7326–7349 *Nucleic Acids Research, 2022, Vol. 50, No. 13* *Published online 1 July 2022*
https://doi.org/10.1093/nar/gkac531

SETDB1 acts as a topological accessory to Cohesin via an H3K9me3-independent, genomic shunt for regulating cell fates

Tushar Warrier[1,2,†], Chadi El Farran[1,2,†], Yingying Zeng[1,3,†], Benedict Shao Quan Ho[1], Qiuye Bao[1], Zi Hao Zheng[1,4], Xuezhi Bi[5], Huck Hui Ng[6], Derrick Sek Tong Ong[4], Justin Jang Hann Chu[7,8], Amartya Sanyal[3], Melissa Jane Fullwood[3,9], James J. Collins[10,11,12,13], Hu Li[14,*], Jian Xu[2,15,*] and Yuin-Han Loh[1,2,4,16,*]

[1]Cell Fate Engineering and Therapeutics Lab, Cell Biology and Therapies Division, A*STAR Institute of Molecular and Cell Biology, Singapore 138673, Singapore, [2]Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore, [3]School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive 637551, Singapore, [4]Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117593, Singapore, [5]Proteomics Group, Bioprocessing Technology Institute, A*STAR, Singapore 138668, Singapore, [6]Gene Regulation Laboratory, Genome Institute of Singapore, Singapore 138672, Singapore, [7]Department of Microbiology and Immunology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117593, Singapore, [8]Infectious Disease Translational Research Programme, National University of Singapore, Singapore 117597, Singapore, [9]Cancer Science Institute of Singapore, National University of Singapore, 14 Medical Drive, Singapore 117599, Singapore, [10]Howard Hughes Medical Institute, Boston, MA 02114, USA, [11]Institute for Medical Engineering and Science Department of Biological Engineering, and Synthetic Biology Center, Massachusetts Institute of Technology, Cambridge, MA 02114, USA, [12]Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA, [13]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA, [14]Center for Individualized Medicine, Department of Molecular Pharmacology & Experimental Therapeutics, Mayo Clinic, Rochester, MN 55905, USA, [15]Department of Plant Systems Physiology, Radboud Institute for Biological and Environmental Sciences, Radboud University, Heyendaalseweg 135, 6525 AJ, Nijmegen, The Netherlands and [16]NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, 28 MedicalDrive, Singapore 117456, Singapore

Received November 13, 2021; Revised May 30, 2022; Editorial Decision June 02, 2022; Accepted June 30, 2022

ABSTRACT

SETDB1 is a key regulator of lineage-specific genes and endogenous retroviral elements (ERVs) through its deposition of repressive H3K9me3 mark. Apart from its H3K9me3 regulatory role, SETDB1 has seldom been studied in terms of its other potential regulatory roles. To investigate this, a genomic survey of SETDB1 binding in mouse embryonic stem cells across multiple libraries was conducted, leading to the unexpected discovery of regions bereft of common repressive histone marks (H3K9me3, H3K27me3). These regions were enriched with the CTCF motif that is often associated with the topological regulator Cohesin. Further profiling of these non-H3K9me3 regions led to the discovery of a cluster of non-repeat loci that were co-bound by SETDB1 and Cohesin. These regions, which we named DiSCs (domains involving SETDB1 and Cohesin) were seen to be proximal to the gene promoters involved in embryonic stem cell pluripotency and lineage development. Importantly, it was found that SETDB1-Cohesin co-regulate target gene expression and genome topology at these DiSCs. Depletion of SETDB1 led to localized dysregulation of Cohesin binding thereby locally disrupting topological structures. Dysregulated gene expression trends revealed the importance of this cluster in ES cell maintenance as well as at gene 'islands' that drive differentiation to other lineages.

*To whom correspondence should be addressed. Tel: +65 6586 9592; Email: yhloh@imcb.a-star.edu.sg
Correspondence may also be addressed to Jian Xu. Email: j.xu@science.ru.nl
Correspondence may also be addressed to Hu Li. Email: li.hu@mayo.edu
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

The 'unearthing' of the DiSCs thus unravels a unique topological and transcriptional axis of control regulated chiefly by SETDB1.

INTRODUCTION

Histone-lysine *N*-methyltransferase SETDB1 is an enzyme encoded by the *Setdb1* gene. It belongs to the Methyltransferase (EC 2.1.1) class of enzymes (1). This protein chiefly mediates the tri-methylation of lysine residue 9 on histone H3 (H3K9me3). This trimethylation event constitutes an epigenetic silencing signal which recruits HP1 proteins to the methylated histones resulting in transcriptional repression (2,3). Histone modifiers and transcription factors are widely seen to be critical regulators of cell fate and are also key determinants of cellular reprogramming (4,5). SETDB1 also teams up with ATF7IP, which stabilises its methyltransferase activity (6,7) and thereby allows it to repress endogenous retroviral elements (ERVs). SETDB1 is recruited by TRIM28, which is in turn recruited by KRAB zinc-finger proteins that can recognize specific sequences on ERVs, Long Terminal Repeats (LTRs) and other Transposable Elements (TEs) (8–10). Among its other H3K9me3-associated functions, SETDB1 is also seen to pair up with OCT4 to promote the silencing of trophoblastic genes in mouse embryonic stem cells (9). SETDB1 also associates with the PRC2 (polycomb repressive complex 2) at developmental gene loci and contributes to their repression (10). Multiple reports pertaining to the repressive role of SETDB1 encompass a variety of regulatory roles ranging from cell fate maintenance, genomic silencing and development (11) and cellular reprogramming (12). Recent studies have implicated other potential functions for SETDB1 ranging from proximal regulation of a large topological domain (13) to methylation of non-histone substrate proteins in cancers (14). Although the regulation of this proximal topological domain still pertains to the H3K9me3-specific role of SETDB1, its association to a topological region was unique. The non-histone associated roles of SETDB1 were revealed via identification of substrate–enzyme relationships in specific signalling cascades. An unbiased assessment of the overall binding profile of SETDB1 across the genome would thus serve as a preliminary clue towards deciphering its multiple roles in regulation.

Genome topology has two key regulatory components—CTCF (CCCTC-binding factor) and Cohesin (15). CTCF is a transcription factor which contains 11 highly conserved zinc-finger domains (16). Depending on its partners, CTCF can function as a transcriptional activator along with RNA Pol II (17) or as a transcriptional repressor if bound to a histone deacytelase complex (18). Additionally, CTCF demonstrates an insulating function between different compartments of the genome through the creation of topologically associating domains (TADs) (19,20). The Cohesin complex is deposited by NIPBL on to the genome, before it translocates along chromatin (21). Cohesin and CTCF also play critical roles in dictating enhancer–promoter (E–P) contacts as a key component of topological regulation of the genome. This phenomenon has been seen to be particularly relevant to the maintenance of mouse embryonic stem cells (mESCs) (22). Further studies have demonstrated additional roles for Cohesin complex in transcriptional regulation, DNA repair and chromatin condensation. Cohesin facilitates transcriptional insulation along with CTCF (23). Studies have also shown that Cohesin is essential for stabilizing genomic loops (24). The Cohesin complex binds in a dynamic manner across the genome wherein it passively slides across chromatin until it encounters CTCF. This sliding phenomenon engages the Cohesin complex in large loop formation (25).

