



Genome-wide binding and mechanistic analyses of Smchd1-mediated epigenetic regulation

Kelan Chen^{a,b}, Jiang Hu^{c,d,1}, Darcy L. Moore^{a,b,1}, Ruijie Liu^{a,1}, Sarah A. Kessans^{e,1}, Kelsey Breslin^a, Isabelle S. Lucet^{a,b}, Andrew Keniry^{a,b}, Huei San Leong^{a,b}, Clare L. Parish^f, Douglas J. Hilton^{a,b}, Richard J. L. F. Lemmers^g, Silvère M. van der Maarel^g, Peter E. Czabotar^{a,b}, Renwick C. J. Dobson^{e,h}, Matthew E. Ritchie^{a,b}, Graham F. Kay^{c,2}, James M. Murphy^{a,b,2}, and Marnie E. Blewitt^{a,b,2,3}

^aThe Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC 3052, Australia; ^bUniversity of Melbourne, Melbourne, VIC 3010, Australia;

^cQueensland Institute of Medical Research Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia; ^dInstitute of Health and Biomedical Innovation, School of Biomolecular Sciences, Faculty of Health, Queensland University of Technology, Brisbane, QLD 4000, Australia; ^eBiomolecular Interaction Centre and School of Biological Sciences, University of Canterbury, Christchurch, Private Bag 4800, New Zealand; ^fThe Florey Institute of Neuroscience and Mental Health, Melbourne, VIC 3010, Australia; ^gDepartment of Human Genetics, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands; and ^hDepartment of Biochemistry and Molecular Biology, Bio21 Institute, Melbourne, VIC 3010, Australia

Edited by Mark Groudine, Fred Hutchinson Cancer Research Center, Seattle, WA, and approved May 27, 2015 (received for review March 2, 2015)

Structural maintenance of chromosomes flexible hinge domain containing 1 (Smchd1) is an epigenetic repressor with described roles in X inactivation and genomic imprinting, but Smchd1 is also critically involved in the pathogenesis of facioscapulohumeral dystrophy. The underlying molecular mechanism by which Smchd1 functions in these instances remains unknown. Our genome-wide transcriptional and epigenetic analyses show that Smchd1 binds *cis*-regulatory elements, many of which coincide with CCCTC-binding factor (Ctcf) binding sites, for example, the clustered protocadherin (*Pcdh*) genes, where we show Smchd1 and Ctcf act in opposing ways. We provide biochemical and biophysical evidence that Smchd1–chromatin interactions are established through the homodimeric hinge domain of Smchd1 and, intriguingly, that the hinge domain also has the capacity to bind DNA and RNA. Our results suggest Smchd1 imparts epigenetic regulation via physical association with chromatin, which may antagonize Ctcf-facilitated chromatin interactions, resulting in coordinated transcriptional control.

Smchd1 | epigenetic control | clustered protocadherins | Ctcf

Structural maintenance of chromosomes flexible hinge domain containing 1 (Smchd1) is an epigenetic repressor that has been shown to play an essential role in autosomal and X-linked gene repression, with critical consequences for normal biology and disease. Smchd1 was originally identified as an epigenetic modifier in an *N*-ethyl-*N*-nitrosourea (ENU) mutagenesis screen (1, 2). The ENU-induced nonsense mutation, termed MommeD1, leads to dramatic reduction of *Smchd1* transcripts and effectively produces a null allele (2). Mice homozygous for this allele display female-specific embryonic lethality due to failure of X chromosome inactivation (2). Although CpG island (CGI) hypomethylation was observed at promoters of a subset of X-linked genes in the absence of Smchd1 (2, 3), Smchd1-dependent gene silencing does not seem to be solely mediated by DNA methylation (4). Indeed, a study of X inactivation in human cells has suggested that SMCHD1 may provide a link between different repressive histone modifications that facilitate heterochromatin formation (5).

In addition to its role in X inactivation, Smchd1 regulates genomic imprinting of a subset of genes within the small nuclear ribonuclear protein N (*SnRNP*)- and *Igf2r*-imprinted clusters (4, 6, 7), and Smchd1/SMCHD1 is involved in regulating the expression of the clustered protocadherin (*Pcdh*) genes (4, 6–8). SMCHD1 has recently been implicated in the pathogenesis of facioscapulohumeral muscular dystrophy, where SMCHD1 is critical for epigenetic repression of the disease causal gene *DUX4* (9, 10). Furthermore, Smchd1 deficiency has been associated with accelerated tumorigenesis in mouse models (6). These findings highlight that Smchd1/SMCHD1 participates in epigenetic regulation at multiple loci in many different cellular scenarios. However, the

precise means by which Smchd1/SMCHD1 modulates gene expression in any of these cases is unknown.

Smchd1 is a noncanonical member of the structural maintenance of chromosomes (SMC) family, comprising an N-terminal ATPase domain and a C-terminal SMC hinge domain (2, 7, 11, 12). SMC proteins form dedicated complexes that play fundamental roles in chromosome dynamics and are implicated in gene regulation, DNA repair, and disease (13–15). The SMC hinge domain mediates dimerization of SMC proteins and confers differential DNA binding properties (14, 16–21). Unlike other SMC proteins, which form a composite ATPase domain from two subdomains that are brought into proximity upon SMC protein heterodimerization, Smchd1 possesses a predicted N-terminal ATPase domain encoded within a single Smchd1 protein. Interestingly, proteins containing an ATPase domain homologous to the ATPase domain of

Significance

Structural maintenance of chromosomes flexible hinge domain containing 1 (Smchd1) is a protein that plays an important role in maintaining gene silencing in many biological circumstances, including facioscapulohumeral muscular dystrophy; however, how it brings about gene silencing is unknown. Understanding the molecular mechanism by which Smchd1 contributes to stable transcriptional silencing is critical to appreciate how it functions in normal biology and when it is mutated in facioscapulohumeral muscular dystrophy. This study reveals, for the first time to our knowledge, where Smchd1 binds genome-wide, its hitherto unappreciated functional interaction with chromatin organizer CCCTC-binding factor in gene regulation, and which part of the protein is required for chromatin binding. These data lead to a new model of Smchd1 function, where it directly binds DNA to mediate 3D chromatin architecture.

Author contributions: K.C., J.H., D.L.M., S.A.K., I.S.L., H.S.L., D.J.H., R.J.L.F.L., S.M.v.d.M., P.E.C., R.C.J.D., M.E.R., G.F.K., J.M.M., and M.E.B. designed research; K.C., J.H., D.L.M., S.A.K., K.B., I.S.L., A.K., C.L.P., and G.F.K. performed research; K.C., J.H., R.L., S.A.K., K.B., I.S.L., A.K., R.C.J.D., M.E.R., G.F.K., and J.M.M. analyzed data; and K.C., R.C.J.D., G.F.K., J.M.M., and M.E.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. [GSE65749](#)).

¹J.H., D.L.M., R.L., and S.A.K. contributed equally to this work.

²G.F.K., J.M.M., and M.E.B. contributed equally to this work.

³To whom correspondence should be addressed. Email: blewitt@wehi.edu.au.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1504232112/-DCSupplemental.

Smchd1 have recently been implicated in heterochromatin compaction and gene silencing in *Arabidopsis* (12, 22).

Previous results from immunofluorescence and ChIP experiments indicated that Smchd1/SMCHD1 is recruited to chromatin, including regions on the inactive X chromosome and at the *DUX4* locus (2, 3, 5, 9). However, the exact site to which Smchd1/SMCHD1 is bound was not delineated, and detailed high-resolution analysis of Smchd1/SMCHD1 occupancy on a genome-wide scale was not available.

To address how Smchd1 affects gene expression at the molecular level, we use genome-wide approaches here and assess Smchd1's chromatin occupancy in conjunction with global gene expression analysis and epigenetic profiling. We find that many Smchd1 binding sites overlap with CCCTC-binding factor (Ctcf) occupancy at both promoters and distal *cis*-regulatory elements. In particular, we demonstrate that Smchd1 and Ctcf display opposing functional effects in regulating the clustered *Pcdh* genes. To investigate how Smchd1 binds to chromatin, we have used a suite of biochemical and biophysical assays, and report that the hinge domain of Smchd1 is capable of direct DNA and RNA binding in vitro. These results indicate that like Ctcf, Smchd1 directly interacts with chromatin but exerts distinct effects on transcription, adding a further layer of complexity in chromatin dynamics and epigenetic regulation.

