to avoid signal contribution from the nuclei of non-neuronal cells) of the cortex from adult CK-Cre;Setdb1<sup>2lox/2lox</sup> and Setdb1<sup>2lox/2lox</sup> mice (Fig. 1b). However, chromosomal-contact mapping analyses at 40-kb resolution (n = 2 in situ Hi-C libraries/genotype, with 250–300 million aligned reads/library) showed that neurons from the mutant mice (hereafter referred to as mutant neurons) were not affected by a generalized disorganization of the 3D genome (Fig. 1b and Supplementary Fig. 2a). For example, length distributions and numbers of autosomal TADs, as assessed by TADtree<sup>1</sup>, were indistinguishable and minimally different between mice of the various genotypes, and the median length was ~200 kb, as expected for mammalian genomes<sup>1</sup> (Fig. 1c and Supplementary Fig. 2b). We then assessed longer-range chromosomal contacts, which spanned >200 kb of linear genome. On a genome-wide basis we identified 110 long-range loop contacts that were affected in the mutant neurons (hereafter referred to as mutant neurons) were not affected by a generalized disorganization of the 3D genome (Fig. 1b and Supplementary Fig. 2a). For example, length distributions and numbers of autosomal TADs, as assessed by TADtree<sup>1</sup>, were indistinguishable and minimally different between mice of the various genotypes, and the median length was ~200 kb, as expected for mammalian genomes<sup>1</sup> (Fig. 1c and Supplementary Fig. 2b). We then assessed longer-range chromosomal contacts, which spanned >200 kb of linear genome. On a genome-wide basis we identified 110 long-range loop contacts that were affected in the mutant neurons, using the DESeq2 package (P < 0.05; Online Methods)....
The large majority of these (84/110 or 76%) represented clustered, locus-specific ‘loop aggregates’ that showed massive weakening or complete loss after neuronal $Setdb1$ ablation (Fig. 1d and Supplementary Fig. 2c). These included a singular hotspot on chromosome 18 with substantial chromosome-wide enrichment (1-Mb sliding window (1Mb$^{sw}$); $P = 1.2 \times 10^{-24}$ by Poisson test; Fig. 1d,e) that fully encompassed the clustered protocadherin (cPcdh) locus, which harbors 77 genes, including 58 encoding cell adhesion molecules, linearly arranged as three gene clusters (Pcdh-$\alpha$, Pcdh-$\beta$ and Pcdh-$\gamma$, which encode the Pcdha, Pcdhb and Pcdhg genes, respectively) that regulate neuronal connectivity$^{13,14}$. Closer inspection of the WT cPcdh TAD structures showed multiple small (~100-kb length) cluster-specific subTADs nested into a massive superTAD that encompassed at least 1.2 Mb of linear genomic DNA. TADtree analyses confirmed that this superTAD was completely lost after $Setdb1$ ablation, leaving behind only subTAD remnants in mutant neurons (Fig. 1e and Supplementary Fig. 2d).

**SETDB1 shields neuronal genomes from excess CTCF occupancy**

We then asked how neuronal $Setdb1$ ablation could trigger such highly localized alterations in chromosomal conformations. To explore the role of $SETDB1$-regulated repressive histone methylation, we analyzed the product of $SETDB1$ activity, trimethylated Lys9 on histone H3 (H3K9me3), in NeuN$^+$ and NeuN$^-$ nuclei that were sorted from the cortex of brains from adult $CK-Cre;Setdb1^{2lox/2lox}$ mutant and $Setdb1^{2lox/2lox}$ control mice. DiffRep-based analysis with a 1-kb sliding window (1kb$^{sw}$) showed that 75% of 2,021 differentially H3K9me3-tagged sequences were hypomethylated in mutant neurons (Fig. 2a).

These deficits were specific, because chromatin immunoprecipitation and sequencing (ChIP-seq) profiling for open-chromatin-associated acetylated Lys27 on histone H3 (H3K27ac), in NeuN$^+$ and NeuN$^-$ nuclei that were sorted from the cortex of brains from adult $CK-Cre;Setdb1^{2lox/2lox}$ mutant and $Setdb1^{2lox/2lox}$ control mice. DiffRep-based analysis with a 1-kb sliding window (1kb$^{sw}$) showed that 75% of 2,021 differentially H3K9me3-tagged sequences were hypomethylated in mutant neurons (Fig. 2a).

**Fig. 2a and Supplementary Tables 1 and 2.** Furthermore, cPcdh emerged genome wide as a top-scoring locus for H3K9me3 hypomethylation (1Mb$^{sw}$: H3K9me3, 35-fold enrichment; observed/expected, 21/0.61; $P = 3.19 \times 10^{-25}$ by Poisson test) (Fig. 2b), with the densely concentrated H3K9me3 deficit readily visible in browser views of chromosome 18 (Fig. 2c). In contrast, nuclei from NeuN$^-$ cells that were sorted from the same cortical specimens were only minimally affected (Fig. 2a, Supplementary Fig. 3 and Supplementary Tables 3 and 4).

We then analyzed motifs in H3K9me3-hypomethylated sequences in $Setdb1$-deficient neurons. Three of the five top-scoring motifs matched to the transcriptional regulator and key 3D genome organizer CTCF and the CTCF paralog CTCF1 (also known as BORIS) (HOMER enrichment$^{17}$, $P < 10^{-60}$) (Fig. 2d and Supplementary Table 5). Furthermore, we analyzed published $SETDB1$ ChIP-seq data from stem cells and CD19$^+$ B lymphocytes and identified marked CTCF motif enrichment (Supplementary Fig. 4a and Supplementary Tables 6 and 7). Of note, other types of H3K9 methyltransferases, including EHMT1 and EHMT2 (also known as GLP and G9a, respectively) did not show any enrichment of CTCF motifs (Supplementary Fig. 4b).

We therefore predicted that there would be altered CTCF occupancy in the $Setdb1$-deficient neuronal genome. ChIP-seq analysis on NeuN$^+$ cells from the cortex of adult $CK-Cre;Setdb1^{2lox/2lox}$ and $Setdb1^{2lox/2lox}$ mice showed that 99.4% (3,059/3,078) of sequences with altered CTCF occupancy represented upregulated (including de novo) binding or ‘peaks’ (Fig. 2e and Supplementary Table 8), which included many promoters and enhancers (Supplementary Fig. 5). There was considerable over-representation for CTCF motifs (HOMER enrichment, $P < 10^{-1,000}$) (Fig. 2f and Supplementary Table 9) that were independent of the filtering conditions (Supplementary Fig. 6 and Supplementary Table 10), and these were present in cis-regulatory elements and H3K9me3-hypomethylated sequences (Supplementary Table 11). Therefore, $SETDB1$ shields mature neuronal genomes from excess CTCF occupancy at cryptic binding sites. Of note, cPcdh again emerged as a top-scoring locus in the genome-wide analysis (1Mb$^{sw}$: CTCF NeuN$^+$ upregulated peaks, 18.7-fold enrichment; $P = 1.32 \times 10^{-21}$ by Poisson test; Fig. 2g).

Additional localized enrichments of excess CTCFoccupancies matched to loci on chromosomes 5 and 7 that were affected by the loss of long-range chromosomal contacts and H3K9 hypomethylation (Fig. 2g).