In this study, we discovered unique regions bound strongly by SETDB1, independent of repressive histone marks, which are co-bound with CTCF and Cohesin, with the enrichment for the latter being particularly strong. Further analysis of these sites, which were referred to as DiSCs (Domains involving SETDB1 and Cohesin) strikingly revealed their co-localization with activating histone marks such as H3K4me3 and H3K27Ac (promoter and enhancer associated marks). Using ChIP-Seq and RNA-Seq, we demonstrated that depletion of SETDB1 resulted in a drastic decrease of Cohesin binding at DiSCs, which was accompanied by the dysregulation of gene expression and genome topology locally. Hi-C data for mESCs revealed a strong enrichment of the DiSCs as topologically enriched regions. Ablation of SETDB1 and Cohesin further disrupted the topological structures at these DiSCs, led to gene compartment switches and contributed substantially to gene dysregulation. Interestingly, these dysregulated genes were key regulators of lineage-specific processes as well as maintenance of ES cell metabolic functions. This was in line with previously attributed functions of SETDB1 in ES cells—such as their survival, maintenance and for differentiation (26). Detailed characterization of the DiSCs reveals a new mechanism of transcriptional and topological regulation by the SETDB1–Cohesin axis.

MATERIALS AND METHODS

Cell culture

E14 mouse embryonic stem cells (mESCs) were cultured using mouse ES medium prepared in DMEM high glucose (Hyclone) supplemented with 15% ES cell FBS (Gibco), 2 mM L-glutamine (Gibco), 1X Pen-Strep (Gibco), 100 μM MEM non-essential amino acids (Gibco), 100 μM β-mercaptoethanol (Gibco), and 1000 U/ml leukemia inhibitory factor (LIF; ESGRO, Millipore). The mouse ES line was cultured on plates pre-coated with 0.1% gelatine (porcine). Mouse embryonic fibroblasts (mEFs) and 293T cells were cultured directly on non-coated, treated culture plates. The media for mEF cells was prepared using DMEM high glucose (Hyclone), along with 10% heat inactivated (HI) FBS, 2 mM L-glutamine (Gibco), 1X Pen-Strep (Gibco) and 100 μM MEM non-essential amino acids. All cultures were maintained at 37°C with 5% CO_2.

Plasmid construction

We used the pGL3-promoter vector with a SV40 promoter upstream of the luciferase gene for cloning DiSC fragments for the Luciferase assays. For the shRNA constructs, we utilized the pSUPER.puro vector for cloning the shRNA

oligos against *Setdb1*, *Smc1a*, *Atf7ip* and *Sumo2*. Additional shRNA oligos targeting the 3′/5′ UTR of the *Setdb1* gene were also cloned into the pSUPER.puro backbone for the purpose of the rescue experiments. For the CRISPR-Cas9 experiments, we used the lentiCRISPR v2 (Addgene plasmid #52961) to clone the sgRNA sequences designed against the DiSC regions. For the constructs used in over-expression and co-immunoprecipitation experiments, we cloned our ORF sequences (*Setdb1, Smc1a*) into pCAG-HA-puro, pCAG-FLAG-puro and pSIN. For rescue experiments, we used the pCAG-HA-hygromycin backbone and the complete *Setdb1* ORF as well as the SET-domain deletion mutant (*mutSET*) were cloned into this vector.

shRNA design

Dharmacon's siDESIGN center (http://dharmacon.gelifesciences.com/design-center/) was used for the design of the shRNA sequences. shRNA sequences were ordered as DNA oligos from IDT and cloned into the pSUPER.puro plasmid (Supplementary Table S1).

Transfection into mESCs

E14 cells were harvested using trypsinization and were re-seeded into six-well plates which were pre-coated with 0.1% gelatine. The seeding density was 300 000 cells per well of the plates. These cells were cultured at 37°C for 20–24 h and the cells were replenished with fresh medium at least 2 h prior to transfection. The cells were then transfected with 3 μg of the plasmid constructs using 4.5 μl of Lipofectamine 2000 (Thermofisher), as per the manufacturer's protocol. Puromycin (1 μg/ml) containing mESC medium was added to the cells 24 h after transfection to select for the transfected plasmids. This process of selection was repeated once more 48 h after transfection. The cells were finally harvested 72 h after transfection. For hygromycin selection, we used mESC media with 250 ug/ml final concentration over a period of 2–3 days. Working stock concentration of hygromycin was 100 mg/ml.

Mithramycin A treatment of mESCs

We titrated the mithramycin A (M6891, Sigma-Aldrich) drug concentration on mESCs after diluting the drug in DMSO at a working stock concentration of 10mM. Mithramycin A at a concentration of 1 μM was seen to substantially deplete *Setdb1* levels in mESCs at 24–48 h post-treatment. For the mithramycin A based rescue system, cells were treated with 1 μM of mithramycin A for a period of 24 h. Mithramycin A was then withdrawn for 12–14 h, followed by transfection with the rescue constructs. Post-transfection (after 24 h), hygromycin selection (200 μg/ml) was coupled with mithramycin A (at 0.5 μM). These cells were cultured for an additional 72–96 h and then harvested.

DZNep treatment of mESCs

Post-seeding of mESCs on gelatine-coated plates, DZNep (diluted in DMSO) (3-deazaneplanocin A hydrochloride) (SML0305, Sigma-Aldrich) was added at a final concentration of 1 μM. After 72 h of culture the DZNep treated cells were harvested.

Chromatin immunoprecipitation

For preparation of chromatin the protocol previously described (27) was utilized. Cells were trypsinized, harvested and the cell number was estimated. Cross-linking was performed using 1% formaldehyde for 10 min at room temperature followed by quenching with 0.125 M Glycine. Cross-linked pellets were washed twice with ice-cold 1× PBS (with 0.1% Triton X-100) and then subjected to lysis with a lysis buffer (10 mM Tris–Cl (pH 8), 100 mM NaCl, 10 mM EDTA, 0.25% Triton X-100 and protease inhibitor cocktail (Roche)). Pellets were resuspended after centrifugation, in 1% SDS lysis buffer (50 mM HEPES–KOH (pH 7.5), 150 mM NaCl, 1% SDS, 2 mM EDTA, 1% Triton X-100, 0.1% NaDOC and protease inhibitor cocktail). Following complete resuspension, the samples were nutated at 4°C for 15 min and the pellets were re-obtained after high-speed centrifugation. This was followed by two washes of all the lysed samples with 0.1% SDS lysis buffer (50 mM HEPES–KOH (pH 7.5), 150 mM NaCl, 0.1% SDS, 2 mM EDTA, 1% Triton X-100, 0.1% NaDOC and protease inhibitor cocktail). For every 10 million cells, 14 cycles of sonication were carried out using the Bioruptor (Diagenode) with 30 s pulses and 60 s halts in each cycle. Cell debris was separated from the sheared chromatin by centrifugation at 15 000 rpm at 4°C for 30 min. Pre-clearing of the chromatin was carried out with 100 ul of Protein G Dynabeads (Life Technologies) for 2 h at 4°C. Simultaneously, 100 ul of Protein G Dynabeads were also bound to 5 ug (per 10 million cells) of the antibodies. After separation of a small amount as input - the remaining pre-cleared chromatin was used to bind to the antibody-bound beads at 4°C, overnight. Elution involved three washes with 0.1% SDS lysis buffer, one with 0.1% SDS lysis buffer/0.35 M NaCl, one with 10 mM Tris–Cl (pH 8.0), 1 mM EDTA, 0.5% NP40, 0.25 LiCl, 0.5% NaDOC, and one with TE buffer (pH 8.0). The immunoprecipitated chromatin was eluted out from the beads by heating the beads re-suspended in 50 mM Tris–HCl (pH 7.5), 10 mM EDTA, 1% SDS, for 1 h at 68°C while shaking at 1400 rpm. Cross-links were reversed by incubating the eluted samples and inputs at 42°C for 2 h and 67°C for 6 h in the presence of Pronase (Sigma) and TE buffer, after which the DNA was purified using the QIAGEN PCR Purification kit (for inputs) and the QIAGEN MinELute PCR Purification kit (for samples) as per the manufacturer's instructions. Quantitative PCR was performed for the purified ChIP-DNA samples by using the CXF384 Real-time System (Bio-Rad), using a Kapa SYBR Fast qPCR kit (Kapa Biosystems). The data represented was normalised to the Inputs as well as the Negative control primers, thus representing an overall fold enrichment.