Results

Transcriptome Analysis of *Smchd1*-Null Neural Stem Cells. To identify differentially expressed (DE) genes in neural stem cells (NSCs), we derived WT and *Smchd1*-null (*Smchd1*^{MommeD1/MommeD1}) NSCs from embryonic day 14.5 (E14.5) mouse brains. Only male cell lines were used in this study because *Smchd1*-null female embryos die around E10.5. We extracted RNA from NSCs ($n = 3$ per genotype) and performed next-generation sequencing, followed by bioinformatic analyses, to quantify the relative transcript levels. We identified 998 up-regulated and 199 down-regulated genes in *Smchd1*-null cells (Fig. 1A and Dataset S1). We performed a gene ontology (GO) analysis of these DE genes using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (23, 24), and found significant up-regulation of genes involved in cell adhesion and other neural-associated functions (Table S1). As

expected, the transcript level of *Smchd1* was dramatically reduced in *Smchd1*-null NSCs, owing to nonsense-mediated decay caused by the MommeD1 mutation. Consistent with previous studies, imprinted genes, including neocidin (*Ndn*), makorin 3 (*Mkrn3*), and paternally expressed gene 12 (*Peg12*), were significantly up-regulated in mutant cells [Fig. 1A and Dataset S1; $P < 0.0001$ by a rotation gene set test (ROAST) (25) of all imprinted genes]. We also found down-regulation of *Grb10*, for the first time to our knowledge, a gene that shows reciprocal imprinting in different areas of the brain (Dataset S1). Strikingly, almost all of the *Pcdh* genes within the alpha and beta clusters, a total of 31 genes, displayed elevated transcript levels (Fig. 1A and Dataset S1; $P < 0.005$ by a ROAST (25) of all *Pcdh* genes), demonstrating a more widespread effect of Smchd1 on these clusters than previously reported in studies performed in whole brain or embryos (4, 7). These differences also likely explain the striking up-regulation of GO terms associated with cell adhesion.

Expression of *Pcdh* Genes Is Altered in *Smchd1*-Null NSCs. The observed up-regulation of *Pcdh* genes in *Smchd1*-null NSCs was verified by quantitative RT-PCR experiments. In agreement with the global analysis, the majority of *Pcdh* genes within the alpha and beta clusters were substantially up-regulated, although expression of genes within the gamma cluster was less perturbed in Smchd1 mutants (Fig. 1B). These data suggest that Smchd1 regulates the expression of *Pcdh* genes, but in a nonuniform manner across the three *Pcdh* gene clusters.

Genome-Wide Occupancy of Smchd1 and Epigenetic Modification Profiling. To uncover the mechanism by which Smchd1 modulates gene expression, we performed ChIP-sequencing (ChIP-seq) with an anti-Smchd1 antibody, using *Smchd1*-null NSCs as a negative control. Model-based analysis for ChIP-Seq 2 (MACS2) peak calling (26) identified 227 highly specific Smchd1 binding sites across the genome (Dataset S1). In parallel, we profiled key genome-wide epigenetic marks in WT and *Smchd1*-null NSCs (H3K4me3 and H3K27me3 ChIP-seq) and assessed CpG methylation via methyl binding domain (MBD)-seq.

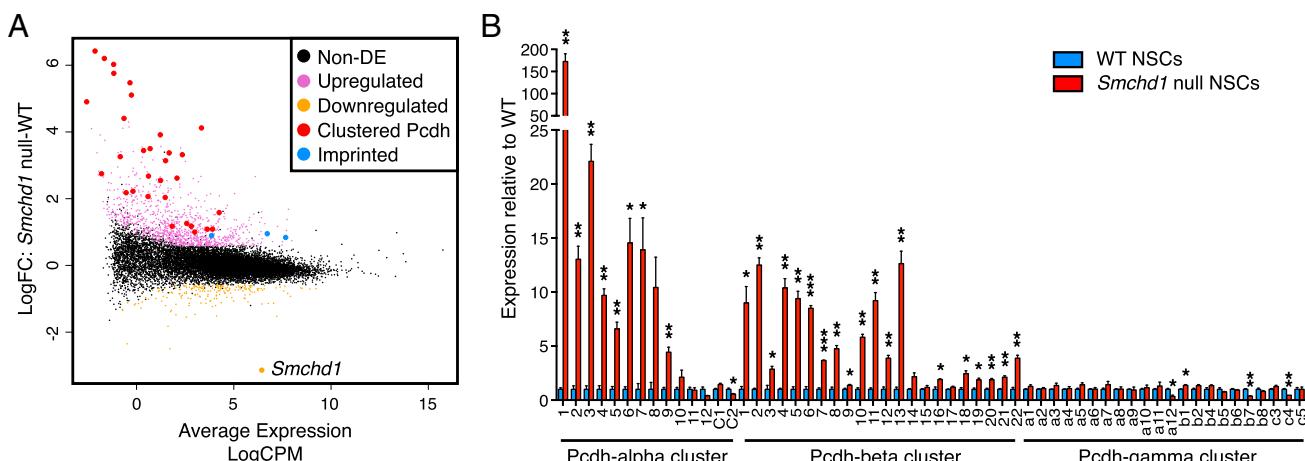


Fig. 1. Gene expression analyses of WT and *Smchd1*-null NSCs. (A) Transcriptome-sequencing data represented by an log ratios versus mean averages (MA) plot showing the log-fold change (LogFC) of normalized expression levels between *Smchd1*-null and WT NSCs ($n = 3$ per genotype) against the average expression in log counts per million (LogCPM). Black dots are non-DE genes. Up-regulated and down-regulated genes (FC > 1.5 , adjusted $P < 0.01$) are shown in pink and orange, respectively. Among the up-regulated genes, genes imprinted with *Ndn*, *Mkrn3*, and *Peg12* are shown in blue. *Pcdh* alpha and beta genes are shown in red. (B) Quantitative RT-PCR (qRT-PCR) quantification of mRNA levels of *Pcdh* genes in alpha, beta, and gamma clusters from WT (blue) and *Smchd1* null (red) NSCs ($n = 3$ per genotype). The qRT-PCR signal was normalized relative to the qRT-PCR signal of *Rala* and plotted relative to the corresponding WT sample. Data are displayed as mean + SEM and were analyzed by an unpaired two-tailed Student's *t* test with a Benjamini-Hochberg correction for multiple testing. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Chromatin Association of Smchd1 and Epigenetic Marks at the *Pcdh* Gene Clusters. The strong altered expression profiles of the clustered *Pcdh* genes prompted us to examine whether Smchd1 was

bound at the *Pcdh* clusters. Initially, we identified a prominent Smchd1 peak (chr18: 37218151–37218474, fold enrichment = 5.07; $P < 5 \times 10^{-4}$) at the HS5-1 site located between the alpha