Given that CTCF, a key regulator of higher-order chromatin structures including domain insulation$^{19}$, was upregulated at thousands of positions in the $Setdb1$-deficient neuronal genome, we assessed genome-wide domain insulation in the *in situ* Hi-C data sets from our mutant and wild-type cortical neurons. We first focused on CTCF-specific de novo peaks, which were filtered for their proximity to the TAD boundary (20% of total TAD length) and for the vicinity (±100 kb) of sequences with altered H3K9me3 levels after $Setdb1$ ablation. Of these, 52–57% of de novo CTCF-specific peaks showed stronger insulation scores in mutants across 8/9 insulation bands, covering 80–1,040 kb of contact distance (Supplementary Fig. 7). At sites with conserved CTCF peaks, insulation scores showed very minimal differences between mutant and control neurons (Supplementary Fig. 8). Therefore, an excess of CTCF on a genome-wide scale conveys a subtle shift toward increased insulation strength in mutant neurons, with the notable exception of $Setdb1$-sensitive superTADs, which are affected by structural collapse and loss of insulation. Our findings, in conjunction with recent genome-scale studies reporting the loss of domain insulation in glioma cells due to decreased CTCF binding$^{17}$, suggest that spatial architectures of chromosomes are highly sensitive to bidirectional changes in CTCF occupancies. Next we explored alterations in the A and B compartments, which are defined as multi-megabase chromosomal segments representing ‘A’ open chromatin and ‘B’ condensed chromatin that tend to interact with other loci sharing similar levels of chromatin accessibility$^{2}$. Because the A and B compartments are defined on a continuum$^{2}$ (as opposed to a biphasic signal), we quantified ‘compartment-ness’ from the intrachromosomal contact matrices generated by Hi-C-Pro program (Online Methods). Of note, the total number of A- and B-specific compartment bins was only minimally different between genotypes. However, 6,032/11,048 (54%) of A and 8,468/12,977 (65%) of B bins had higher ‘compartment-ness’ scores in mutant neurons than in WT neurons (Fishier’s exact $P < 10^{-210}$) (Supplementary Fig. 9a). However, $SETDB1$-sensitive superTADs did not follow this genomewide trend, as exemplified by the weakened ‘B’ signal at the cPcdh locus in mutant neurons (Supplementary Fig. 9b).
We then asked which molecular mechanisms contributed to the excess amount of CTCF at the hypomethylated H3K9me3 sites. Of note, the majority of cPcdh sequences affected showed coordinate increases in CTCF binding and histone hyperacetylation (Fig. 2h), suggesting a shift toward open (permissive) chromatin states. To this end, alterations in cytosine methylation in DNA—which reduce CTCF’s affinity for DNA—via interaction with the seventh of CTCF’s 11 zinc fingers—could play a key role, as SETDB1 functions as an upstream regulator for DNA methylation. To explore this, we used bisulfite sequencing (Bis-seq) to quantify levels of 5-methylcytosine (mC5) with 43 PCR amplicons that targeted 13 cPcdh sites. We assayed cortical and striatal NeuN+ and NeuN− nuclei from forebrains, plus cerebellar tissue as an additional control, from adult CK−;Cre;Setdb1Δlox/Δlox and Setdb1Δlox/Δlox mice, which altogether comprised 46 individual samples (Supplementary Table 12). As expected, mC5 levels in the non-neuronal forebrain nuclei and in the cerebellum remained unaltered; however, in Setdb1-deficient neurons, mC5 significantly increased in neuronal nuclei (Fig. 3a, b). This is consistent with previous findings in other contexts, including gene deserts, the enhancerome, and non-coding RNA genes (Fig. 3). Taken together, our findings suggest that CTCF binding is reciprocally regulated; CTCF occupancy is reduced in hypermethylated regions, whereas DNA methylation levels are lower in CTCF-enriched areas (Fig. 3). The mechanism behind this reciprocal regulation is likely to be complex, involving multiple molecules, including SETDB1. To shed light on this, we performed ChIP-seq experiments to determine whether mC5 enrichment is accompanied by changes in CTCF binding (Fig. 3c, d). We found that CTCF binding is reduced at hypomethylated H3K9me3 sites, whereas mC5 is increased at these sites (Fig. 3e, g). This suggests that CTCF binding and DNA methylation are inversely regulated (Fig. 3). Therefore, excess amounts of CTCF occupany in Setdb1-deficient neuronal genomes is associated with conversion to an ‘open’ chromatin state at cryptic CTCF-binding sites. This includes reduced DNA methylation levels and weakening of regulatory mechanisms designed to prevent excess CTCF binding.

**TAD-specific regulation of gene expression**

To explore whether disintegration of superTADs affects gene expression, we mapped transriptomes and neuronal H3K27ac in the cortex of CK−;Cre;Setdb1Δlox/Δlox and Setdb1Δlox/Δlox mice. Consistent with SETDB1’s repressor function, the majority of transcripts that
were altered in the mutant (208/321) showed an upregulation of expression (Supplementary Table 14). Notably, 20% of the entire pool of Setdb1-sensitive genes were located in the Pcdh locus and affected both protocadherins and non-protocadherins, resulting in a unique, 543-fold enrichment for upregulated transcripts at this locus, affecting both protocadherins and non-protocadherins, resulting in a specific transcriptional control across the cPcdh domain specifically in mature neurons. However, this regulatory layer is not specific for other types of H3K9 methyltransferase, or for the prenatal Setdb1-deficient brain.

Of note, 31/53 stochastically expressed (S-type) isoforms of protocadherins from the Pcdh-α, Pcdh-β and Pcdh-γ gene clusters, which are critically important for neuronal diversity and connectivity, showed increased expression after neuronal Setdb1 ablation (Fig. 4b,c). We studied S-type protocadherin expression, which is defined by random expression in a very small subset of neurons dispersed within each specific brain region, by in situ hybridization...
at single-cell resolution using probes specific for the individual S-type-protocadherin-encoding genes Pcdha1, Pcdha8, Pcdhb22 and Pcdhga7. Brain sections from CK–Cre;Setdb1Cre–/- conditional mutants but not from any of the three control genotypes, including the transgenic rescue line, showed massively increased numbers of robustly stained neurons that were diffusely distributed across cortical layers II–VI and the hippocampus (Fig. 4c). The cerebellar cortex, which in contrast to the forebrain lacks CK–Cre expression, remained unaffected in the brains of the conditional mutant mice (Fig. 4c and Supplementary Figs. 12–15). Therefore, S-type single-neuron
stochastic constraint is severely compromised in Setdb1-deficient neurons, which contributes to upregulated expression at the cPcdh locus. Given the critical importance of orderly cPcdh expression—including single-cell stochastic constraint of S-type Pcdh genes—for neuronal morphology and connectivity,13,33–35, we quantified spine densities and diameters from layer III apical dendrites from Setdb1 WT/WT and Setdb1 2lox/2lox mice that were crossed into a conditional line expressing membrane-bound GFP (GFP-F)16 for Golgi-like labeling after virus vector (AAVg9SNY1-CreGFP)25-mediated Cre recombination to adult prefrontal cortex (PFC). Indeed, spines from Setdb1-deficient neurons showed 40–50% increased density and overall decreased size (Supplementary Fig. 1g), which provided a morphological correlate for dysregulated cPcdh expression.