Sequential chromatin immunoprecipitation (sequential ChIP)

Protein A and Protein G Dynabeads were combined in an equimolar ratio (50 μl per sample) and washed thrice with 1× PBS (with 0.1% Triton X-100). The beads were then resuspended in 600 μl of Pre-Adsorption buffer (equal amounts of ChIP lysis buffer and ChIP dilution buffer (1% Triton X-100, 2mM EDTA, 20 mM Tris–HCl, 150 mM NaCl, 1X Protease inhibitor cocktail)). BSA was added to the resuspended beads at a final concentration of 200

μg/ml. This process of pre-adsorption and blocking of the beads to allow specificity in the pull-down was carried out overnight at 4°C. After pre-adsorption, the beads were washed with ChIP dilution buffer and the first antibody (either SMC1A or CTCF) to 100 μl of the bead mixture. The antibody was bound to the beads at room temperature over a period of 3 h. Sonicated E14 DNA (see Chromatin Immunoprecipitation protocol) was pre-cleared with 100 μl of the beads for 3 h at 4°C. After pre-clearing, 50 μl of the chromatin was separated out as input and the rest was added to the antibody-bound beads and then incubated at 4°C, overnight. For the first elution, the beads were washed thrice with 0.1% SDS lysis buffer, once with 0.1% SDS lysis buffer/0.35 M NaCl, once with 10 mM Tris–Cl (pH 8), 1 mM EDTA, 0.5% NP40, 0.25 LiCl, 0.5% NaDOC, and once with TE buffer (pH 8.0). The beads were separated by a short spin at 800g for 1 min. The beads were resuspended in 75 μl of elution buffer (50 mM Tris–HCl (pH 7.5), 10 mM EDTA, 1% SDS) and kept at 37°C for 30 min. The DNA was eluted from the beads by a brief spin at 1000g for 2 min. 15 μl of the eluted sample was separated out for validation of the first ChIP. The remaining 60 μl was diluted with ChIP dilution buffer to 1200 μl. The second antibody (SETDB1 or SMC1A) was added to freshly pre-adsorbed 100 μl Protein A and G Dynabeads and allowed to bind at room temperature for 3 h. To maintain a 1:19 Input:IP ratio, 63 μl of the first eluted sample was separated out. The rest of the sample was added to the antibody-bound beads and incubated overnight at 4°C. The elution steps for the second ChIP were the same as the first one and then eluted at 68°C for 60 min while shaking at 1400 rpm. Following, this the samples were de-crosslinked, purified and subsequently used for qPCR.

ChIP-Seq library preparation

ChIP samples and the corresponding inputs were prepared as mentioned before (see Chromatin Immunoprecipitation in Methods). Following qPCR validation of the samples, libraries were prepared by using Illumina reagents with TruSeq adapters as per the manufacturer's instructions. The quality and concentration of each sample was assessed by using Agilent High Sensitivity chips on the Agilent 2100 Bioanalyzer. High throughput sequencing for the samples was performed on the HiSeq 4000 (Illumina).

ChIP-Seq analysis

Sequenced ChIP-Seq libraries were first run through the FastQC software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to determine the quality of the libraries. The libraries were then mapped to the mm9 genome assembly by using the STAR aligner while setting the –alignIntronMax option to 1 and the –alignEndsType option to EndToEnd to ensure compatibility of the ChIP-Seq files to STAR. Published library datasets were mapped in a similar manner. To include repeat regions and multi-mapped regions of the genome, we set the –outFilterMultimapNmax option to 500. The makeTagDirectory script of HOMER was executed while using the -keepOne option to retain the repeat elements that were mapped to the genome. Peaks were called for the libraries by using the findPeaks script of HOMER with the -style being set to factor for proteins such as CTCF. For the histone marks, the -style was set to histone. The same was done for the detection of the broad peak binding profile of SETDB1 and SMC1A in the study. All the discovered peaks were annotated by using the annotatePeaks.pl script. For visualization of the ChIP-Seq libraries we generated UCSC bedgraph files by using the makeUCSCfile script on HOMER. The following libraries were downloaded for analysis from the Gene Expression Omnibus (GEO) (28): E14-SETDB1 (GSM440256) (9), V6.5-SETDB1 (GSM459273, GSM459274, GSM459275) (26), H3K9me3 (GSM440257) (9), H3K27me3 (GSM1327220) (29), H3K27Ac (GSM1000126) (30), H3K9Ac (GSM1000123) (31), H3K4me3 (GSM1000124) (31), CTCF (GSM699165) (20), NIPBL (GSM560349) (32), SMC3 (GSM560344) (32), OCT4 (GSM288346) (33), NANOG (GSM288345) (33), SUMO2 (GSM1819197) (34), SUV39H1 (GSM1375157) (35), G9A (GSM1215219) (36), EZH2/SUZ12/JARID2 (GSE18776) (37), HP1A/HP1B/HP1G (GSE97945) (38).

RNA-sequencing

Total RNA was extracted for each of the transfected cells using Trizol reagent (Ambion). DNA contamination for the samples was minimized by using the QIAGEN RNeasy Kit. The RNA samples were processed using a TruSeq Stranded mRNA Library Prep Kit (RS-122-2101, Illumina). This kit was used for mRNA selection, fragmentation, cDNA synthesis and library preparation. The library quality was analyzed on an Agilent Bioanalyzer using the Agilent DNA 1000 kit. High throughput sequencing was then performed on a HiSeq4000 instrument.

RNA-Seq analysis

For the RNA-Seq analysis, the quality control of the libraries was performed by using the RSeQC package (39) and SeqMonk (https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/). The libraries were mapped to the mm9 genome assembly by using the STAR aligner (40). All reads that were mapped to more than one genomic locus or had multiple mismatches were filtered out. Gene expression levels and differential gene expression were assessed by using the cuffdiff tool (41). For visualizing the RNA-Seq libraries on the UCSC browser, the bedgraph files were generated by using the makeUCSCfile script with the addition of the option -style rnaseq. For the gene lists, all genes below FPKM values of 0.5 were filtered out.