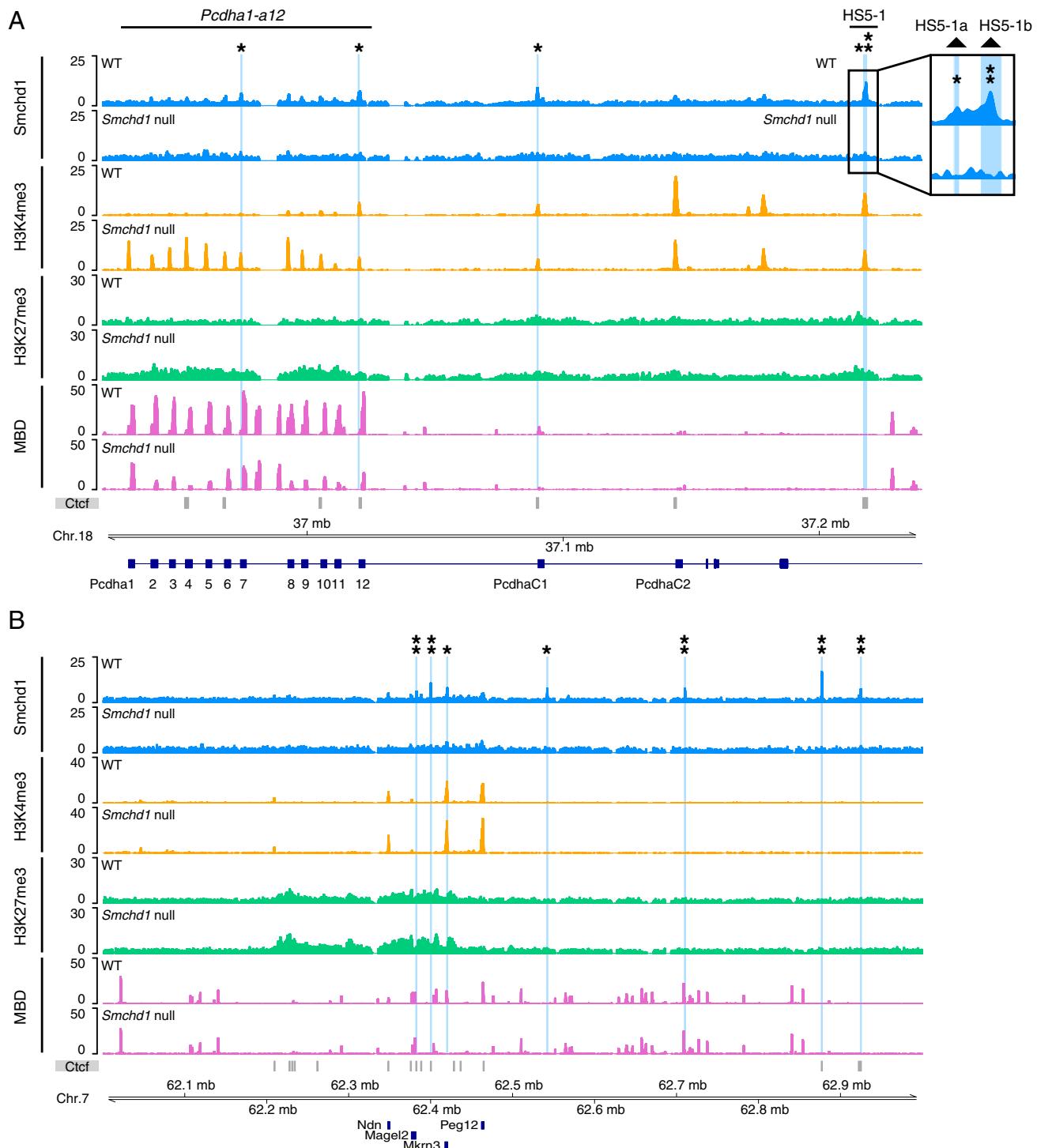


Fig. 2. Aligned ChIP-seq and MBD-seq tracks showing localization of Smchd1 (blue), H3K4me3 (orange), H3K27me3 (green), and CpG methylation (pink) in WT and *Smchd1*-null NSCs at the *Pcdh* alpha cluster (A) and the *Snrpn* cluster (B). Sequenced reads from two (ChIP-seq) or three (MBD-seq) biological replicates of each genotype were combined and plotted as normalized read coverage on the y axis against the genomic location along the horizontal axis. Smchd1 peaks identified by MACS2 are annotated as follows: ** $P < 5 \times 10^{-4}$, $q < 0.1$; * $P < 5 \times 10^{-3}$. (Inset) Zoomed-in view of Smchd1 peaks. Smchd1 peak locations are shown using blue lines that extend through all tracks. Ctcf peaks from the ENCODE E14.5 mouse brain dataset [University of California, Santa Cruz (UCSC) accession no. wgEncodeEM002595] are represented as gray bars below the tracks. The positions of Ctcf peaks at HS5-1a (chr18: 37217137–37217798) and HS5-1b (chr18: 37218262–37218753), as previously published by Monahan et al. (29), were converted from mm9 to mm10 mouse genome and are marked by the black triangles.

and beta clusters (Fig. 2A and [Dataset S1](#)). HS5-1 refers to the *cis*-regulatory DNaseI hypersensitive site, which acts as a transcriptional enhancer of *Pcdh* alpha genes in the nervous system and is required for their repression in nonneuronal lineages (27, 28). We noticed that subtle Smchd1 peaks might not be detected by our stringent genome-wide analysis because our Smchd1 ChIP-seq had considerable background. When we focused on the region spanning the clustered *Pcdh* genes and relaxed the cutoff set for MACS2 peak calling to $P < 5 \times 10^{-3}$, four additional Smchd1 peaks were recovered, including three at promoter regions of the *Pcdh* genes and one at the HS5-1 site (Fig. 2A and [Dataset S1](#)). Previous studies have demonstrated that the Ctcf/Cohesin complex (which contains a Smc1/Smc3 heterodimer) is localized at the HS5-1 site, as well as promoters of actively transcribed *Pcdh* alpha genes (27, 29, 30). Smchd1 binding at the HS5-1 site and the *Pcdh* alpha promoter region was reminiscent of Ctcf/Cohesin occupancy compared with Ctcf peaks from the Encyclopedia of DNA Elements (ENCODE) E14.5 mouse brain dataset. In particular, two Ctcf/Cohesin binding sites within the HS5-1 element, HS5-1a and HS5-1b (29, 30), were both occupied by Smchd1 (Fig. 2A and [Dataset S1](#)). However, we did not detect significant Smchd1 enrichment within the region spanning the beta and gamma clusters (Fig. S1).

We assessed whether Smchd1 deficiency altered the epigenetic modifications at the *Pcdh* clusters. We noticed an acquisition of the active mark H3K4me3 and a reduction of repressive CpG methylation at the promoters of individual *Pcdh* genes in *Smchd1*-null cells, predominantly for *Pcdha1-a12* (Fig. 2A and [Fig. S1](#)). This result was confirmed by read quantification comparing the *Smchd1*-null and WT NSC data. The calculated log₂-fold change for H3K4me3 and CpG methylation at the region encompassing *Pcdha1-a12* promoters was ~2.23 and -0.74, respectively ([Table S2](#)). Similar analysis also revealed a discernible increase in H3K27me3 abundance in *Smchd1*-null cells at this region ([Table S2](#)). Because H3K27me3 is usually associated with gene silencing, this finding was unexpected, potentially indicating some form of epigenetic compensation, albeit inadequate to maintain *Pcdh* repression. At the beta and gamma clusters, differential enrichment for H3K4me3 and CGI methylation was less apparent and H3K27me3 was not significantly changed in *Smchd1*-null NSCs (Fig. S1 and [Table S2](#)).

These data demonstrate that Smchd1 is physically associated with regulatory elements that are implicated in Ctcf/Cohesin-mediated transcription of *Pcdh* alpha genes. In *Smchd1*-null NSCs, altered epigenetic modifications, featuring increased H3K4me3 and reduced CpG methylation at promoters, are concomitant with the induced expression, mostly evident for the *Pcdh* alpha genes.

Chromatin Association of Smchd1 and Epigenetic Marks at the *Snrpn*-Imprinted Gene Cluster. *Smchd1*-null cells show failed genomic imprinting at the *Snrpn* cluster, specifically for the distal half of the cluster encoding *Ndn*; melanoma antigen, family L, 2 (*Magel2*); *Mkra3*; and *Peg12*, which are biallelically expressed in *Smchd1*-null cells (4, 7). We again find up-regulation of *Ndn*, *Mkra3*, and *Peg12* here (Fig. 1A), indicative of loss-of-imprinting and biallelic expression. Analysis of chromatin marks and CpG methylation in this region showed the expected enrichment of H3K4me3 at the promoters of the up-regulated genes (*Ndn*, *Mkra3*, and *Peg12*) and loss of CpG methylation. No significant change in H3K27me3 enrichment was observed (Fig. 2B and [Table S2](#)). We next examined Smchd1 localization in this region using MACS2 with higher ($P < 5 \times 10^{-4}$) and lower ($P < 5 \times 10^{-3}$) stringency settings, as we did for the *Pcdh* cluster. This analysis revealed seven Smchd1 peaks of enrichment: three within the cluster of genes and four more distally located. Interestingly, two Smchd1 peaks near *Magel2* and *Mkra3*, and two at the distal end of the region, overlap with Ctcf binding sites, suggesting that Smchd1 and Ctcf also have some relationship at the *Snrpn* locus. However, at three other Smchd1 peaks (at the *Mkra3* promoter and at two sites downstream of the genes), no such

overlap exists, demonstrating that Smchd1 and Ctcf can bind unique sites (Fig. 2B).

Chromatin Association of Smchd1 and Epigenetic Marks at Homeobox (*Hox*) Gene Clusters. Unexpectedly, our genome-wide analysis also identified Smchd1 binding sites at all four *Hox* gene clusters, some of which overlap with Ctcf peaks from the ENCODE E14.5 mouse brain dataset ([Fig. S2](#)). *Hox* genes encode homeodomain transcription factors that are expressed in a spatially and temporally restricted manner, regulated partly via epigenetic mechanisms (31–33). In embryonic forebrain, the region from which the NSCs were derived, *Hox* genes are repressed via polycomb group protein-mediated silencing (34). We did not detect expression of *Hox* genes in our forebrain-derived NSCs irrespective of the presence or absence of Smchd1. Furthermore, in both WT and *Smchd1*-null NSCs, all four *Hox* gene clusters were decorated with extensive H3K27me3 methylation ([Fig. S2](#)). Thus, Smchd1 binding is dispensable for *Hox* gene silencing in NSCs.

Smchd1 Occupancy Coincides with Ctcf Binding Sites. To annotate the identified Smchd1 binding sites functionally across the genome, we defined the distribution of Smchd1 peaks with respect to gene transcription start sites (TSSs) using the Genomic Regions of Enrichment Analysis Tool (GREAT) algorithm (35). We noticed that only 24% of Smchd1 peaks occur within 5 kb of annotated TSSs (Fig. 3A). In contrast, there was a tendency for Smchd1 to be localized at regions distant from the TSSs (Fig. 3A). These data suggest Smchd1 is not only involved in transcriptional repression at the TSSs but may also bind more distal regulatory elements, such as enhancers or insulators. To test this possibility, we examined Smchd1 occupancy in relation to *cis*-regulatory elements in mouse E14.5 brain, defined by their transcription factor and chromatin mark profiles (36). Strikingly, we observed that overlap between Ctcf and Smchd1 at the *Pcdh* and *Snrpn* clusters extended genome-wide: 138 of the 227 Smchd1 binding sites overlapped with Ctcf, of which 75 were devoid of promoter or enhancer elements (Fig. 3B and [Dataset S1](#)). To rule out the possibility that those 75 sites contained inactive promoters (without H3K4me3 mark or Pol II binding), we applied the GREAT algorithm on those 75 sites and found that the majority of them were distant from TSSs ([Fig. S3A](#)). Thus, our comparative analysis indicated that genome-wide Smchd1 occupancy coincided with Ctcf binding sites, featuring both promoters that are within close proximity of the TSSs and distal enhancers and insulators. This finding was further supported by our de novo motif analysis for Smchd1, which identified two statistically significant motifs corresponding to the Ctcf consensus sequence within 119 of the Smchd1 peaks (Fig. 3C and [Dataset S1](#)). Additionally, a less common motif containing a sequence recognized by RE1 silencing transcription factor/neuron-restrictive silencer factor (Rest/Nrsf), a neuronal gene-specific transcription repressor (37, 38), was identified in 38 Smchd1 peaks (Fig. 3C and [Dataset S1](#)), whereas the other motifs identified do not represent known transcription factor binding sites. The top 10 Smchd1 motifs identified in this analysis are given in [Fig. S3B](#).