Balanced facilitative and repressive conformations at cPcdh

Next we wanted to gain deeper mechanistic insight into the molecular mechanisms that mediate the unique position of the cPcdh locus within the SETDB1-sensitive transcriptional and epigenome space. Because CTCF associates with RNA polymerase subunits37 and transcriptional activators38,39, the observed increase in CTCF occupancy at H3K27ac hyperacetylated S-type Pcdh-α, Pcdh-β and Pcdh-γ gene promoters in Setdb1-deficient neurons could facilitate expression, including loss of single-cell stochastic constraint. However, promoter-bound CTCF alone was not sufficient to upregulate transcription. From 63 genes with an excessive presence of CTCF near transcription start sites genome wide, only transcripts within the cPcdh locus showed increased expression (Supplementary Tables 8 and 14). Of note, promoter–enhancer loopings that are mediated by the CTCF–cohesin scaffolding complex contribute to transcriptional regulation of cPcdh genes20, and therefore, excess CTCF occupancy in cPcdh sequences from Setdb1-deficient neurons could trigger alterations in higher-order chromatin structure. Indeed, excessive CTCF binding at the cPcdh locus in Setdb1-mutant neurones was not limited to promoters, because multiple new CTCF peaks emerged in intergenic DNA upstream from the Pcdh-α cluster and within the Pcdh-γ cluster (peaks A–C in Fig. 2h). Notably, these de novo peaks were surrounded by broad (>100–200 kb) stretches of H3K9me3-tagged chromatin that underwent substantial ‘shrinkage’ after neuronal Setdb1 ablation (labeled ‘R1’ and ‘R2’ in Fig. 5a). Of note, R1 and R2 marked the anchor regions of massive blocks of long-range chromosomal contactings in wild-type neurons. Thus, densely spaced H3K9me3-tagged R1 loopings, which emanated from 100–200-kb-wide blocks of repressive chromatin upstream of the Pcdh-α genes, radiated toward many sites within cPcdh and even reached the distal-most Pcdh-γ sequences. However, these long-range loopings became completely dissolved after structural disintegration of the superTAD Pcdh (Fig. 1e). Among the H3K9me3-tagged contactings lost after neuronal Setdb1 deletion were multiple loopings that interconnected R1 and R2 with two DNase-I-hypersensitive enhancer elements, HS16 and HS5–1, which have been previously shown to broadly facilitate cPcdh expression23–29. These defects in HS16- and HS5-1-bound long-range contactings were highly specific, because mutant neurones fully maintained shorter-range loopings from protocadherin gene promoters to HS16 and HS5-1-enhancers within the subTADs (Fig. 1e). We confirmed these Hi-C findings, including specific weakening of long-range R1–HS16, R1–R2 and R2–HS5–1 interactions and preservation of shorter-range contactings, in neuron-specific chromosome conformation capture (3C)-PCR assays from the cortex of adult mutant and control mice (Fig. 5b and Supplementary Fig. 16a). These studies, taken together, would suggest that in neurones from wild-type mice, HS16 and HS5-1 enhancer sequences are ‘locked’ into H3K9me3-tagged repressive chromatin. After neuronal deletion of Setdb1, loss of R1–R2 repressive loop formations could release the ‘epigenomic brake’, thereby shifting the balance from repressive toward facilitative contactings furnished by HS16- and HS5–1-bound promoter–enhancer loopings, thus triggering increased expression across the cPcdh locus (Fig. 5c). To test this hypothesis, we transfected mouse NG108 neuroblastoma cells with small-RNA-guided Cas9–SunTag protein scaffolds40 that were designed to load ten copies of the potent transcriptional activator, VP64, onto a single HS16 site (hereafter referred to as HS16 Cas9–SunTag(10×VP64)) (Fig. 5d). Therefore, such type of ‘epigenomic superactivation’ could, like the loss of R1–HS16 and R2–HS5–1 repressive loopings after Setdb1 ablation, increase transcription at multiple positions across the entire 1-Mb cPcdh locus via promoter–enhancer contacts and other mechanism. Cells epigenomically edited at the HS16 position in the cPcdh locus (HS16 Cas9–SunTag(10×VP64)) were compared to controls expressing exactly the same types of vectors but without the small guide RNAs (sgRNAs) (Fig. 5d). Indeed, HS16 epigenomic superactivation was associated with increased expression of three of six cPcdh transcripts (pre-selected for consistent baseline expression in neuroblastoma cells), which closely mimicked the transcriptional phenotype observed in the cortex of Setdb1-mutant mice (Fig. 5e).

Conserved regulation of human and mouse superTAD cPCDH

The linear arrangement of S- and C-type protocadherin genes in the Pcdh-α, Pcdh-β and Pcdh-γ clusters is highly conserved across vertebrate genomes15. We showed that higher-order chromatin structures, including broad >100- to 200-kb stretches of intergenic SETDB1-sensitive H3K9me3-tagged sequence associated with repressive loop bundles, critically regulate transcription across the cPcdh locus. We asked whether such types of 3D genome contactings, just like the linear genome, could be conserved across mammalian lineages. To examine this, we generated in situ Hi-C interaction matrices in human glutamatergic neurones that were differentiated from induced pluripotent stem cell (iPSC)-derived neural precursors by controlled expression of neurogenin 2, and we compared the 3D genome map to that in NeuN+ nuclei in the cortex of wild-type mice. Indeed, TAD landscapes surrounding the cPcdh locus in mice (mm10 chromosome 18) and the orthologous PCDH locus in humans (hg19 chromosome 5) showed startling similarities between the NeuN+ nuclei in the cortex of mice and the human neurones, including complete preservation of cluster–specific subTADs that were nested in a large megabase-scale superTAD cPCDH. In addition, human and mouse neuronal chromatin showed highly similarly shaped H3K9me3 landscapes, including the broadly stretched previously mentioned SETDB1-sensitive R1 at the superTAD’s 5’ end and R2 near the 5’ end of the Pcdh-γ cluster (Fig. 6a). Of note, R1 near-perfectly matched a risk haplotype (chr5: 140,023,664–140,222,664) of the Psychiatric Genomics Consortium41. This haplotype (number 108 in ref. 41, hereafter referred to as PGC) significantly contributes, independently from another 107 loci genome wide, to schizophrenia heritability41, with a small insertion–deletion (indel) being the lead polymorphism (rs11896713, chr5: 140,143,664; ref. 41). This risk-associated polymorphism matched robust SETDB1 peaks that were conserved in human and mouse cells including those in brain, but were ‘replaced’ by de novo CTCF peaks after Setdb1 ablation (Fig. 6a and Supplementary Fig. 16c,d). Therefore, we predicted that higher-order chromatin organization at these positions would be highly conserved in human neurones, with long-range loopings radiating from ~200 kb R1 toward the cPCDH promoter and enhancers, anchored in chromatin at and near the SETDB1 peak. To explore this, we surveyed (40-kb resolution) the cPCDH-bound chromosomal contacts in our
Figure 5 Epigenomic editing at the cPcdh locus. (a) Schematic representation of ~2 Mb of mouse chromosome 18, showing the cPcdh locus and surrounding sequences. Also shown are the TADs called (in TADtree) and H3K9me3 tracks for neuronal nuclei in cortex from WT and KO mice.

(b) Cell-type-specific 3C-PCR analysis, with cropped gels showing specific loop products for cPcdh locus and the β2-microglobulin (B2m) control, and graphs summarizing the 3C-PCR cPcdh loops 1–3. Graphs summarizing 3C-PCR. Each circle represents an individual mouse, and data were normalized to the B2m 3C-PCR analysis. *P<0.05, Mann–Whitney, one-tailed; **P<0.014, Mann–Whitney, two-tailed. Loop defects in KO include A/R1–HS16 (de novo CTCF peak A in R1) and A/R1–B/R2 (de novo CTCF peak B in R2). Shorter-range Pcdha8 promoter–HS5 enhancer loop is maintained in KO neurons. No lig, 3C without DNA ligase; L, 100-bp DNA ladder; CX, cortex. Uncropped gels are shown in Supplementary Figure 16b.