Chromosome conformation capture (3C)

As per the recommended protocol (42), 40 million cells were harvested and washed twice with ice-cold PBS. Cells were then cross-linked by using 1% formaldehyde (Sigma) for 10 min at room temperature. The reaction was quenched with 0.266 M of Glycine followed by washing twice with ice-cold PBS. The pellets were lysed on ice for 15 min after resuspending in 1 ml of 3C lysis buffer (10 mM Tris–HCl (pH 7.5), 10 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1× protease

inhibitor). The lysates were further dounced by using a tight and loose douncer, to ensure effective lysis. Nuclei were pelleted by centrifugation and were resuspended in 500 μl of NEB Buffer 3.1 (1×) and washed twice in this buffer and then split into 20 tubes with 25 μl each and spun down. After this, 362 μl of 1× Buffer 3.1, 33 μl of 1% SDS were added to each tube and then incubated at 65°C, 10 min. This was followed by addition of 44 μl of 20% Triton X-100 (Sigma-Aldrich) and 200 U of BglII for enzymatic digestion overnight. Some of the lysate was separated out before and after digestion for using as a control to estimate digestion efficiency by qPCR. The digested DNA was pooled together and treated with the ligation mix (40 ml nuclease-free water, 5 ml 10× Ligation Buffer, 2.687 mL 20% Triton X-100) for 1 h at 37°C. After cooling the mix on ice for 10 min, 67 kilo-units (KU) of NEB T4 DNA Ligase was added and incubated overnight at 16°C. 15 μl of 20 mg/ml Proteinase K (Promega) was then added to the mixture and incubated at 65°C overnight. The samples were cooled to room temperature and 30 μl of 10 mg/ml RNase A (QIAGEN) was added and the samples were incubated at 37°C for 1 h. The 3C libraries were purified by phenol-chloroform extraction and PCR purification kits (QIAGEN). The 3C interactions were then detected by qPCR with the use of a Kapa SYBR Fast qPCR kit.

Co-immunoprecipitation

pCAG-FLAG-*Smc1a* and pSIN-V5-*Setdb1* were co-transfected into 293T cells by using Lipofectamine 2000. These cells were cultured for a period of 72 h after transfection and then harvested and lysed by using the lysis buffer (50 mM Tris–HCl, pH 8.0, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1× protease inhibitor cocktail). Immunoprecipitation (IP) was carried out for FLAG-SMC1A by using 100 μl of Protein A Dynabeads, overnight at 4°C, and then for V5-SETDB1. For elution the beads were washed thrice with the lysis buffer and then resuspended in 2× Laemmli sample buffer (BioRad) and boiled for 10 min, prior to loading on to an 8% SDS-PAGE gel.

Western blot

Cells were lysed by using lysis buffer (containing protease inhibitor cocktail and PMSF) and the lysate was separated at 12,000g, 20 min. As per the requirement, the concentrations of the protein samples were estimated by Bradford assay. The proteins in the lysate are denatured by boiling for 10 min with 2× Laemmli buffer. The protein samples were loaded onto an SDS–PAGE gel and transferred onto a polyvinylidene difluoride membrane (BioRad). The membrane was blocked with 3% BSA at room temperature overnight, followed by incubation with primary antibodies for 1.5 h at room temperature. The secondary horseradish peroxidase (HRP)-conjugated anti-mouse IgG, HRP-conjugated anti-rabbit IgG or HRP-conjugated anti-goat IgG (1:10,000) antibodies were then added to the membrane at room temperature for 1 h. For signal detection, we used the SuperSignal West Dura Extended Duration Substrate (Thermo Scientific) and captured the blots on CLXposure films (Thermo Scientific). The primary antibodies used were anti-V5 (1:1000, ThermoFisher Scientific #R960-25), anti-ACTIN (1:1000, Santa Cruz #sc-10731), anti-FLAG (1:500, Sigma-Aldrich #F1804), anti-ESET (SETDB1) (1:2000, Santa Cruz #sc-66884), anti-SMC1A (1:2000, Bethyl #A300-055A) and anti-H3K9me3 (1:1000, Abcam #ab8898).

Mass spectrometry

E14 cells were cultured in three 15-cm Corning treated, cell culture dishes. After culturing for 72 h, the cells were harvested, crosslinked with 1% formaldehyde and lysed by using lysis buffer. SETDB1 antibody (home-made) (9) was incubated with 100 μl of Protein A Dynabeads for 3 h at room temperature. The lysate was then added to this mix and incubated overnight. The protein was eluted out of the antibody-bead complex by washing 3 times with lysis buffer, followed by boiling for 30 min with Elution buffer 1 (20 mM citrate acid, pH 6 + 0.1% Tween-20). This was followed by another round of elution with Elution buffer 2 (100 mM TEAB, pH 8.5 in 10% SDS). Peptides were extracted and evaporated to dryness in SpeedVac (Savant Instruments, Holbrook, NY, USA), and then dissolved in LC–MS buffer: 2% (v/v) acetonitrile–1% (v/v) formic acid. Mass spectrometry analysis was performed on LTQ-Orbitrap Elite Mass Spectrometer (Thermo Fisher Scientific, San Jose, CA), which was equipped with a nanoACQUITY UPLC system (Waters Milford, MA), Thermo Xcalibur 3.063 and LTQ Tune Plus 2.7.0.1112 SP2 instrument control. Mobile phases composed of A: 0.1% (v/v) formic acid in water, and B: 0.1% (v/v) formic acid in acetonitrile. Peptides (5 μl) were desalted on Symmetry C18 trapping column, 5 μm, 180 μm × 20 mm (Waters) for 5 min with a 1% mobile phase B at 8 μl/min. The desalted peptides were separated on nanoACQUITY UPLC BEH130 C18 column, 1.7 μm, 75 μm × 200 mm, over a 90-min linear gradient from 5% to 40% mobile phase B. Ionization was achieved by nanospray in the positive ion mode at 1.8 kV. Spectra were obtained by data-dependant scanning tandem MS, in which one full MS scan at 120 000 resolution from 350 to 1600 m/z was followed by HCD Orbitrap tandem MS scans of the 15 most intense peptide ions, fragmented with normalized collision energy of 35.0%, at a resolution of 15 000.

Hi-C Seq

Cells treated with the shControl, sh*Setdb1* and sh*Smc1a* constructs were harvested at the 72-h time point post transfection. Assessment for the depletion levels of the *Setdb1* and *Smc1a* transcripts was performed and seen to be >75% depleted. The cells were fixed using 2% formaldehyde and subsequently quenched with 0.125 M glycine. The amount of input DNA per million cells of each type was determined using the protocol recommended by the Arima-HiC + kit. Using the recommended kit, the cells equivalent to 3–5 ug of DNA were used as starting material and were lysed. The DNA released was RE digested, ends were filled-in using biotin and then ligated. The proximally ligated DNA was then sheared using a Covaris sonicator in the range of 300–700 bp and was enriched for biotin-bound DNA prior to ligation to indexed adapters. Finally, the library was amplified

and sequenced using an Illumina HiSeq 4000 instrument with one sample sequenced across three lanes for sufficient depth.

Quantitative PCR

For cDNA samples to be quantitated by qPCR, RNA was converted to cDNA by using the 5X iScript Reverse Transcriptase mix (BioRad). The cDNA samples, after dilution, were run on the CXF384 Real-time System (Bio-Rad), using a Kapa SYBR Fast qPCR kit (Kapa Biosystems). *Actin* or *Gapdh* were used as control primers to normalize gene expression. For ChIP DNA samples, negative control primers were designed based on lack of binding of factors (gene desert regions). Most ChIP data was represented as fold enrichment, with internal normalization to sample input as well as a negative control primer. ChIP data has also been presented as Input Percentage (%) with normalization to the sample input.