Smchd1 and Ctcf Mediate Opposite Effects on the Expression of *Pcdh* Genes. Given the potential co-occupancy of Smchd1 and Ctcf at *cis*-regulatory elements, we next investigated whether Smchd1 and Ctcf could coordinately regulate gene expression, focusing on the clustered *Pcdh* genes. We performed shRNA-mediated knockdown of Ctcf in NSCs with two independent hairpins, validated at both the mRNA and protein levels ([Fig. S4](#)). We selected a set of 13 *Pcdh* genes to test, spread throughout the alpha, beta, and gamma clusters. Upon knockdown of Ctcf, there was noticeably reduced expression of a number of the *Pcdh* genes, compared with the negative control (Fig. 4A), similar to the down-regulation of clustered *Pcdh* genes observed in Ctcf-deficient mouse brains

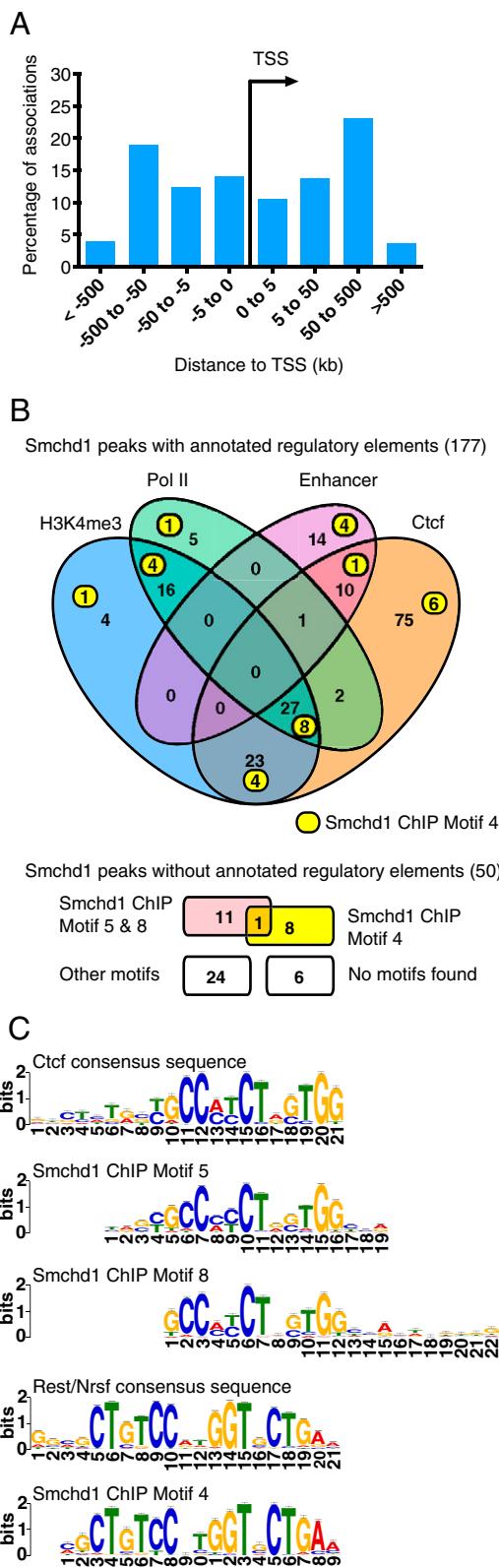


Fig. 3. Genome-wide analysis of Smchd1 chromatin occupancy. (A) Distribution of 227 Smchd1 binding sites relative to TSSs calculated by the GREAT algorithm. The percentages of association are plotted on the y axis, and categories of the distances between Smchd1 peaks and assigned TSSs are plotted on the x axis. (B) Venn diagram showing the overlap between 227 Smchd1 binding sites and the E14.5 mouse brain-specific *cis*-regulatory elements, including promoters (denoted by H3K4me3 and Pol II sites), enhancers, and insulators (Ctcf binding sites), as previously published by Shen et al. (36).

(39). We found that the up-regulated expression of several *Pcdh* genes in *Smchd1*-null NSCs was partially reversed upon Ctcf depletion (Fig. 4*A*). Together, these results suggest that *Smchd1* and Ctcf evoke opposing effects on transcription of *Pcdh* genes, particularly those *Pcdh* genes of alpha and beta clusters. Whether this effect is coordinated through protein–protein interactions between *Smchd1* and Ctcf is less clear, because we did not detect apparent interactions by immunoprecipitation under native conditions (Fig. S4).

To elucidate the observed opposing effects of *Smchd1* and Ctcf, we compared the chromatin localization of Ctcf in WT and *Smchd1*-null cells by performing Ctcf ChIP-seq, using NSCs with Ctcf knockdown as the negative control. Intriguingly, we identified additional Ctcf peaks at the promoters of *Pedha1-a12* genes in *Smchd1*-null cells (Fig. 4*B* and Dataset S1), concurrent with the gain of H3K4me3 and loss of CpG methylation as shown in Fig. 2*A*. These data are highly suggestive that *Smchd1* binding is essential for stabilizing a repressive chromatin environment and potentially involved in antagonizing Ctcf binding at the *Pcdh* alpha promoters, resulting in coordinated regulation.

Smchd1 Hinge Domain Binds to DNA in Vitro. Although our ChIP experiments successfully captured *Smchd1*–chromatin interactions, it was unclear how *Smchd1* bound to chromatin. Given that the hinge domains of other SMC proteins have been implicated in DNA binding (14, 16–19, 21), we produced a recombinant *Smchd1* hinge domain to test whether it too bound to DNA to direct *Smchd1*’s interaction with chromatin (Fig. 5*A*). In parallel, we generated a mutant *Smchd1* hinge domain with a single amino acid substitution, R1867G (Fig. 5*A*), mimicking the mutation present in a facioscapulohumeral muscular dystrophy type 2 family caused by a missense SMCHD1 mutation (40). EMSA with ssDNA demonstrated a shift of 15-mer poly-dT and poly-dC by the WT *Smchd1* hinge domain in a concentration-dependent manner, indicating binding, but there was no binding of poly-dA ssDNA (Fig. 5*B*). The highly positive charge of the *Smchd1* hinge domain meant it remained in the wells, or even ran toward the anode on these gels, similar to the pattern observed for the hinge domains of other SMC proteins (18). Binding activity was dramatically diminished with the pathogenic mutation (Fig. 5*B*).

To verify this finding, we adopted the thermal shift assay (TSA), which monitors the thermal denaturation of sample protein by the fluorescence intensity of the dye SYPRO orange upon its binding to denatured protein (41, 42). In agreement with the EMSA results, we observed an elevated melting temperature of the WT *Smchd1* hinge domain in the presence of ssDNA 30-mer poly-dT, indicating increased thermal stability induced by oligonucleotide binding. Furthermore, although the melting temperature of the R1867G mutant was indicative of a stable and correctly folded protein, such a shift was not observed for the R1867G mutant in the presence of oligonucleotides, reflecting its reduced capacity to bind DNA (Fig. 5*C*).

To assess the *Smchd1* hinge domain–oligonucleotide interactions quantitatively, we performed analytical ultracentrifugation (AUC). Using sedimentation velocity experiments and analyzing the data via the continuous size [c(s)] distribution or species analysis method (as implemented in the program SEDFIT), the concentration of bound 6-carboxyfluorescein (6-FAM) fluorescently labeled oligonucleotide in relation to the known concentration of the free

The number of peaks corresponding to motif 4 (Rest/Nrfs consensus) from the de novo motif analysis is shown. Below the Venn diagram, we show the frequency with which other motifs are observed in those peaks without annotated regulatory motifs. (C) De novo motif analysis based on *Smchd1* peak sequences identified motif 5 and motif 8, which are highly similar to the Ctcf consensus sequence, and motif 4, which is highly similar to the Rest/Nrfs consensus.