(c) Summary presentation of 3C-PCR. (d) dCas9–SunTag superactivation of the HS16 cPcdh enhancer with a U6-sgRNA cassette upstream of the CK-dCas9-10xGCN4 tetO-BFP cassette, with the CK-sFv-sGFP-VP64 cassette on a separate vector. Representative FACS sort shows dually labeled BFP+GFP+ NG108 cells. NC, negative control. (e) RT–PCR quantification of Pcdha3, Pcdha8, Pcdhb16, Pcdhgb2 and Pcdhgb8 transcripts (arrows in red mark genomic positions), normalized to Gapdh RNA. Top, BFP+GFP+ NG108 cells with (HS16–VP64) and without (VP64) the sgRNAHS16 cassette. Bottom, expression data for adult KO and WT PFC. Each circle represents an individual mouse or cell culture. n = 4 control and n = 3 HS16–VP64 cultures, *P<0.05, Mann–Whitney, two-tailed; **P<0.01, Mann–Whitney, two-tailed. See Supplementary Figure 16a for additional quantifications.
Figure 6  Regulatory mechanisms at human and mouse TAD\textsuperscript{pCDH}. (a) Top, neuronal in situ Hi-C interaction matrices, and H3K9me3 landscapes, for ~2 Mb of mouse and human pCDH, including the superTAD that spans across the α, β and γ clusters. Bottom left, SETDB1 peaks in mouse embryonic stem cells and lymphocytes match to de novo CTCF peaks in Setdb1-ablated neurons. Bottom right, PGC risk haplotype chr5: 140,023,664–140,222,664 with lead polymorphism rs111896713, which matches to the SETDB1, KAP1 and ZNF143 peaks. Note the epigenomic similarities between ‘R1’ (mouse) and ‘PGC’ (human). (b) Top, representative images (n = 2 cell lines) for NPC differentiation into neurons and astrocytes, with phenotypic markers as indicated. Scale bars, NPC 100 μm (NPC), 50 μm (neurons, astrocytes). Bottom, conformations for three representative 40-kb bins from 200 kb of the PGC haplotype, with the bin harboring the index polymorphism (PGC-3) showing increased pCDH contact. (c) pCDH gene expression in epigenomically edited NPCs, with PCDHG6 (but not PCDHG8B) transcript decreased by sgRNA-guided dCas9-KRAB in 3/3 experiments. dCas9-VP64 elicited increased expression of a subset of protocadherin transcripts. Scr, scrambled control. All data were normalized to 18S rRNA levels and are shown as fold change. For dCas9–VP64, n = 5/scrambled control, n = 6 for each PGC2 and PGC3 target site. For dCas9–KRAB, n = 7/scrambled control, n = 8 for each PGC2 and PGC3 target site. P < 0.05, Mann–Whitney, two-tailed. (d) ZNF-specific motif enrichments in CTCF–specific upregulated sequences. Graphs (1 circle/cell culture) summarize expression of specific Pcdha, Pcdhb and Pcdhg genes after shRNA-induced Zfp143 knockdown in NG108 neuroblastoma cells. Unpaired t-test, two-tailed; n = 3 per treatment, *P < 0.05. (e) Schematic summary of TAD\textsuperscript{pCDH} epigenomic architectures in WT and KO neurons. Loss of repressive long-range contacts in KO shifts the balance toward facilitative shorter-range promoter–enhancer loopings.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.
in situ Hi-C data sets that were generated from human neurons and their isogenic neural precursors cells (NPCs) and NPC-differentiated astrocytes. 40-kb bins within PGC showed a step-wise progression in contact intensities with cPCDH sequences, which culminated in massively increased contact frequencies at the bin harboring a robust SETDB1 peak (Pgc-3 in Fig. 6b). This effect was pronounced in neurons and NPCs, whereas the corresponding loopings were much weaker or missing altogether in our contact maps from astrocytes, indicating that there was strong cell-type-specific regulation of local 3D genome architectures (Fig. 6b). Next we asked whether these within-PGC haplotype differences in cPCDH interaction frequencies translated into differential repressive potential. To this end, we introduced sgRNAs into two stable NPC lines, one expressing a dCas9–KRAB fusion protein tethering a Kap1 (KRAB-associated protein 1)-SETDB1 repressor complex8,42, and the other expressing dCas9–VP64 to dock the VP64 activator at different positions within the haplotype-associated with increased risk of schizophrenia. Then measured expression levels for the genes in the S-type Pcdh-γ clusters that were expressed in NPCs at comparatively high levels at baseline (data not shown). Notably, the KRAB that was recruited to sequences close to the SETDB1 peak at the risk haplotype’s lead polymorphism (Pgc-3 in Fig. 6c) was consistently associated in 3/3 experiments with a robust multifold decrease in expression of Pcdhgb6 (but not Pcdhga3). In contrast, dCas9–KRAB docked to a non-SETDB1 binding site (Pgc-2 in Fig. 6c) or to a scrambled control sequence (‘Scr’ in Fig. 6c) remained ineffective and did not suppress Pcdhgb6 and Pcdhga3 expression (Fig. 6c). Of note, VP64 epigenomic editing at Pgc-3 and the neighboring Pgc-2 was associated with increased expression of a subset of cPcdh genes (Fig. 6c). These findings, taken together, suggest that repressive effects on protocadherin gene expression are specific for loop-bound KRAB that is positioned at intergenic Pgc-3 sequences upstream of the cPcdh gene clusters.

This shows that KRAB—a critical module in KRAB zinc finger proteins (KRAB-ZNF) that is important for sequence-specific docking of the Kap1–SETDB1 repressor complex8,42—inhibits cPcdh expression via long-range loopings, bypassing 644 kb of linear genome in the case of Pcdhgb6 (Fig. 6c). Therefore, intergenic R1–risk haplotype-bound SETDB1 is likely to function as a key transcriptional regulator at the cPcdh locus. These intergenic sequences harbor mating peaks for SETDB1, Kap1 and multiple KRAB-ZNF proteins in the ENCODE database, including Znf274 (ref. 43) and Zfp143 (Fig. 6a). Of note, ZFP143 recognition sequences emerged as top-scoring zinc finger motifs that were enriched at sites with excessive CTCF binding in Setdb1-deficient mouse neurons (Fig. 6d and Supplementary Tables 9 and 10). Znf143, considered to be a key organizer for the 3D genome44,45 similar to CTCF and cohesin, co-assembles with positive and negative regulators of transcription depending on local chromatin context46. Thus, ZNF143 at the cPcdh locus occupies not only R1 repressive chromatin but also promoters and the HS16 and HS5-1 enhancers (Fig. 6a). Therefore, alterations in ZNF143 supply, which affect facilitative and repressive chromatin, could destabilize cPcdh gene expression. Indeed, small-RNA-mediated Zfp143 knockdown in mouse neuroblastoma cells was associated with decreased expression of multiple cPcdh genes (Fig. 6d). Taken together, our studies suggest that (i) regulatory 3D genome architectures at the cPcdh locus are highly conserved between mouse and human, (ii) SETDB1–KRAB–ZNF143 and CTCF are key organizers of local repressive and facilitative chromosomal conformations, and (iii) ‘bundles’ or ‘aggregates’ of SETDB1-dependent long-range repressive loopings radiate from intergenic DNA (R1 and R2) and function as ‘epigenomic brakes’ for transcriptional control, counterbalancing facilitative shorter-range promoter–enhancer contacts (Fig. 6e).