Luciferase assay

DiSC fragments were cloned into the pGL3-promoter vector flanked by the KpnI and XhoI restriction sites. E14 cells were co-transfected with pGL3-promoter vector clones and pRL Renilla luciferase control vector using Lipofectamine 2000. Cells were harvested 72 h post-transfection using the Passive lysis buffer (provided with the Dual Luciferase Reporter Assay system (Promega)). Luciferase assay was performed as per the the GloMAX Explorer system protocol.

sgRNA design

Sequences of the SETDB1 and Cohesin binding sites at DiSCs were imported into the CRISPR sgRNA designer (43) and the outputs of the same were checked for their off-target scores. The sgRNA sequences with the lowest off-target scores were chosen and then cloned into the lenti-CRISPR v2 vector and the sgRNA constructs were transfected into the E14 cells by using Lipofectamine 2000.

SURVEYOR assay

E14 cells were transfected with the cloned sgRNA constructs using Lipofectamine 2000 and the transfected cells were then selected using Puromycin (1 μg/ml) from 24 h post-transfection. After two rounds of selection, the cells were harvested 72 h after transfection. Genomic DNA was extracted from the harvested cells using QIAGEN DNeasy Blood & Tissue Kit following manufacturer's protocol. The CRISPR target sites for each of the sgRNA was PCR amplified and purified using QIAquick PCR Purification Kit following manufacturer's protocol. Purified PCR products (400 ng) were mixed with 2 μl 10× Taq polymerase PCR buffer (QIAGEN) and nuclease free water to a final volume of 20 μl. The PCR products were annealed: 95°C for 10 min, 95°C to 25°C (ramping at –0.25°C/s). After re-annealing, products were treated with SURVEYOR nuclease and SURVEYOR enhancer (IDT), following the manufacturer's recommended protocol. These products were analyzed on a 2% agarose gel and was imaged with Gel Doc imaging system (BioRad). The observation of multiple bands implied successful targeting of the sequences by the sgRNA.

Construction of FLAG-*Setdb1* cell line

To generate the 3XFLAG-*Setdb1* knock-in cell line, we designed sgRNA targeting the upstream 5′ region of the *Setdb1* gene using the CRISPR sgRNA designer. This sgRNA was cloned into lentiCRISPR v2. A donor plasmid pCAG-puro was used for cloning the homologous DNA sequence for *Setdb1* with 3XFLAG. The CRISPR and donor plasmids were co-transfected into E14 cells and two rounds of puromycin selection were carried out. The colonies arising out of a single cell were used for immunostaining with FLAG and further confirmed by Sanger sequencing. The confirmed colonies were then expanded to attain a stable cell line.

Flow cytometry

For validating the homogeneity of the FLAG-*Setdb1* cell line, we cultured the cells in gelatine-coated six-well plates and fixed the cells with 4% PFA (paraformaldehyde) after washing with 1× phosphate buffered saline (PBS). The cells were then treated with blocking buffer (0.5% BSA, 1× PBS), followed by the addition of the primary Anti-FLAG antibody (1:500, Sigma-Aldrich #F1804) in the blocking buffer. This was followed by treatment with the Anti-Mouse Alexa Fluor 488 (1:1000, 715-545-150, Jackson Laboratories) as the secondary antibody. Unstained, primary stained and secondary stained as well as dual stained E14 mESCs were used as a control for the assay. The BD FACS Canto instrument was used for the final flow cytometry assay and analysis.

Immunofluorescence

Cells were fixed for 30 min at room temperature using 4% paraformaldehyde solution. This was followed by permeabilization using 0.25% Triton X-100 for 15 min. Blocking was performed by using 1% BSA/0.2% Tween 20 solution for 30 min. The cells were then treated with the anti-FLAG antibody (1:500, Sigma, F1804) or anti-HA probe (1:500, (F-7) sc-7392, Santa Cruz) followed by the Alexa 488 conjugated anti-mouse secondary antibody (Thermofisher) and counter-stained with Hoechst 33342 (Thermofisher). A Zeiss LSM700 confocal microscope was used for capturing the images.

Knock down Hi-C analysis

The Hi-C reads were aligned to the mm9 genome using BWA with runHiC pipeline using the default settings and set the enzyme as Arima (https://pypi.org/project/runHiC/). Valid pairs were retained and output as .cool file which used for downstream analysis. Cooler file were input to the HiC-Peaks (https://github.com/XiaoTaoWang/HiCPeaks) with HiCCUPS algorithm for loop detection using 5 kb resolution (44). TADs were called by Juicer 'Arrowhead' (45) function at 10 kb resolution. Differential compartments were

called using dcHiC using 100 kb resolution (https://github.com/ay-lab/dcHiC) (45). Pile-up analysis of the loops were performed by coolpup.pl using the .cool file (https://github.com/open2c/coolpuppy) (46). The genes enriched in differential compartments were defined by the overlap between the gene regions and differential compartments regions using bedtools. Insulating boundaries and dots were called by cooltools (https://github.com/open2c/cooltools). The contact matrix heatmap of detected TADs and chromatin loops in Figure 6I were shown by GENOVA (47) (https://github.com/robinweide/GENOVA). The genome browser visualization of chromatin loops in Figure 6K was done by IGV (48).

Cohesin Hi-ChIP analysis

Cohesin HiChIP paired-end reads in mESC (GSE80820) (49) were aligned to the mm9 genome using HiC-Pro software (50). Reads were assigned to the MboI restriction enzyme fragments and duplicates were removed from analysis. Valid interaction pairs were generated and converted to the .hic file using Juicer tools (51). HiCCUPs (44) from the Juicer package was used to call the significant loops and the Knight-Ruiz (KR) normalization method (52) was used for the correction of the contact matrix. And the parameters used for calling loops were set as follows: -m 500 -r 5000,10000 -f 0.1,0.1 -p 4,2 -i 7,5 -d 20000,20000. DiSC genomic regions were overlapped with loop anchors by using bedtools, with a threshold setting of minimal 1bp overlap. Finally, aggregate peak analysis (APA) (44) was used to evaluate the enrichment of putative SMC1A loops on the DiSC. The genome interactions in Supplementary Figure S5M and N were visualized by JuiceBox (53).

Differential ChIP-Seq analysis

By using the getDifferentialPeaks function for Homer, we were able to identify differential peaks between shControl Smc1a and sh*Setdb1* Smc1a ChIP-Seq libraries. The .bed file supplied for this contained either the cDiSC or ncDiSC co-ordinates to obtain the differential peaks specifically from these subsets. The fold change option -F was set to 2 to identify the sites with the most significant effects.