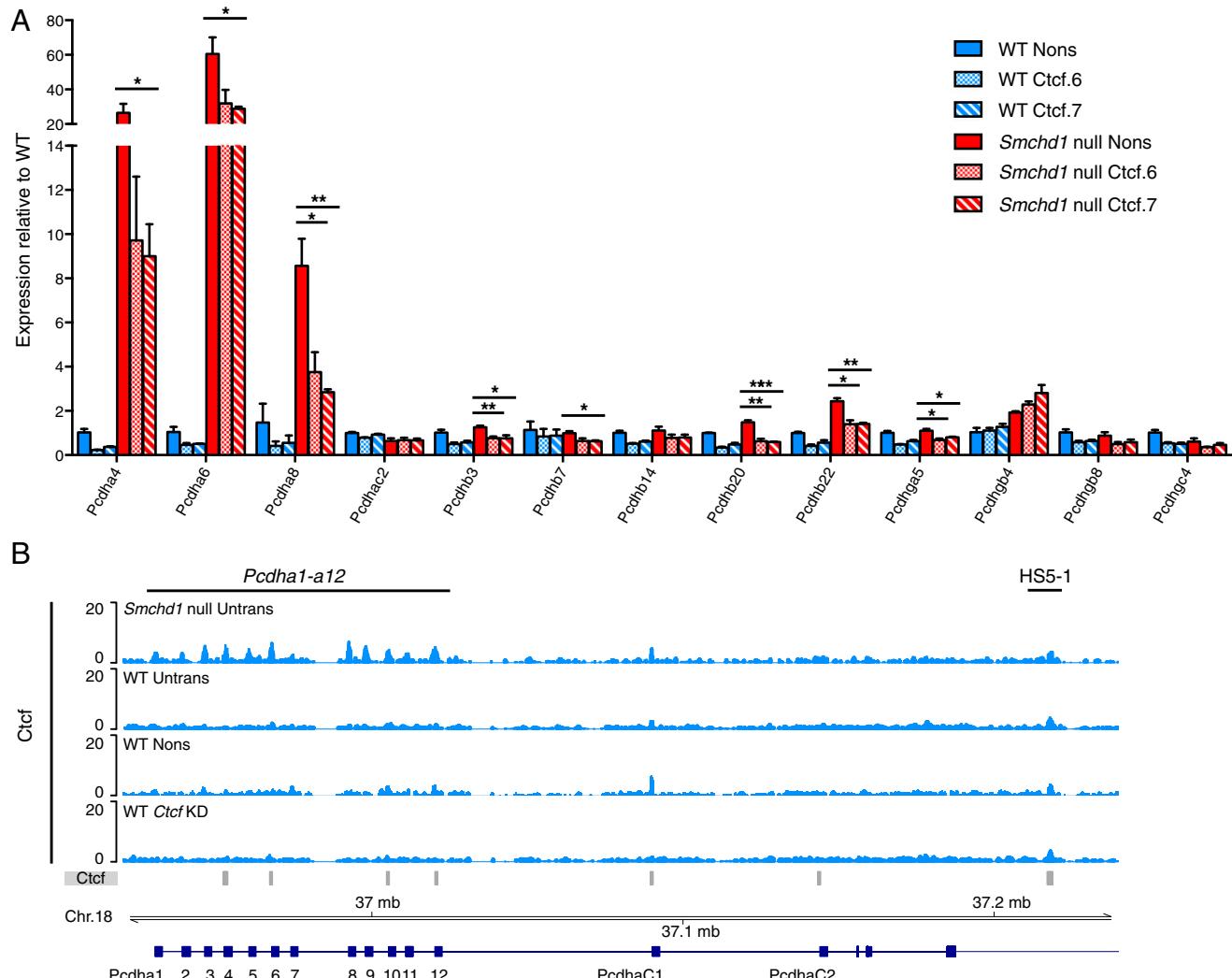


Fig. 4. *Smchd1* and *Ctcf* play opposing roles in regulating *Pcdh* gene expression. (A) qRT-PCR quantification of mRNA levels of a subset of *Pcdh* genes in WT (blue) and *Smchd1* null (red) NSCs with *Ctcf* knockdown by two validated shRNAs, *Ctcf.6* (dotted bar) and *Ctcf.7* (dashed bar), and the nonsilencing negative (Nons; plain bar) control ($n = 3$ per genotype and treatment). The qRT-PCR signal was normalized relative to the averaged signal of *Gusb* and *Hprt*, and plotted relative to the corresponding WT samples. Data are displayed as mean + SEM. Comparison between expression levels in *Ctcf* knockdown and Nons control samples from *Smchd1*-null NSCs was analyzed by an unpaired two-tailed Student's *t* test. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. (B) Aligned ChIP-seq tracks showing localization of *Ctcf* at the *Pcdh* alpha cluster. ChIP-seq experiments were performed in untransduced (*Untrans*) *Smchd1*-null and WT NSCs, Nons control, or *Ctcf* knockdown (*Ctcf KD*) transduced WT NSCs. Sequenced reads from two replicates of each genotype and treatment were combined and plotted as normalized read coverage on the y axis against the genomic location along the horizontal axis. *Ctcf* peaks from the ENCODE E14.5 mouse brain dataset (UCSC accession no. wgEncodeEM002595) are represented as gray bars below the tracks.

oligonucleotide and protein was determined. From this experiment, the dissociation constant (K_d) was calculated (details are provided in *SI Methods* and Fig. S5). This method has the advantage of also reporting the oligomeric state of the free protein and the protein/oligonucleotide complex. Within the concentration range of 1–40 μ M, both the WT and R1867G mutant *Smchd1* hinge domain proteins are dimeric, either free in solution or in complex with the oligonucleotides tested here.

Consistent with the EMSA results, poly-dC displayed higher binding affinity than poly-dA. This result can be seen qualitatively when comparing the c(s) distributions of WT *Smchd1* in the presence of either poly-dC or poly-dA (Fig. S5 C and D). The poly-dC oligonucleotide distribution (Fig. S5D) clearly shows a peak at ~3.5 Svedberg, consistent with the *Smchd1*/oligonucleotide complex, whereas this peak is absent in the distribution with poly-dA (Fig. S5C). The K_d for poly-dC was determined to be 2.5 ± 0.2 μ M, and the K_d for poly-dA was 66 ± 9 μ M. The

weak binding evident in the poly-A continuous size distribution and EMSA experiments also demonstrates that the 6-FAM label does not significantly contribute to protein binding. As expected based on EMSA and TSA data (above), the binding affinity was substantially decreased for the R1867G mutant hinge domain with both oligonucleotides (Fig. S5 C and D).

We next tested oligonucleotides corresponding to the summit of the *Smchd1* peak at the HS5-1b site. We found that the K_d of 20-mer ssDNA containing a *Ctcf* motif in sense orientation exhibits comparable binding to 15-mer poly-dC and, again, that the R1867G mutant hinge domain had dramatically decreased binding affinity (Fig. 5 D and F). In contrast, the binding affinity for the anti-sense-orientated ssDNA was considerably compromised, with the K_d being about 10-fold higher (Fig. 5F and Fig. S5F). The dsDNA produced by annealing those two oligonucleotides was also bound by the WT hinge domain, albeit at an intermediate affinity (Fig. 5F and Fig. S5H).