DISCUSSION

Neuronal Setdb1 ablation triggers structural disintegration of mega-base-scale TADs, including those at the Pcdhα-, Pcdhβ- and Pcdhγ-containing locus as the only SETDB1-sensitive TAD that harbors a gene cluster. TADs that were affected in mutant neurons showed shrinkage of broadly stretched H3K9me3-tagged chromatin, in conjunction with localized hotspots of excess and de novo CTCF binding. SETDB1-regulated long-range repressive cPcdh loopings were highly enriched in neurons, as compared to those in their isogenic precursors, and carried DNA polymorphisms conferring liability for schizophrenia. 3D genome conformations at the cPcdh locus could have even broader relevance for neuropsychiatric disease, given that micro-deletions in Setdb1 and structural variants in the protein are associated with neurodevelopmental delay47,48, with CpG hypermethylation having been reported for orthologous CTCF binding sites within the Pcdh gene cluster in Down syndrome (trisomy 21), including in the mouse model 49, and that cPcdh DNA promoter methylation has been linked to depression and anxiety50–52. Furthermore, mice that are exposed to chronic variable stress, a preclinical paradigm frequently implied in psychiatric disease, show hyperexpression of Pcdh genes53 (Supplementary Fig. 17).

We show that SETDB1, which maintains high levels of DNA methylation and low levels of histone acetylation at sequences in close vicinity or that partially overlap with potential CTCF binding sites, critically shields neuronal genomes from uncontrolled CTCF docking at thousands of cryptic binding sites genome wide. However, after neuronal Setdb1 ablation, the shield becomes deficient, triggering collapse of vulnerable TADs. Our findings on excessive cPcdh CTCF occupancies and increased cPcdh gene expression, and the resulting increase in spine densities, in Setdb1-deficient neurons are contrary to the previously reported decreases in cPcdh expression and spine densities after neuronal Ctf1 ablation54. Likewise, switching ‘reverse-forward’ strand orientations in CTCF binding sequences55 could disrupt promoter–enhancer loopings and broadly dampen transcription across the genes in the Pcdh-α, Pcdh-β and Pcdh-γ clusters (Fig. 2d in ref. 25). These findings strongly point to delicate regulatory mechanisms governing chromosomal conformations, with genomes excessively populated by CTCF, including the Setdb1-deficient neurons in our study, showing disintegration of higher-order chromatin in a locus-specific manner. The structural collapse of the Setdb1-sensitive superTADs was highly specific, given that the genome-wide excess of CTCF binding in mutant neurons triggered a genome-wide increase in insulation strength and compartment-ness.

Future work will clarify whether neuronal Setdb1 overexpression30,55, or loss of other proteins assigned with regulation of cPcdh gene expression, including the Dnmt3b cytosine methyltransferase34 and the Smcdh1 (ref. 56) and Wiz257 repressors, could trigger TAD-specific 3D genome changes in neurons. We note that SETDB1 is primarily located toward the 5’ and 3’ ends of superTADs (Fig. 6a,e and Supplementary Fig. 18). In any case, the TAD-specific phenotypes in Setdb1-mutant neurons point toward unexplored modular complexities in the regulatory mechanisms that govern the 3D genome. Thus, rewiring or disintegration of specific TAD units may not be exclusive to chromosomal microdeletion and duplication events58,59, because as shown here, loss of SETDB1 function triggers the disintegration of highly select subset of neuronal TADs. With each chromosome furnishing hundreds of TAD-like structures, it will be an exciting and challenging task to dissect ‘TAD by TAD’ and in a cell-type-specific
manner, the multilayered mechanisms governing locus-specific higher-order chromatin in highly differentiated brain cells.

URLs. Optimized CRISPR Design tool, http://crispr.mit.edu.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank J. Gonzalez-Maeso (Virginia Commonwealth University) for kindly providing the Camk2α promoter plasmid. The work was supported by US National Institutes of Health (NIH) grants R01MH065065 (S.A.), P50MH096890 (E.I.N.), R01MH101454 (K.J.B.), NIA U0154005318-30 (K.J.B.), U01AG046170 (K.J.B.), R01AG059086 (P. Roussos), R01MH199677 (P. Roussos) and R01NS091374 (A.S.), NIH training grant T32 AG049688 (S.C.) and NIH fellowship award 1F30MH13330 (P. Rajarajan). Additional support was provided by a Grant-in-Aid for AMED-CREST, AMED, Japan (T.Y.), the Japan–US Brain Research Cooperation Program (T.Y.), the Veterans Affairs Merit Grant BX002395 (P. Roussos), the Brain and Behavior Research Foundation (Y.J. and P. Roussos), the Alzheimer’s Association (P. Roussos), the New York Stem Cell Foundation (K.J.B.) and the Brain Research Foundation (S.A.).

AUTHOR CONTRIBUTIONS

Y.J., P. Rajarajan, T.H., B.S.K., B.J.H., S.-M.H., B.J., L.K., R.B.P., S.C.D., C.D.J.P., J.T.C.W. and B.M.S. performed experiments; Y.J. and S.A. conceived and designed experiments; Y.J. performed statistical analyses; Y.-H.E.L., P. Rajarajan, W.L., P. Roussos and I.S.P. performed bioinformatics and genomic analyses; A.S., B.R.R., G9a and GLP transcriptome data; B.L. and E.J.N. (mouse stress model and P. Roussos) performed bioinformatics and genomic analyses; A.S., B.R.R., Camk2a et al. wrote the paper with contributions from the other co-authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.
47. Xu, Q. et al. Chromosomal microarray analysis in clinical evaluation of neurodevelopmental disorders—reporting a novel deletion of SETDB1 and illustration of counseling challenge. 
Pediatr. Res. 80, 371–381 (2016).

48. Cukier, H.N. et al. The expanding role of MBD genes in autism: identification of a MECP2 duplication and novel alterations in MBD5, MBD6 and SETDB1. 
Autism Res. 5, 385–397 (2012).

49. Mendioroz, M. et al. Trans effects of chromosome aneuploidies on DNA methylation patterns in human Down syndrome and mouse models. 
Genome Biol. 16, 263 (2015).

50. Garafola, C.S. & Henn, F.A. A change in hippocampal protocadherin-γ expression in a learned helpless rat. 
Brain Res. 1593, 55–64 (2014).

51. Suderman, M. et al. Conserved epigenetic sensitivity to early-life experience in the rat and human hippocampus. 
Proc. Natl. Acad. Sci. USA 109 (Suppl. 2), 17266–17272 (2012).

52. McGowan, P.O. et al. Broad epigenetic signature of maternal care in the brain of adult rats. 
PLoS One 6, e14739 (2011).

53. Labonté, B. et al. Sex-specific transcriptional signatures in human depression. 
Nat. Med. (in the press).

54. Hirayama, T., Tarusawa, E., Yoshimura, Y., Galjart, N. & Yagi, T. CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. 
Cell Rep. 2, 345–357 (2012).

55. Bharadwaj, R. et al. Conserved higher-order chromatin regulates NMDA receptor gene expression and cognition. 
Neurot. 84, 997–1008 (2014).

56. Chen, K. et al. Genome-wide binding and mechanistic analyses of Smchd1-mediated epigenetic regulation. 
Proc. Natl. Acad. Sci. USA 112, E3535–E3544 (2015).

57. Isbel, L. et al. Wiz binds active promoters and CTCF-binding sites and is required for normal behavior in the mouse. 
eLife 5, e15082 (2016).

58. Franke, M. et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. 
Nature 538, 266–269 (2016).

59. Redin, C. et al. The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. 
Nat. Genet. 49, 36–45 (2017).
ONLINE METHODS

Human stem cell lines. All work with human induced pluripotent stem cell lines (hiPSCs) has been approved by the Institutional Review Board of the Mount Sinai School of Medicine, in accordance with Mount Sinai’s Federal Wide Assurances (FWA#00005656, FWA#00005651) to the Department of Health and Human Services. No new stem cell lines were generated for the work presented here. Informed consent was obtained from all participating subjects. See Supplementary Note for protocols for differentiation of hiPSCs into neural progenitors, glutamatergic neurons and astrocytes.