Reprogramming

Retroviral vectors purchased from Addgene - pMXs-*Oct3/4* (#13366), pMXs-*Sox2* (#13367), pMXs-*Klf4* (#13370) and pMXs-*c-Myc* (#13375) were co-transfected individually with the pVSV-G and pCMV-intron packaging plasmids into 293T cells using TransIT-LT1 transfection reagent (Mirus). After 48 h, the supernatant (with virus) was harvested and fresh fibroblast medium was added to the culture dishes. After an additional 24 h, the supernatant was collected again and pooled with the pervious supernatant. The total supernatant was filtered through a 0.45 μm pore-size membrane filter (17574-K, Sartorius) and concentrated by ultracentrifugation at 23,000g (Beckman), and stored at −80°C in small 50–100 ul aliquots. Immortalized mEF (imEF) cells were cultured on the Corning cell culture plates. These cultured imEF cells (2000) were seeded onto Corning 12-well plates and infected with OSKM virus simultaneously. We used a feeder free reprogramming medium (DMEM/F12 (Gibco) supplemented with 15% Knockout Serum Replacement (Gibco), 2 mM L-glutamine (Gibco), 1× Pen-Strep (Gibco) 100 μM MEM non-essential amino acids (Gibco), 100 μM β-mercaptoethanol (Gibco), and 1000 U/ml LIF (ESGRO, Millipore), 50 μg/ml ascorbic acid (Sigma), 10 ng/ml basic fibroblast growth factor (bFGF, Gibco), 3 μM CHIR99021 (Stemgent) and 0.5× N-2 Supplement (Gibco)) which was changed daily for the cells. After 16 days of reprogramming, the iPSC-like cells were harvested and cross-linked and subsequently utilized for ChIP using the SETDB1 and SMC1A antibodies.

Enrichment heatmaps

The deeptools software package was utilized for plotting heatmaps based on the ChIP-Seq data. The bamCoverage function was used to create bigwigs out of the mapped files and normalization was performed using the RPGC option with the -effectiveGenomeSize being set for the mm9 assembly. A matrix was generated using the computeMatrix function with the scale-regions and –skipZeros options being utilized. Finally, heatmaps were generated by using the plotHeatmap function.

Clustering analysis of ChIP-Seq libraries

Clustering analysis of ChIP-Seq libraries was performed by first generating Jaccard indices of the bound loci. This was calculated by using BEDTools (54). The Jaccard indices were then clustered by using R for the generation of heatmaps.

Gene ontology analysis

For performing gene ontology analysis, any duplicates in the gene list of interest were removed and the filtered list was input into Metascape (http://metascape.org) (55), while selecting *Mus musculus* as the input species. The list of gene ontology terms was then ordered by the [−log(*P*-value)] for representation.

Interaction networks

The gene names were input into the STRING database (56) and the protein-protein interaction networks were formulated by setting the stringency to low (score = 0.150) and filtering out any text-mined interactions. The interaction networks were saved as .tsv files. For determining relevant sub-networks within the large networks, we input the list of interacting proteins into the MCODE (57) tool within Cytoscape (58).

RESULTS

SETDB1 and Cohesin co-bind to DiSCs devoid of repressive histone marks

To assess the genome-wide binding profile of SETDB1 across ES (embryonic stem) cells, we prepared SETDB1

ChIP-Seq libraries in the E14 mouse embryonic stem cell (mESC) line, using an anti-SETDB1 antibody previously used for ChIP-Seq datasets (9). SETDB1 is known to be associated with the repressive histone marks H3K9me3 and H3K27me3, which are required for endogenous retroviral (ERV) element silencing in ES cells (59,60). Our ChIP-Seq analysis revealed that a significant proportion (∼48%) of SETDB1-bound sites (non-H3K9me3 cluster) were devoid of the repressive H3K9me3 epigenetic marks (Figure 1A, Supplementary Figure S1A). To understand the molecular specifications of this cluster of SETDB1-bound sites, we performed DNA-binding motif analysis on these sites and identified CTCF-binding motif as the most significantly enriched motif exclusively in the non-H3K9me3 cluster (Supplementary Figure S1B).

Since CTCF regulates genome topology along with the Cohesin complex (61,62), we next examined whether the Cohesin subunits SMC1A,SMC3 were bound to the non-H3K9me3 cluster of SETDB1-bound sites. Similar binding profiles could be seen for the ring subunits SMC1A, SMC3 across this cluster of SETDB1-bound sites as well as for the Cohesin loading protein - NIPBL (Figure 1B). CTCF enrichment was also seen at this cluster of SETDB1-bound sites, albeit not as prominent as the Cohesin subunit enrichment profile. This unique cluster of SETDB1 and SMC1A co-bound sites was also found to be completely free from H3K9me3 binding (Supplementary Figure S1C, D). We totally found 2644 such sites (∼30% of non-H3K9me3 sites) were highly enriched with SMC1A binding (Supplementary Table S2) (Supplementary Figure S1E).

A similarity coefficient-based analysis (63) between several factors across cluster 1 SETDB1-bound sites revealed that SETDB1 and SMC1A were grouped together at these sites (Figure 1C), with SETDB1 partnering with SMC1A over conventional SETDB1 partners such as SUMO2. Visually the binding of SETDB1 and SMC1A at these sites was highly co-localized with overlapping peaks covering similar sized stretches over the genome (Figure 1D)—which could also be validated via multiple prepared libraries as well as public datasets. As a result of this co-localization, these sites were annotated as Domains involving SETDB1 and Cohesin (DiSCs). Further visualization of the DiSCs revealed that some of the sites exhibited broad binding profiles for SETDB1 and SMC1A (Figure 1D, Supplementary Figure S1F). To statistically profile these broad regions, a modified peak-calling paradigm for the ChIP-Seq libraries allowed the identification of all broad peaks at the DiSC sites and expanded the list of such sites to 11,947 loci (Figure 1E) (Supplementary Table S3). It is known that histone modifiers such as SETDB1 and topological modifiers such as Cohesin translocate across the genome and hence the broad peak profiles are indicative of an insight into their dynamic locomotion across the genome. Next, a broad versus narrow peak analysis was performed for SETDB1 and Cohesin complex sub-units RAD21 and SMC1A, as well as for CTCF (Supplementary Figure S1G). In both cases a substantial percentage of sites were observed to have broad profiles. In contrast, CTCF demonstrated a very high overlap between the two forms of peaks. This suggests that CTCF binding is highly localized, without the characteristic translocating profile displayed by both SETDB1 and SMC1A. Strikingly, SMC1A binding at SETDB1-bound sites was also seen to be at least 2-fold higher as compared to sites without SETDB1 across the entire genome (Figure 1F). This was the first indication suggesting the strong molecular association between these two proteins. A detailed dissection of the non-H3K9me3, SMC1A-bound loci, i.e. the DiSCs further revealed ∼73% of the sites to be non-repeat regions that were distinct from the usual repeats that SETDB1 populates (Supplementary Figure S1H). Furthermore, motif analysis of the genes proximal to this non-repeat cluster of DiSCs revealed a high prevalence of CTCF and Cohesin sub-units RAD21, SMC3 as compared to other regions (Figure 1G). Other enriched motifs like UBTF at these loci indicate these DiSC-proximal genes to be essential components of embryogenesis and homeostasis (64,65).

Additionally, we were able to validate the DiSCs by displaying binding of SETDB1 across canonical H3K9me3-bound loci (Supplementary Figure S1I) and DiSCs (Supplementary Figure S1J, K) in two public SETDB1 ChIP-Seq datasets (9,26). An overall correlation of our shControl E14 SETDB1 ChIP-Seq with individual public datasets revealed a significant correlation (Supplementary Figure S1L). Comparison of wild-type (WT) E14 SETDB1 ChIP-Seq with shControl SETDB1 ChIP-Seq also revealed a significant correlation (Supplementary Figure S1M), whereas the overall correlation between the individual public datasets seemed to be significantly poorer (Supplementary Figure S1N). We were also able to show conserved binding profiles for SETDB1 and SMC1A across public datasets and for different replicate libraries prepared by using multiple antibodies, cell lines etc. (Figure 1D, Supplementary Figure S1F), thus imparting reproducibility to our findings.