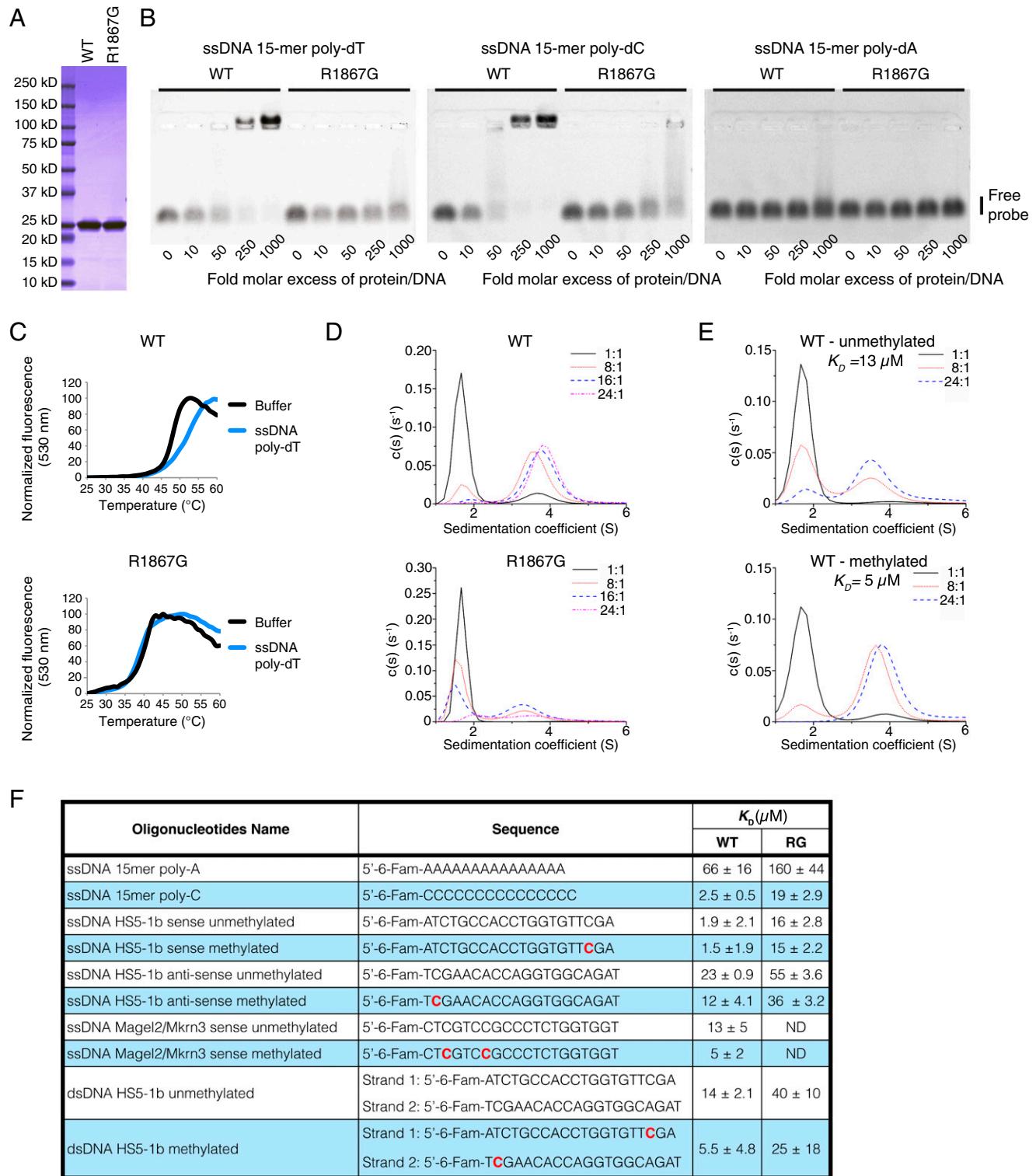


Fig. 5. Hinge domain of Smchd1 binds to DNA in vitro. (A) Coomassie-stained 4–12% (wt/vol) Bis-Tris SDS/PAGE gel of purified WT recombinant Smchd1 hinge domain (WT) and mutated protein with the R1867G substitution. (B) EMSA with 50 nM 15-mer single-stranded poly-dT (Left), poly-dC (Middle), and poly-dA (Right) using protein at increasing concentrations as indicated. (C) Thermostability of the WT (Top) and R1867G mutant (Bottom) Smchd1 hinge domain measured in the presence or absence of ssDNA 30-mer poly-dT. Binding with oligonucleotides increased the stability of the WT protein, thus inducing a shift in the melting temperature. The plot is representative of two independent experiments. (D and E) The $c(s)$ distribution, plotted as a function of Svedberg (S) and detected at 488 nm with the ratio of protein/oligonucleotide concentration as indicated in individual plots. (D) Profiles of $c(s)$ distribution for the Smchd1 hinge domain WT (Top) and R1867G mutant (Bottom) with sense ssDNA unmethylated oligonucleotide corresponding to the HS5-1b Smchd1/Ctfc binding site. (E) Profiles of $c(s)$ distribution for the Smchd1 hinge domain WT bound to sense ssDNA unmethylated oligonucleotide corresponding to the Smchd1/Ctfc binding site near *Magel2* in the *Snrrn* cluster (Top) or the methylated equivalent (Bottom). (F) Summary of K_D values of the WT and R1867G (RG) mutant Smchd1 hinge domain binding to the listed oligonucleotides. The ssDNA containing Ctfc consensus sequence in either sense or antisense orientation and dsDNA produced by annealing the two oligonucleotides were tested individually. Oligonucleotides with an unmethylated or methylated Cyt residue (highlighted in red) were tested separately. The K_D was calculated based on at least three independent experiments and is displayed as mean \pm SE.

Previous studies demonstrated that CpG methylation at a single site could block Ctcf binding to 20-mer dsDNA probes (30). Interestingly, similar CpG methylation did not abrogate binding of the Smchd1 hinge domain to the oligonucleotides (Fig. 5F and Fig. S5 E, G, and I). We next tested the effect of DNA methylation of two CpG sites next to a Ctcf motif within a Smchd1 binding site, between *Magel2* and *Mkm3* in the *SnRpN*-imprinted cluster (Fig. 2B). Here, methylation resulted in enhanced affinity of the WT hinge domain for the oligonucleotide (Fig. 5E and F). Although the exact binding mode is unclear and subject to further investigation, together, these results demonstrate that the Smchd1 hinge domain binds to DNA in vitro, and therefore is a likely candidate for recruiting Smchd1 to its chromatin binding sites.

Because the Smchd1 hinge domain displayed the highest affinity for ssDNA, we used EMSA and AUC to analyze its RNA binding capacity. We found that the WT hinge domain shifted both 15-mer polyU and polyA RNA oligonucleotides in a concentration-dependent manner in EMSA (Fig. S6A). AUC with the RNA oligonucleotides corresponding to the forward and reverse orientations of the *Pcdha12* exonic Ctcf motif (29) demonstrated approximately equivalent binding as for the ssDNA oligonucleotides of the same motif (Fig. S6B). In each case, binding was compromised for the R1867G mutant protein (Fig. S6). These data raise the possibility that Smchd1, like Ctcf (43, 44), may bind both DNA and RNA moieties to achieve its epigenetic function.

Discussion

Here, we present the first, to our knowledge, high-resolution ChIP-seq analysis of Smchd1 binding, which we performed in male murine NSCs. Previous studies have predominantly used immunofluorescence and ChIP/quantitative PCR to describe colocalization of Smchd1 with H3K27me3, H3K9me3, and DNA methylation (2–5, 7, 9). Where ChIP-seq was reported (5), the data were analyzed in 150-kb nonoverlapping regions, possibly due to high background with their anti-SMCHD1 antibody, meaning that the data are of low resolution and would not identify sharp peaks such as we observed. Our high-resolution analysis of Smchd1 binding was enabled by using *Smchd1*-null samples as antibody controls; however, the background still may conceal more subtle features of Smchd1's pattern of binding. Our results are consistent with previous reports but offer deeper insight, which, in concert with our biochemical and biophysical assays, allows us to propose models for Smchd1 binding to chromatin.

We find that Smchd1-bound regions are not restricted to promoters and that many are distant from the TSSs. A significant

proportion of Smchd1 binding sites overlap with Ctcf binding sites, with many being putative *cis*-regulatory elements, including promoters, enhancers, and insulators, indicating that Smchd1 serves a broad array of roles in regulating gene expression. Specifically, Smchd1 binding at the HS5-1 enhancer site and promoters of the clustered *Pcdh* genes (particularly *Pcdha1-a12*), along with specific Ctcf-bound promoters and enhancers in the *SnRpN* cluster, correlate with alterations to other epigenetic marks and expression of genes within these clusters. In the absence of Smchd1, we observed both acquisition of active histone mark H3K4me3 and loss of repressive CpG methylation at individual promoters of these genes, concomitant with their elevated gene expression. However, Smchd1 may not be the sole factor in determining transcriptional outcome because gene regulation often involves multiple redundant pathways. This fact is exemplified by the finding that despite loss of Smchd1 binding at the *Hox* gene clusters, the polycomb group protein-mediated H3K27me3 and gene repression were maintained in *Smchd1*-null male NSCs.

We used the *Pcdh* cluster as a sensitive model to examine the genetic interaction between Smchd1 and Ctcf. We demonstrated a potential functional interaction between Smchd1 and Ctcf by showing their opposing action in regulating *Pcdh* gene expression. Ctcf is known to be involved in regulating gene expression via mediating long-range chromatin interactions and partners with an array of transcription factors, chromatin proteins, and RNA molecules (43, 45–47). It has been proposed that the Ctcf/Cohesin complex recruits selected *Pcdh* alpha promoters to the HS5-1 site, and thus creates an active transcriptional hub for maximal expression (30) (Fig. 6, Left). In accordance with previous studies (30, 48), we observed stable CpG methylation at *Pcdha1-a12* promoters in WT NSCs that could potentially hinder Ctcf binding. On the other hand, additional Ctcf binding was observed at the hypomethylated promoters in *Smchd1*-null NSCs. Together with the discovery of Smchd1's localization at promoter regions and the HS5-1 site, these results have led us to hypothesize that Smchd1 may facilitate a repressive domain that potentially shields promoters of *Pcdh* genes from Ctcf binding (Fig. 6, Middle). This model is further supported by our *in vitro* DNA binding analyses using EMSA together with complementary biophysical assays. However, it is unclear whether Smchd1 and Ctcf act in a competitive manner or whether they could simultaneously co-occupy the HS5-1 site in a cooperative way, where Smchd1 restrains the inactive promoters from the active transcriptional hub (Fig. 6, Right). We propose that Ctcf selectively binds the unmethylated promoters, whereas Smchd1 engages with the methylated

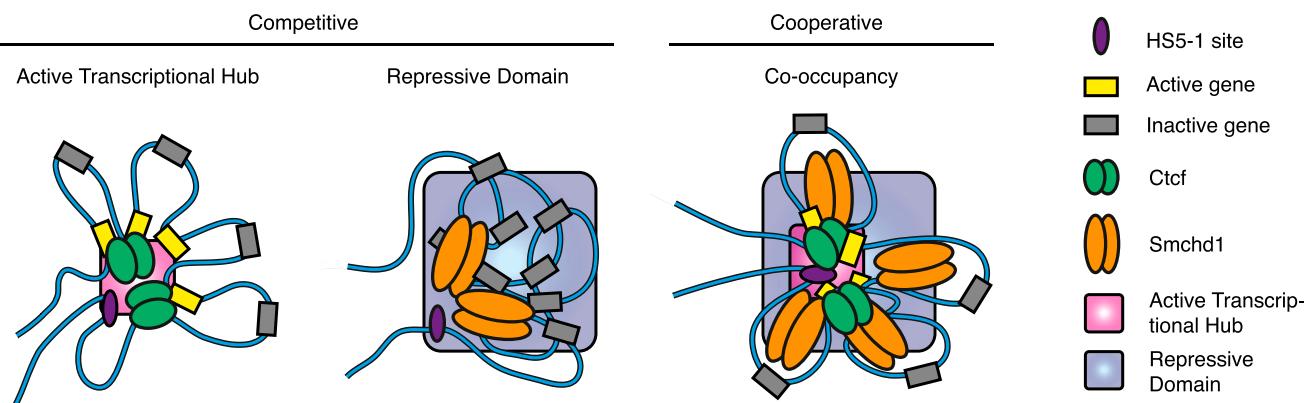


Fig. 6. Schematic diagrams for coordinated regulation of *Pcdh* alpha genes by Smchd1 and Ctcf through competitive (Left and Middle) or cooperative (Right) models. (Left) Active transcriptional hub model is based on a previous study by Guo et al. (30), where Ctcf/Cohesin-mediated long-range chromatin interactions bring active *Pcdh* alpha genes into close proximity to the HS5-1 enhancer site. (Middle) Smchd1 binding may facilitate or maintain a repressive domain antagonizing Ctcf binding. (Right) Alternatively, Smchd1 and Ctcf could co-occupy the HS5-1 site, and Smchd1 might restrain the inactive promoter from the active transcriptional hub.

promoters to maintain their DNA methylation and repressed status; this model may also hold true at the *SnRp*-imprinted cluster, given that Smchd1 binding is enhanced by DNA methylation for at least one site in this cluster. Our data support the idea that without Smchd1, *Pcdh* promoters are unleashed from a repressive chromatin environment, with pursuant engagement of Ctcf/Cohesin resulting in activation, at least for many of the *Pcdh* alpha genes.