Mouse studies. All mouse work was approved by the Institutional Animal Care and Use Committee of the Icahn School of Medicine at Mount Sinai. Mice were held under specific pathogen-free conditions with food and water being supplied ad libitum in an animal facility with a reversed 12-h light–dark cycle (lights off at 7:00 a.m.) under constant conditions (21 ± 1°C; 60% humidity). All mice were group-housed (2–5 mice/cage).

Generation of Setdb1 conditional mutant and rescue mice. Setdb1±/lox mice were generated by Orgene, Australia. In brief, two loxP sites were inserted into the endogenous Setdb1 locus (Setdb1ENSMUSG0000015697; SET domain, bifaceted 1, MGI: 1934229) flanking exon 3. To generate conditional Setdb1-knockout mice, Setdb1±lox/lox mice were first crossed with CK-Cre+ transgenic mice to generate Setdb1±lox+/+CK-Cre+ heterozygous mice, which were further crossed with Setdb1±lox/lox mice to generate CK–Cre°/°;Setdb1±lox/lox homozygous mice. Cre-mediated excision of Setdb1 exon 3 caused a frame shift event and generated a stop codon at the new junction of exons 2 and 4, which resulted in early termination in Setdb1 translation. Gender- and age-matched litters with the genotype CK–Cre°/°;Setdb1±lox/lox were used as controls for WT SETDB1 levels. CK–Setdb1 transgenic mice, described previously30, express full-length mouse Setdb1 cDNA driven by the CK promoter in postnatal and adult mouse forebrain. To generate Setdb1 rescue mice, the CK–Setdb1 transgene was introduced into the Setdb1±lox/lox conditional-knockout background and CK–Setdb1°/°;Setdb1±lox/lox mice were crossed with CK–Cre°/°;Setdb1±lox/lox mice to generate CK–Cre°/°;Setdb1±lox/lox–CK–Setdb1°/° rescue mice. Gender- and age-matched litters with the genotype CK–Cre°/°;CK–Setdb1°/°;Setdb1±lox/lox were used as WT controls. All genetically engineered lines were backcrossed to the C57BL/6J line for at least ten generations.

See the Supplementary Note for information on nestin–Cre conditional mutagenesis, mendelian survival ratios, histology, comet assays and RNA quantifications, including RNA-seq analysis.

Chromatin assays. Chromatin assays (ChIP-seq, in situ Hi-C and 3C-PCR) and RNA-seq were conducted in young adult mice at 3 months of age (±2 weeks). Supplementary Tables 1–4, 8 and 12 provide additional information for each chromatin assay, including number of animals and sex ratios.

Preparation of nuclei, immunotagging and fluorescence-activated sorting. For fluorescence-activated sorting of nuclei, nuclei were extracted from mouse cerebral cortex or human prefrontal cerebral cortex (control, PFC, male, post-mortem interval (PMI) 17 h) and anterior cingulate cortex (control, ACC, female, PMI 27 h) as described previously60. In brief, brain tissue was homogenized in hypotonic lysis solution, purified by ultracentrifugation and then resuspended in 1 ml Dulbecco’s phosphate buffered saline (DPBS) containing 0.1% BSA and 1:000 anti-NeuN antibody (clone A60, Alexa Fluor 488 conjugated; EMD Millipore Corp., MAB377X). Samples were incubated for at least 45 min by rotation in the dark at 4 °C. DAPI was added before FACS to label all of the nuclei. Sorting was done at the Flow Cytometry Center at Mount Sinai. Nuclei were separated into NeuN+ and NeuN− populations and then pelleted for the following applications. For XChIP 3C and in situ Hi-C experiments, immediately after brain homogenization the tissue was fixed in 1% formalin for 10 min at room temperature. The cross-linking was quenched by incubation with 125 mM glycine. Nuclei were then purified, stained, and sorted as described above.

Chromatin immunoprecipitation and sequencing (ChIP-seq) analysis. Native immunoprecipitation (NChIP) was performed as previously described60. In brief, NeuN+(neuronal) and NeuN−(non-neuronal) nuclei were pelleted after FACS and then resuspended in 300 μl of micrococcal nuclease digestion buffer (10 mm Tris, pH 7.5, 4 mm MgCl2, and 1 mM CaCl2) and digested with 3 μl of MNano (0.2 μl) for 5 min at 28 °C to obtain mononucleosomes. The reaction was stopped with 50 mM EDTA, pH 8. Nuclear were swollen to release chromatin after addition of hypotonicization buffer (0.2 mm EDTA, pH 8, containing PMSE, DTT and benzamidine). Chromatin was incubated with anti-H3K9me3 (Abcam AB8898) and anti-H3K27ac (Active Motif, #39133) antibodies overnight at 4 °C. The DNA–protein–antibody complexes were captured by Protein AG Magnetic Beads (Thermo Scientific 88803) by incubation at 4 °C for 2 h. Magnetic beads were then washed with low-salt buffer, high-salt buffer and TE buffer. DNA was eluted from the beads and treated with RNase A, followed by proteinase K digestion. DNA was purified by phenol–chloroform extraction and ethanol precipitation.

For XChIP on cross-linked preparations, formaldehyde-fixed NeuN+ nuclei after FACS were resuspended in lysis buffer containing 0.1% SDS and sonicated (Bioruptor Plus sonication device, Diagenode) at the ‘high’ setting for 30 min on ice. The sizes of the DNA fragments were between 100 bp and 500 bp, with an average size of 300 bp. Chromatin was then incubated with anti-CTCF (EMD Millipore, # 07729), anti-SETDB1 (Santa Cruz H-300X ##sc-66884 X) or anti-SETDB1 (Thermo Fisher SH6A12 #MA5-15722) and captured with Protein AG Magnetic Beads. After washing and elution, the DNA was incubated at 65 °C overnight to reverse the cross-links, which was followed by RNase A and proteinase K treatment and DNA precipitation. SETDB1 occupancies were measured by conventional ChIP-PCR. For CTCF ChIP-seq library preparation, ChIP DNA was end-repaired (End-it DNA Repair kit; Epicentre) and A-tailed (Klenow Exo minus; Epicentre). Adaptors (Illumina) were ligated to the ChIP-DNA (Fast-Link kit; Epicentre) and then PCR-amplified using the Illumina TruSeq ChIP Library Prep Kit. Library DNA with the expected size (NChIP = ~275 bp; XChIP, 350 bp to 500 bp) was selected by Pippin and submitted to the New York Genomic Center and sequenced with an Illumina HiSeq 2000. Each single fragment was sequenced from both ends at each 75-bp length. The ChIP-seq data were then first checked for quality using the various metrics generated by FastQC (v0.11.2). The raw sequencing reads were then aligned to the mouse mm10 genome (or Hg19 for human) using the default settings of Bowtie (v2.2.0). Uniquely mapped reads were retained, and the alignments were subsequently filtered using the SAMTools package (v0.19.1) to remove duplicate reads. Differential analysis between mutant and control samples was performed by using diffReps with a window size of 1,000 bp, a moving step size of 500 bp and a FDR < 5% as significance cutoff42, and data were visualized on the genome using the Integrative Genomics Viewer (IGV) program42. For H3K27ac and CTCF ChIP-seq, peak-calling was performed using MACS (v2.1.1) with a FDR cutoff of 0.05. Gene Ontology enrichment of annotated genes, with significant hits from diffReps within gene bodies or within 3 kb upstream and downstream of the transcription start sites, was further analyzed using DAVID Functional Annotation Bioinformatics tools (Resources 6.7, National Institute of Allergy and Infectious Diseases, NIH). Significant hits from diffReps for a decrease in H3K9me3 and an increase in CTCF ChIP-seq were subjected to motif analysis using the Homer package (v4.8.3) at default settings47. Manhattan plots for genome-wide differential epigenetic profiling of conditional mutants and controls were constructed after the genome was divided into non-overlapping 1-Mb bins, including the 1-Mb bin spanning the clustered Pcdh genes on chr18: 36,870,001–37,870,000, based on the mm10 genome. The number of occurrences of each signal was tabulated within each bin. The probability of the number of occurrence of each signal per 1-Mb bin was then modeled using a Poisson distribution with the maximum-likelihood estimator for the lambda parameter given by the calculated mean number of occurrences. The Poisson models for each signal were used to calculate the probability of occurrence of the signal observed in every 1-Mb bin (including the Pcdh bin).