DiSCs exhibit the presence of a putative SETDB1–Cohesin complex

The binding trends of both SETDB1 and Cohesin (represented by its ring sub-unit SMC1A) were validated by ChIP-qPCR (Figure 2A, B, Supplementary Figure S2A). We further validated the SETDB1 binding at DiSCs by engineering a FLAG-*Setdb1* cell line across several tested DiSCs (Supplementary Figure S2B). ChIP-qPCR validation experiments using the FLAG-*Setdb1* cell line showed consistent SETDB1 binding at the DiSC sites (Figure 2C and Supplementary Figure S2C). Additional evidence for SETDB1 and SMC1A co-binding at these sites was provided by sequential ChIP experiments, with no signal seen at loci that were only SETDB1-bound (*Polrmt*) or SMC1A-bound (*Vav2*), as compared to the SMC1A-IgG mock experiment (Figure 2D). SETDB1 and SMC1A were also seen to strongly interact with each other in over-expression Co-IP experiments performed in HEK293T cells (Figure 2E), thereby delineating the presence of a potential regulatory complex. Furthermore, we performed IP-MS for endogenous E14 cells for SETDB1 which revealed the presence of SMC1A and SMC3 (Figure 2F) (Supplementary Table S4). For this IP-MS dataset, we could also detect proven partners of SETDB1 viz. TRIP12, MCM2, TRIM28, HNPRNK as high ranked partners. Other proteins such as HDAC1, CHD4, SMARCA4—which are also proven part-

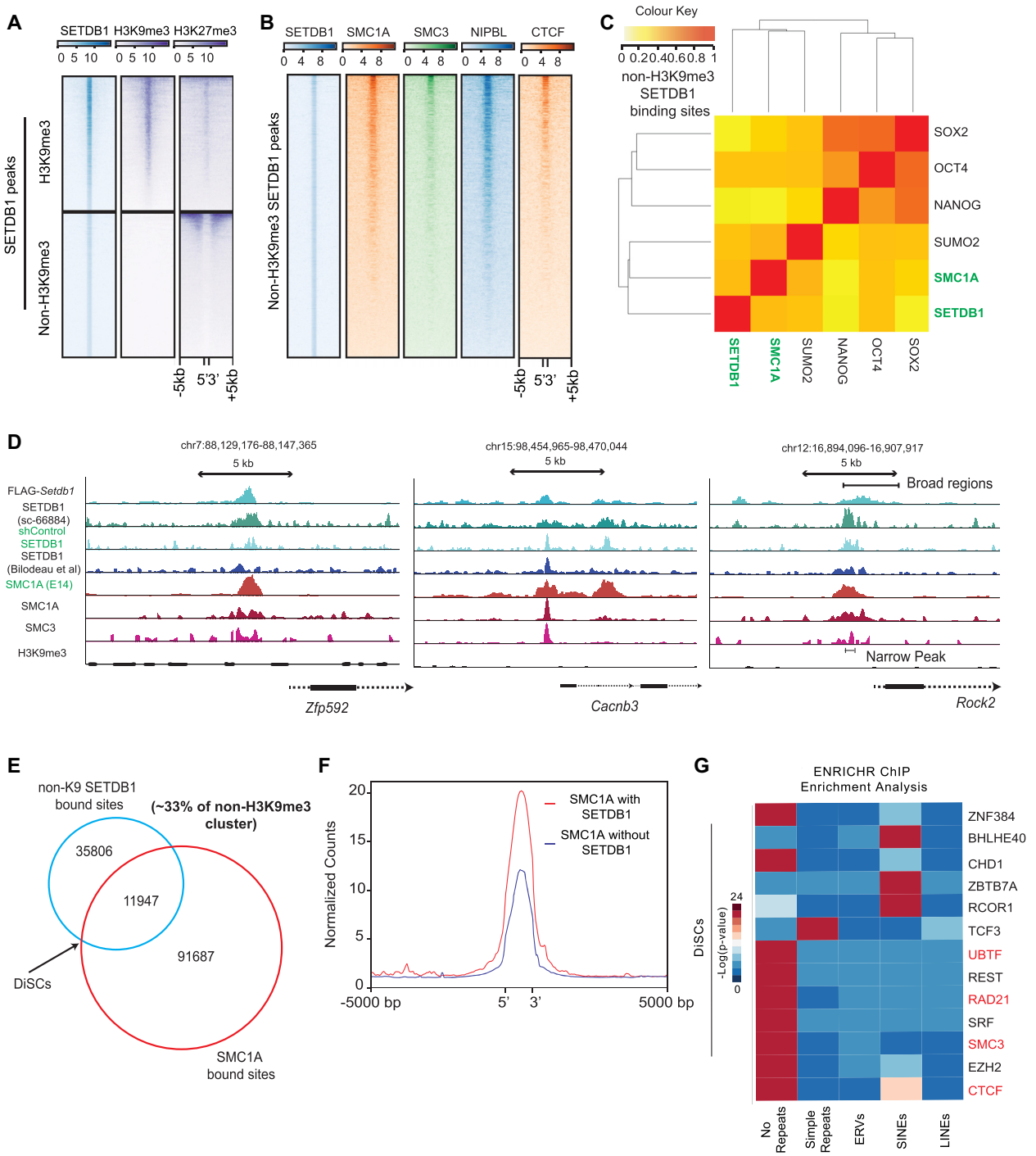

Figure 1. H3K9-independent SETDB1 is enriched at Cohesin-bound sites. (**A**) Heat map representing non-canonical SETDB1 binding sites free of repressive H3K9me3 and H3K27me3 (annotated as non-H3K9me3 cluster) as well as the canonical H3K9me3-bound sites. (**B**) Heat map indicating co-binding of Cohesin complex components – SMC1A, SMC3 and Cohesin loading factor – NIPBL, and CTCF on non-H3K9me3 sites. (**C**) Clustering analysis for non-H3K9me3 sites bound by SETDB1 and Cohesin (SMC1A) compared to other factors—SUMO2, NANOG, SOX2, OCT4 (Jaccard similarity coefficient analysis). The colour intensity signifies the strength of the correlation. Red indicates strong correlation; orange indicates medium and significant correlation and yellow represents weak correlation. (**D**) UCSC browser view of DiSCs indicating SETDB1 and SMC1A binding profiles (with the use of multiple public datasets as well as replicate ChIP-Seq experiments), with absence of H3K9me3. The right-most panel represents the presence of broad binding regions for both SETDB1 and SMC1A at DiSCs, contrasted with conventionally detected narrow peaks. Libraries represented from top to bottom—FLAG-*Setdb1* (FLAG ChIP), E14 SETDB1 (sc-66884), shControl SETDB1 (SETDB1 home-made), SETDB1 (Bilodeau *et al.*), E14 SMC1A, SMC1A (GSM560344), SMC3 (GSM560344), H3K9me3 (GSM440257). (**E**) Venn diagram for enhanced DiSC numbers based on broad peak calling for SETDB1 and SMC1A. (**F**) Average binding profiles indicating higher levels of SMC1A binding at regions bound by SETDB1 in comparison to genomic regions without SETDB1. (**G**) ENRICHR ChIP enrichment analysis for delineation of factors bound to DiSC proximal genes, classified into repeat regions (from right to left—LINEs, SINEs, ERVs, simple repeats) and non-repeat regions. Red indicates high levels of enrichment and blue indicates low enrichment.