In addition to overlap with Ctcf sites, we have identified a subset of Smchd1 binding sites that overlap with Rest/Nrfs occupancy. Rest/Nrfs is a transcription repressor that suppresses neural-specific gene expression in nonneuronal tissues (49–51). It would be interesting to examine if Smchd1 cooperates with Rest/Nrfs in this process in the future. Beyond the clustered *Pcdh* genes, the functional significance of Smchd1 binding at other genomic loci, for example, at the *SnRp* locus, is the subject of ongoing investigation.

To characterize how Smchd1 binds to chromatin, we used a variety of biochemical and biophysical analyses. These analyses have provided the basis for characterizing the exact DNA binding mode of the hinge domain of Smchd1. Our AUC data have revealed several previously unrecognized features of Smchd1: first, the Smchd1 hinge domain dimerizes, similar to the hinge domains from other SMC proteins (16, 17, 20); second, Smchd1 might directly associate with DNA via its hinge domain; third, Smchd1 has the potential to bind methylated DNA, consistent with its predicted function in maintaining DNA methylation (2, 4, 7); and, finally, Smchd1 has the potential to bind RNA. These data are consistent with our model of Smchd1 and Ctcf binding at the *Pcdh* alpha cluster, and they raise the intriguing possibility that like Ctcf (43), Smchd1 may also partner with RNA molecules for part of its epigenetic function. Given Smchd1's unique protein domain arrangement, with a C-terminal SMC hinge domain that directs chromatin interactions and an N-terminal putative ATPase domain, we are keen to investigate the possibility that Smchd1 exerts active manipulation at the chromatin level, and whether this function somehow requires RNA interaction.

In conclusion, our results reveal a potential involvement of Smchd1 in long-range chromatin interaction-mediated epigenetic regulation. We have established a functional link between Smchd1 and Ctcf in regulating expression of the *Pcdh* genes. The coincidence of Smchd1's genome-wide occupancy with a subset of Ctcf binding sites raises the intriguing possibility that Ctcf and its plethora of interacting factors may have some bearing on how Smchd1 regulates transcription. Interestingly, CTCF may also have the opposite effect of SMCHD1 in facioscapulohumeral muscular dystrophy (52). Therefore, further characterization of Smchd1's DNA binding sites and in vivo RNA binding partners in other cellular contexts will provide deeper insights into the molecular mechanism of Smchd1-mediated gene regulation and the underlying implications for development and disease.

Methods

All experimental animals were treated in accordance with the Australian Government National Health and Medical Research Council guidelines under approval from the Animal Ethics Committees of the Walter and Eliza Hall Institute (WEHI AEC 2011.027) and the Queensland Institute of Medical Research (QIMR AEC A0812-610M). Additional details are provided in *SI Methods*.

Derivation and Culture of NSCs. Brains from E14.5 embryos derived from C57BL/6 *Smchd1*^{MommeD1/+} males mated with FVB/N *Smchd1*^{MommeD1/+}

- Blewitt ME, et al. (2005) An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proc Natl Acad Sci USA* 102(21):7629–7634.
- Blewitt ME, et al. (2008) Smchd1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat Genet* 40(5):663–669.
- Gendrel AV, et al. (2012) Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Dev Cell* 23(2):265–279.
- Gendrel AV, et al. (2013) Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes. *Mol Cell Biol* 33(16):3150–3165.
- Nozawa RS, et al. (2013) Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBIX1 pathway. *Nat Struct Mol Biol* 20(5):566–573.
- Leong HS, et al. (2013) Epigenetic regulator Smchd1 functions as a tumor suppressor. *Cancer Res* 73(5):1591–1599.
- Mould AW, et al. (2013) Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation. *Epigenetics Chromatin* 6(1):19.
- Massah S, et al. (2014) Epigenetic characterization of the growth hormone gene identifies Smchd1 as a regulator of autosomal gene clusters. *PLoS ONE* 9(5):e97535.

females were dissected out in Leibovitz's L-15 Medium (Gibco) using NSCs as described previously, and as detailed in *SI Methods* (53).

Transcriptome Sequencing and Analysis. RNA and DNA were extracted from WT and *Smchd1*^{MommeD1/MommeD1} male NSCs using an AllPrep DNA/RNA Mini Kit (Qiagen) according to the manufacturer's instructions and were quantified using a NanoDrop 1000 Spectrophotometer (Thermo Scientific). RNA integrity was assessed with an Agilent Bioanalyzer 2100 (Agilent Technologies) according to the manufacturer's instructions. Sequencing libraries were prepared with a TruSeq Stranded Total RNA Sample Preparation Kit (Illumina) according to the manufacturer's instructions. Sequencing and analysis were as described in *SI Methods*.

MBD-Seq. Genomic DNA was extracted from WT and *Smchd1*^{MommeD1/MommeD1} male NSCs. After purification and fragmentation, methylated DNA was isolated using the MethylMiner Methylated DNA Enrichment Kit (Life Technologies) according to the manufacturer's instructions. The DNA was eluted from MBD-coupled beads in two salt concentration cuts: first, a 600 mM NaCl cut to remove poorly methylated DNA, followed by a 2 M NaCl cut to elute highly methylated DNA; the latter was used for the preparation of libraries for next-generation sequencing. Details are provided in *SI Methods*.

ChIP-Seq and Analysis. ChIP was performed with the fast ChIP protocol (54) with modifications as detailed in *SI Methods*. Antibodies used were as follows: Smchd1 (31865; Abcam), H3K4me3 (07-473; Millipore), H3K27me3 (07-449; Millipore), and Ctcf (07-729; Millipore). Library preparation and sequencing were performed using standard protocols (*SI Methods*). Analysis details are provided in *SI Methods*.

Cloning, Expression, and Purification of Recombinant Protein. cDNA encoding amino acids 1683–1899 of Smchd1 was PCR-amplified from a full-length cDNA clone and cloned into a pPROEX HTb vector (Life Technologies) for expressing 6-His-tagged recombinant protein. The R1867G mutation was introduced by oligonucleotide-directed PCR mutagenesis. Details of the expression and purification procedure are provided in *SI Methods*.

EMSA. The EMSA was performed with 6-FAM fluorescence-labeled and HPLC-purified oligonucleotides (Integrated DNA Technologies) dissolved in 10 mM Tris-HCl (pH 8.5). The dsDNA was annealed by mixing equal volumes of 100 μM ssDNA (strands 1 and 2), incubated at 95 °C for 5 min, and then gradually cooled down to room temperature. The EMSA was performed using the method described by Griesse et al. (18) and is detailed in *SI Methods*.

TSA. The TSA was performed using the method described by Murphy et al. (42), with modifications as detailed in *SI Methods*.

AUC. AUC experiments were conducted using a Beckman Coulter model XL-I instrument at 20 °C. Protein and 6-FAM fluorescently labeled oligonucleotides were mixed in 100 mM NaCl and 20 mM Hepes (pH 7.5), loaded into double-sector quartz cells, and mounted in a Beckman Coulter eight-hole An-50 Ti rotor. Details of the experiments and analyses are included in *SI Methods*, and additional details are shown in *Figs. S5* and *S6*.

ACKNOWLEDGMENTS. We thank Aliaksei Holik for technical assistance. We thank The Dyson Bequest and The DHB Foundation for philanthropic funding (to M.E.B.). This work was supported by grants from the Australian National Health and Medical Research Council (Grant APP1045936 to M.E.B., J.M.M., and M.E.R. and Grant APP1020871 to G.F.K. and M.E.B.). M.E.B. is a Queen Elizabeth II Fellow of the Australian Research Council (DP1096092), and J.M.M. is an Australian Research Council Future Fellow (FT100100100). C.L.P. was supported by a Senior Medical Research Fellowship provided by the Vierlet Charitable Foundation, Australia. This work was made possible through Victorian State Government Operational Infrastructure Support and the Australian National Health and Medical Research Council Research Institute Infrastructure Support Scheme.

- Nozawa RS, et al. (2013) Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBIX1 pathway. *Nat Struct Mol Biol* 20(5):566–573.
- Leong HS, et al. (2013) Epigenetic regulator Smchd1 functions as a tumor suppressor. *Cancer Res* 73(5):1591–1599.
- Mould AW, et al. (2013) Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation. *Epigenetics Chromatin* 6(1):19.
- Massah S, et al. (2014) Epigenetic characterization of the growth hormone gene identifies Smchd1 as a regulator of autosomal gene clusters. *PLoS ONE* 9(5):e97535.