Chromosome conformation capture (3C) analysis. 3C analysis was performed using standard protocols with minor modifications43. In brief, nuclei were fixed and extracted from mouse cerebral cortex and FACS-sorted as described above. NeuN+(neuronal) nuclei were then pelleted and digested with the HindIII restriction enzyme (New England Biolabs) at 37 °C overnight, washed and treated with T4 DNA ligase at room temperature for 4 h.
3C DNA was then incubated at 65 °C overnight to reverse the cross-links, and the DNA was then purified and precipitated. The 3C primers are listed in Supplementary Table 16. Sequence-verified PCR products were measured semi-quantitatively with UVP Bioimaging system/Labworks 4.5 software. Neighboring primers at the B2m gene locus were used for normalization.

In situ Hi-C analysis, including bioinformatic analyses. Nuclei were fixed and extracted from mouse cerebral cortex, and these were sorted into NeuN+ (neuronal) and NeuN- (non-neuronal) populations, which were then processed using an in situ Hi-C protocol1, with minor modifications. Cultured cells (human NPCs, neurons and astrocytes) were processed similarly. Briefly, the protocol involves a restriction digest of the cross-linked chromatin within intact nuclei, followed by biotinylation of the strand ends, re-ligation, sonication and size selection for 300–500-bp fragments, followed by standard library preparation for Illumina sequencing. The resulting data were mapped, filtered and normalized using HiC-Pro64 (v2.7.8) and visualized on the Washington University Epigenome Browser. To explore localized enrichments in the in situ Hi-C data sets, for each 40-kb bin along chromosome 18 we tabulated the number of long-range interactions >200 kb that were disrupted (those significantly decreased in conditional CK-Cre mutants versus control, as detected using DESeq2 at P < 0.05). The probability of observing the number of disrupted interactions at each bin was then modeled using a Poisson distribution with a maximum likelihood of mean (6.165) calculated from the data. TADs were predicted by using TADtree1 with the 20-kb HiC-Pro data as input and the following parameter settings: maximum size of TAD in bins (S) = 60; maximum number of TADs in each TAD-tree (M) = 10; boundary index parameter (p) = 6; boundary index parameter (q) = 24; balance between boundary index and squared error in score function (p) = 500; and number of TADs to use (N) = 400 (chromosome 18) or 700 (chromosome 5).

In addition, initial processing of the raw 2 × 125 bp pair FASTQ files was performed using the HiC-Pro analysis pipeline. In brief, Hi-C-Pro performs four major tasks: aligning short reads, filtering for valid pairs, binning and normalizing contact matrices. HiC-Pro implements the truncation-based alignment strategy using Bowtie v2.2.3 (ref. 64), mapping either full reads end to end or the 5’ portion of reads preceding a GATGCATC ligation site that results from restriction enzyme digestion with MboI followed by end ligation. Invalid interactions such as same-strand, dangling-end, self-cycle or single-end pairs are not retained. Binning was performed in 40-kb and 100-kb non-overlapping, adjacent windows across the genome, and the resulting contact matrices were normalized using iterative correction and eigenvector decomposition (ICE), as previously described.

Starting with the 20-kb-resolution intrachromosomal contact matrices generated by HiC-Pro, we first generated 100-kb-resolution contact matrices by summing the interaction frequencies of the 20-kb bins within each 100-kb bin. We next generated the corresponding log(observed/expected) matrices, where the observed/expected values were the ratio of the contact values of each interaction bin to the average contact values of all interaction bins the same distance apart. The Pearson’s correlation matrices were then calculated from the log(observed/expected) matrices, and principal component analysis (PCA) was performed on them. The first principal component (PC1) was then used to differentiate the compartments. When the first principal component value was positively correlated to gene density and gene expression (we found that PC1 always correlated with both gene density and gene expression in the same direction), bins with positive PC1 values were assigned as compartment A, whereas bins with negative PC1 values were assigned as compartment B. Conversely, when PC1 negatively correlated to gene density and gene expression, bins with negative PC1 values were assigned as compartment A, whereas bins with positive PC1 values were assigned as compartment B.

Higher-resolution TAD calls were made following the procedure described by Dixon et al.65 using the directionality index (DI) metric. DI was calculated using raw interaction counts between 40- and 100-kb bins, and the window sizes of 2 Mb or 5 Mb, respectively, to capture observed upstream or downstream interaction bias of genomic regions. A hidden Markov model (HMM) was then trained to infer true bias states. TADs were defined by pairing adjacent regions of inferred downstream or upstream bias states.

To identify significantly enriched interactions involving a bin of interest, the expected interaction counts for each interaction distance were estimated by calculating the mean of all intrachromosomal bin–bin interactions of the same separation distance throughout the entire ICE-normalized contact matrix. We estimated the probability of observing an interaction between a bin-of-interest and some other bin by calculating the expected interaction between those two bins divided by the sum of all expected interactions between the bin-of-interest and all other intrachromosomal bins. We then calculated the P value of observing the observed number of interaction counts or more between the bin-of-interest and some other bin by using a binomial test where the number of successes was defined as the observed interaction count, the number of tries as the total number of observed interactions between the bin of interest and all other intrachromosomal bins, and the success probability as the probability of observing the bin–bin interaction estimated from the expected mean interaction counts. To control the FDR, the R package ‘qvalue’ function was used to estimate q-values from the calculated binomial P values.

The insolation analysis was performed with reference to Crane et al.66 and Vietri Rudan et al.18. Briefly, using the 20-kb-resolution Hi-C matrix, we calculated (at each 20-kb bin) the average interaction frequency of the chromosomal bins within a certain distance band. The normalized ‘insulation’ score along the chromosome for each band was then calculated as the log(ratio of average interaction frequency at each 20-kb bin to the average of all 20-kb bins in that band. Regions along the chromosome that display a dip of normalized insolation values represent regions of reduced interactions and can be interpreted as TAD boundaries or regions of high local insulation.

DNA methylation. Targeted Bis-seq was used for the fine-mapping of methylation patterns in candidate cPcdh sequences. Genomic DNA (1 µg) was bisulfite-converted using the Epicentre Bisulfite Kit (Qiagen). Primers were designed in MethPrime67, and bisulfite-converted DNA was amplified and multiplexed by high-throughput PCR with a Fluidigm AccessArray instrument. PCR was performed in duplicate, and the duplicates were pooled. Primers are listed in Supplementary Table 16. The library was diluted to a final concentration of 10 pM with 35% of the PhiX library (Illumina, FC-110-3001). Paired-end reads (250 bp) were generated with an Illumina MiSeq sequencer. Fastq files were generated by the MiSeq sequencer. After trimming off low-quality bases (Phred score < 30), Illumina and Fluidigm adaptors and reads with a length <40 bp using the TrimGalore script, the reads were aligned to the mouse genome (mm10) using Bismark68 and the following settings: -D 50 -R 10--score_min L 0, -0.6. Because the sequences are PCR-based, reads were not deduplicated. Methylation calling was performed using Bismark extractor68. Net methylation was assessed when the coverage was at least 100x and reported by CpG, and averaged across amplicon. Graphical representations of random samples of 50 sequenced DNA fragments were generated using R. Brieﬂy, using CpG context output files generated by Bismark methylation extractor68, which reports CpG methylation status for each individual sequenced DNA fragment (taking into account paired-end reads), methylation patterns for each DNA fragment were reconstructed based on the coordinates of the covered CpGs and their methylation status. Off-target reads that mapped outside the amplicon coordinates were discarded. After random sampling of 50 sequences using the R sample function, their methylation patterns were plotted using the R plot function. Only CpGs present in the reference genome were represented, and sequences were represented on the positive strand. Circles represented consecutive CpGs, with each line being a unique DNA fragment. White circles are unmethylated CpGs, and black circles are methylated CpGs.

To ensure sufficient library complexity, we pooled two PCRs for each amplicon. The median coverage per library ranged from 1,000x to 2,600x (Supplementary Fig. 19a). Because we used a PCR-based targeted Bis-seq approach, the sequence start points were constrained for each amplicon. Therefore, the constrained start sites generated duplication levels that should not be treated as technical duplicates nor removed by bioinformatics deduplication. These duplication levels will be reflected in the distribution of specific sequences that are identified using start sites and therefore not informative to assess the library complexity. Therefore, because some level of randomness of DNA methylation patterns in a cell population is present, to estimate the library complexity, we assessed the number of distinct methylation patterns (specific sequences based on C>T conversion) observed for each amplicon in a given library. Because of biological duplicates (genuinely distinct DNA molecules with the same methylation pattern), this metric provides highly
stringent coverage information. In addition, to take into account sequencing errors, only methylation patterns representing more than 1% of the total reads covering a given amplicon were counted. Overall, the median of methylation patterns per amplicon was 5 (range:1.0–12.5), with 81.2% of the amplicons with less than five methylation patterns covering fully methylated (>80% of methylation) or fully unmethylated (<20%) regions, which were expected to show lower randomness level (Supplementary Fig. 19b).

Lenti-shRNA knockdown of Zfp143. Mouse shRNA lentiviral particles targeting Zfp143 (four unique 29-mer target-specific shRNAs, one scramble control) were purchased from Origene (TL502149V); sequences in the shRNA expression cassettes were verified by the manufacturer to correspond to the target gene Zfp143 (Gene ID https://www.ncbi.nlm.nih.gov/gene/?term=2084120841) with 100% identity and to produce ≥70% gene expression knockdown; provided a minimum transfection efficiency of 80%. NG108 cells (NG108-15 #108CC15, ATCC; organism: Mus musculus (neuroblastoma); Bottas norvegicus (glioma)) were seeded in 12-well plates 12 h before viral transduction. 72 h after transduction, cells were lysed with 500 µl of Trizol, total RNA was extracted and reverse-transcribed. Pcdh gene expression was quantified by real-time PCR. Primers are listed in Supplementary Table 16.

dCas9-SunTag (10xVP64) epigenomic editing. The CRISPR–dCas9_SunTag VP64 two-plasmid system66 was used for genomic editing of the Pcdh locus in the NG108 neuronal cell line. For the plLV-U6-sgRNA-CamK2a-10xGCN4-BFP construct, the lentiviral backbone and dCas9 were cassette were cloned from plasmid Lenti-dCAS-VP64_Blast (Addgene, #61425), the BFP construct, the lentiviral backbone and dCas9 cassette were cloned from the NG108 neuronal cell line. For the pLV-U6-sgRNA-CamK2a-10xGCN4-BFP plasmid, the genome-wide expression knockdown strategy was validated67 with less than five methylation patterns covering fully methylated (>80% of methylation) or fully unmethylated (<20%) regions, which were expected to show lower randomness level (Supplementary Fig. 19b).

Generation of stable dCas9–KRAB NPCs. 3.5 ×10⁶ NPCs per well were seeded on 6-well plates in NPC medium. The following day lentiviruses, generated as described above using either the lentiviral vector dCas9-VP64-T2A-puro or dCas9:KRAB-T2A-puro, were added, and the cultures were spinfected (1 h, 1,000g; 25 °C). Following spinfication, the plates were transferred to a cell culture incubator for 3 h. Medium was then removed and replaced with fresh NPC medium. The following day, fresh NPC medium containing 1 µg/ml puromycin (Sigma, #P2755) was added, and the cells were maintained in NPC medium containing 1 µg/ml puromycin for the remainder of the experiment. Stable NPC lines were validated via FACS using Alexa Fluor 488 (AF488)-conjugated anti-Cas9 (Cell Signaling Technologies, #34963S; 5 µl per 1 × 10⁶ cells).

sgRNA design and cloning. The sgRNAs were designed using the Optimized CRISPR Design tool (see URLs) at the genomic regions of interest. Guide RNAs were selected based on their specific locations at decreasing distances from region of interest, as well as on strand specificity and lack of predicted off-targets. Synthetic oligonucleotides (Supplementary Table 16) were annealed (95 °C for 5 min, ramp down to 25 °C at 5 °C per min), diluted 1:100 and then ligated into BsmB1-digested lentiGuide-dTomato.

NPC lentiviral transduction and FACS. 100,000 dCas9–KRAB NPCs per well were seeded onto growth-factor-reduced Matrigel–coated 24-well plates in NPC medium containing 1 µg/ml puromycin. The following day lentiviruses expressing the scrambled sgRNAs and pooled sgRNAs targeting the PGC-1, PGC-2 or PGC-3 region were added to cultures in the presence of puromycin. 48 h after transduction, NPC cells were FACS-sorted, and the live cell population with a dTomato signal was collected directly into Trizol LS (Thermo Fisher, 10296028). Total RNA was extracted for RT–PCR analysis.

Data availability. All next-generation sequencing data for genome-scale analysis in this paper have been deposited in NCBI’s Gene Expression Omnibus71 and are accessible through GEO series accession number GSE93963. All other data discussed are included in the publication and are available from the authors upon request.

60. Kundakovic, M. et al. Practical guidelines for high-resolution epigenomic profiling of nucleosomal histones in post-mortem human brain tissue. Biol. Psychiatry 81, 162–170 (2017).
61. Shen, L. et al. diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. PLoS One 8, e65598 (2013).
62. Robinson, J.T. et al. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26 (2011).
63. Servant, N. et al. Hi-C-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. 16, 259 (2015).
64. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).
65. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat. Methods 9, 999–1003 (2012).
66. Crane, E. et al. Condensin-driven remodeling of X chromosome topology during dosage compensation. Nature 523, 240–244 (2015).
67. Li, L.C. & Dahiya, R. MethPrimer: designing primers for methylation PCRs. Bioinformatics 18, 1427–1431 (2002).
68. Krueger, F. & Andrews, S.R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27, 1571–1572 (2011).
69. Xu, J. et al. Inhibition of STEP1 ameliorates deficits in mouse and hPSC-based schizophrenia models. Mol. Psychiatry http://dx.doi.org/10.1038/mp.2016.163 (2016).
70. Topal, A. et al. Dysregulation of miRNA-9 in a subset of schizophrenia patient–derived neural progenitor cells. Cell Rep. 15, 1024–1036 (2016).
71. Edgar, R., Domrachev, M. & Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 30, 207–210 (2002).