9. Lemmers RJ, et al. (2012) Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat Genet* 44(12):1370–1374.
10. Sacconi S, et al. (2013) The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. *Am J Hum Genet* 93(4):744–751.
11. Iyer LM, Abhiman S, Aravind L (2008) MutL homologs in restriction-modification systems and the origin of eukaryotic MORN ATPases. *Biol Direct* 3:8.
12. Lorkovic ZJ, Naumann U, Matzke AJ, Matzke M (2012) Involvement of a GHKL ATPase in RNA-directed DNA methylation in *Arabidopsis thaliana*. *Curr Biol* 22(10):933–938.
13. Hagstrom KA, Meyer BJ (2003) Condensin and cohesin: More than chromosome compactors and glue. *Nat Rev Genet* 4(7):520–534.
14. Hirano T (2006) At the heart of the chromosome: SMC proteins in action. *Nat Rev Mol Cell Biol* 7(5):311–322.
15. Wood AJ, Severson AF, Meyer BJ (2010) Condensin and cohesin complexity: The expanding repertoire of functions. *Nat Rev Genet* 11(6):391–404.
16. Hirano M, Hirano T (2002) Hinge-mediated dimerization of SMC protein is essential for its dynamic interaction with DNA. *EMBO J* 21(21):5733–5744.
17. Chiu A, Revenkova E, Jessberger R (2004) DNA interaction and dimerization of eukaryotic SMC hinge domains. *J Biol Chem* 279(25):26233–26242.
18. Gries JJ, Witte G, Hopfner KP (2010) Structure and DNA binding activity of the mouse condensin hinge domain highlight common and diverse features of SMC proteins. *Nucleic Acids Res* 38(10):3454–3465.
19. Sun M, Nishino T, Marko JF (2013) The SMC1-SMC3 cohesin heterodimer structures DNA through supercoiling-dependent loop formation. *Nucleic Acids Res* 41(12):6149–6160.
20. Haering CH, Löwe J, Hochwagen A, Nasmyth K (2002) Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol Cell* 9(4):773–788.
21. Gruber S, et al. (2006) Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. *Cell* 127(3):523–537.
22. Moissiard G, et al. (2012) MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* 336(6087):1448–1451.
23. Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44–57.
24. Huang W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37(1):1–13.
25. Wu D, et al. (2010) ROAST: Rotation gene set tests for complex microarray experiments. *Bioinformatics* 26(17):2176–2182.
26. Feng J, Liu T, Qin B, Zhang Y, Liu XS (2012) Identifying ChIP-seq enrichment using MACS. *Nat Protoc* 7(9):1728–1740.
27. Kehayova P, Monahan K, Chen W, Maniatis T (2011) Regulatory elements required for the activation and repression of the protocadherin-alpha gene cluster. *Proc Natl Acad Sci USA* 108(41):17195–17200.
28. Ribich S, Basic B, Maniatis T (2006) Identification of long-range regulatory elements in the protocadherin-alpha gene cluster. *Proc Natl Acad Sci USA* 103(52):19719–19724.
29. Monahan K, et al. (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 USA* 109(23):9125–9130.
30. Guo Y, et al. (2012) CTCF/cohesin-mediated DNA looping is required for protocadherin α promoter choice. *Proc Natl Acad Sci USA* 109(51):21081–21086.
31. McGinnis W, Krumlauf R (1992) Homeobox genes and axial patterning. *Cell* 68(2):283–302.
32. Kmita M, Duboule D (2003) Organizing axes in time and space: 25 years of colinear tinkering. *Science* 301(5631):331–333.
33. Pearson JC, Lemons D, McGinnis W (2005) Modulating Hox gene functions during animal body patterning. *Nat Rev Genet* 6(12):893–904.
34. Noordermeer D, et al. (2011) The dynamic architecture of Hox gene clusters. *Science* 334(6053):222–225.
35. McLean CY, et al. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 28(5):495–501.
36. Shen Y, et al. (2012) A map of the cis-regulatory sequences in the mouse genome. *Nature* 488(7409):116–120.
37. Chong JA, et al. (1995) REST: A mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* 80(6):949–957.
38. Schoenherr CJ, Anderson DJ (1995) Silencing is golden: Negative regulation in the control of neuronal gene transcription. *Curr Opin Neurobiol* 5(5):566–571.
39. Hirayama T, Tarusawa E, Yoshimura Y, Galjart N, Yagi T (2012) CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. *Cell Reports* 2(2):345–357.
40. van den Boogaard ML, et al. (2015) Double SMCHD1 variants in FSHD2: The synergistic effect of 2 SMCHD1 variants on D4Z4 hypomethylation and disease penetrance in FSHD2. *Eur J Hum Genet*, 10.1038/ejhg.2015.55.
41. Lo MC, et al. (2004) Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal Biochem* 332(1):153–159.
42. Murphy JM, et al. (2014) A robust methodology to subclassify pseudokinases based on their nucleotide-binding properties. *Biochem J* 457(2):323–334.
43. Kung JT, et al. (2015) Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF. *Mol Cell* 57(2):361–375.
44. Su D, et al. (2015) Interactions of chromatin context, binding site sequence content, and sequence evolution in stress-induced p53 occupancy and transactivation. *PLoS Genet* 11(1):e1004885.
45. Phillips JE, Corces VG (2009) CTCF: Master weaver of the genome. *Cell* 137(7):1194–1211.
46. Zlatanova J, Caiafa P (2009) CTCF and its protein partners: Divide and rule? *J Cell Sci* 122(Pt 9):1275–1284.
47. Ong CT, Corces VG (2014) CTCF: An architectural protein bridging genome topology and function. *Nat Rev Genet* 15(4):234–246.
48. Kawaguchi M, et al. (2008) Relationship between DNA methylation states and transcription of individual isoforms encoded by the protocadherin-alpha gene cluster. *J Biol Chem* 283(18):12064–12075.
49. Ballas N, Grunseich C, Lu DD, Speh JC, Mandel G (2005) REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell* 121(4):645–657.
50. Greenway DJ, Street M, Jeffries A, Buckley NJ (2007) RE1 Silencing transcription factor maintains a repressive chromatin environment in embryonic hippocampal neural stem cells. *Stem Cells* 25(2):354–363.
51. Johnson R, et al. (2008) REST regulates distinct transcriptional networks in embryonic and neural stem cells. *PLoS Biol* 6(10):e256.
52. Ottaviani A, et al. (2009) The D4Z4 macrosatellite repeat acts as a CTCF and A-type lamins-dependent insulator in facio-scapulo-humeral dystrophy. *PLoS Genet* 5(2):e1000394.
53. Wang TY, Forsythe JS, Nisbet DR, Parish CL (2012) Promoting engraftment of transplanted neural stem cells/progenitors using biofunctionalised electrospun scaffolds. *Biomaterials* 33(36):9188–9197.
54. Nelson JD, Denisenko O, Bomsztyk K (2006) Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat Protoc* 1(1):179–185.
55. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta Ct) Method. *Methods* 25(4):402–408.
56. Vert JP, Foveau N, Lajaunie C, Vandebrouck Y (2006) An accurate and interpretable model for siRNA efficacy prediction. *BMC Bioinformatics* 7:520.
57. Dickins RA, et al. (2005) Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. *Nat Genet* 37(11):1289–1295.
58. Majewski J, et al. (2008) Polycomb repressive complex 2 (PRC2) restricts hematopoietic stem cell activity. *PLoS Biol* 6(4):e93.
59. Liao Y, Smyth GK, Shi W (2013) The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* 41(10):e108.
60. Liao Y, Smyth GK, Shi W (2014) featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30(7):923–930.
61. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139–140.
62. Ritchie ME, et al. (2015) limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res* 43(7):e47.
63. Liu R, et al. (April 29, 2015) Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. *Nucleic Acids Res*, 10.1093/nar/gkv412.
64. Law CW, Chen Y, Shi W, Smyth GK (2014) voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 15(2):R29.
65. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—A practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 57(1):289–300.
66. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
67. Lawrence M, et al. (2013) Software for computing and annotating genomic ranges. *PLOS Comput Biol* 9(8):e1003118.
68. Carlson M (2014) org.Mm.eg.db: Genome wide annotation for Mouse. R package version 3.0.0. Available at bioconductor.org/packages/3.0/data/annotation/html/org.Mm.eg.db.html. Accessed November 1, 2014.
69. Hahne F, et al. (2014) Gviz: Plotting data and annotation information along genomic coordinates. R package, version 1.10.5. Available at www.bioconductor.org/packages/release/bioc/html/Gviz.html. Accessed November 1, 2014.
70. Consortium EP; ENCODE Project Consortium (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* 9(4):e1001046.
71. Bailey TL, et al. (2009) MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res* 37(Web Server issue):W202–W208.
72. Babon JJ, Murphy JM (2013) In vitro JAK kinase activity and inhibition assays. *Methods Mol Biol* 967:39–55.
73. Laue TM, Shah BD, Ridgeway TM, Pelletier SL (1992) *Computer-Aided Interpretation of Analytical Sedimentation Data for Proteins* (The Royal Society of Chemistry, Cambridge, UK).
74. Perugini MA, Schuck P, Howlett GJ (2000) Self-association of human apolipoprotein E3 and E4 in the presence and absence of phospholipid. *J Biol Chem* 275(47):36758–36765.
75. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Iamm equation modeling. *Biophys J* 78(3):1606–1619.
76. Schuck P, Perugini MA, Gonzales NR, Howlett GJ, Schubert D (2002) Size-distribution analysis of proteins by analytical ultracentrifugation: Strategies and application to model systems. *Biophys J* 82(2):1096–1111.