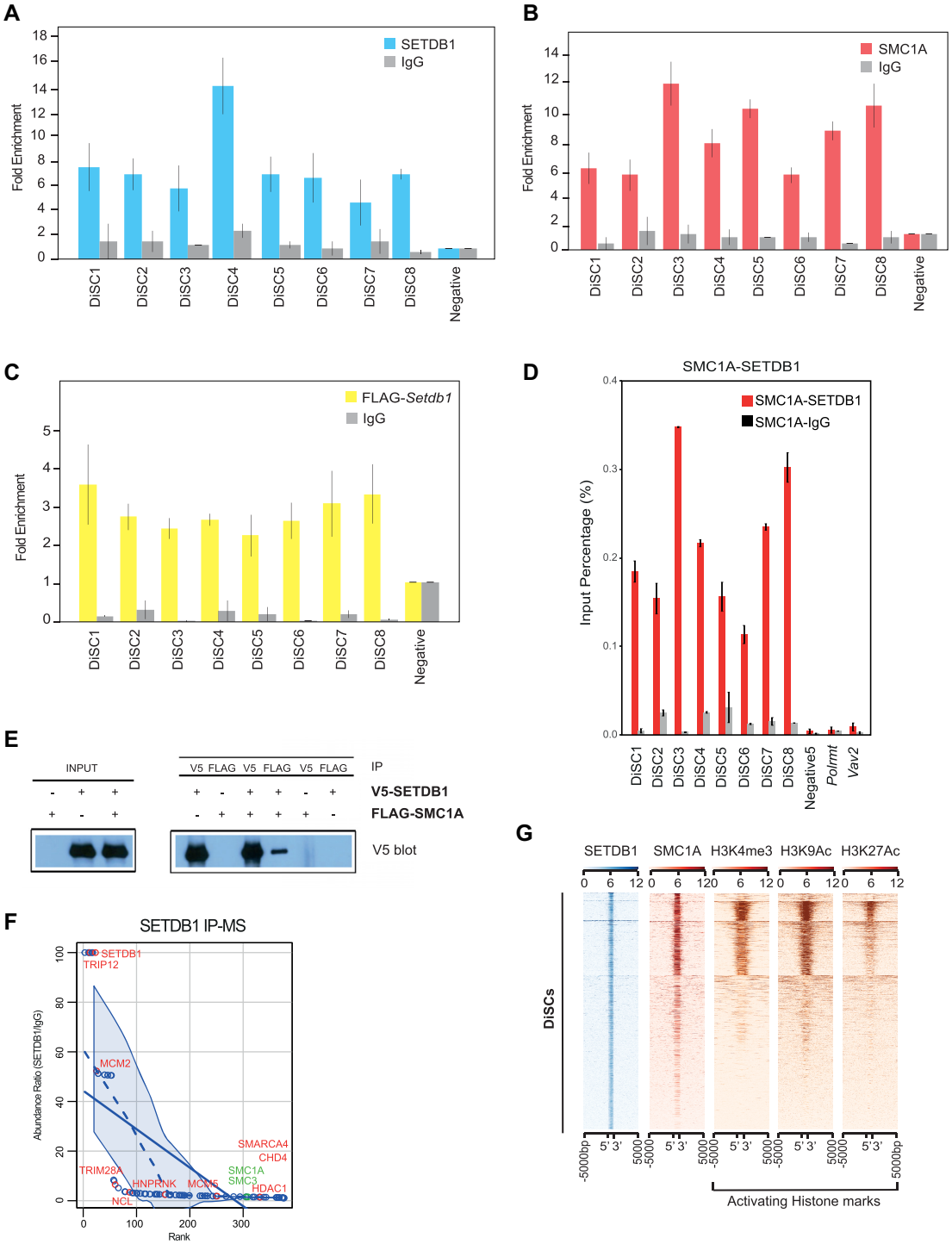

Figure 2. SETDB1 and SMC1A strongly co-exist at DiSCs. (**A**) ChIP-qPCR analysis of SETDB1 binding at DiSCs ($n = 3$, Error bars represent standard deviation). (**B**) ChIP-qPCR analysis of SMC1A binding at DiSCs ($n = 3$, Error bars represent standard deviation). (**C**) ChIP-qPCR analysis of SETDB1 binding at DiSCs as seen by using the FLAG-*Setdb1* cell line ($n = 3$, Error bars represent standard deviation). (**D**) Sequential ChIP-qPCR indicating the co-binding of SMC1A and SETDB1 at DiSCs. *Polrmt* and *Vav2* indicate SETDB1 and SMC1A only binding sites respectively, thereby serving as sequential ChIP controls ($n = 3$, Error bars represent standard deviation). (**E**) Western blot for V5-SETDB1 following overexpression of pSIN-V5-*Setdb1* and pCAG-FLAG-*Smc1a* in the HEK293T cells followed by co-immunoprecipitation indicating the strong interactions between SETDB1 and SMC1A. (**F**) IP-MS dataset for endogenous SETDB1 IP indicating the abundance ratios (compared to IgG) and rank (on x-axis) for SETDB1 partners indicating the presence of SMC1A and SMC3 as prominent partners. TRIP12, MCM2, TRIM28, HNPRNK are known prominent partners of SETDB1 that are also replicated in our dataset. (**G**) Heat maps depicting the binding profiles of active histone marks (H3K4me3, H3K9Ac and H3K27Ac) at the DiSCs.

ners of SETDB1, were also detected and were in fact ranked lower than SMC1A, SMC3. Ablation of *Setdb1* using an shRNA construct, showed substantial reduction in SETDB1 binding at the DiSCs, indicating that the binding was specific (Supplementary Figure S2D, E). ChIP experiments performed after *Smc1a* knock-down (Supplementary Figure S2D, F) also confirmed the binding specificity of Cohesin at DiSCs.

As further validation of the importance of these sites in pluripotent cell types, we carried out reprogramming of mEF (mouse embryonic fibroblasts) cells to induced pluripotent stem cells (iPSCs) (66) and confirmed SETDB1 and SMC1A binding at these sites in the reprogrammed cells (Supplementary Figure S2G). This was a clear indication towards the potentially conserved nature of these sites in pluripotent cell types. *Setdb1* and *Smc1a* were also seen to co-express significantly across varied tissue and cell types in the ENCODE dataset (31) ($r = 0.75$) (Supplementary Figure S2H) and moreover, the DiSCs were also seen to be moderately conserved in terms of genomic sequence characteristics across a set of 30 vertebrate species as per their PhastCon scores (Supplementary Figure S2I). These findings successfully demonstrate an unprecedented but strong association between SETDB1 and Cohesin at the DiSCs.

To prove the strong association of SETDB1, SMC1A via DiSCs in the mESC state, we profiled the binding of these proteins across a panel of DiSC sites ($n = 62$) using chromatin from the 2i naïve cells, immortalized mouse embryonic fibroblasts (iMEF) and wild-type mouse embryonic fibroblasts (MEF) and E14 mESCs. Both SETDB1 and SMC1A enrichment was seen to be significantly stronger in the mESC state across the DiSCs as compared to the other cell types (Supplementary Figure S2J). Additionally, this cell type specificity was also seen to extend to the cell specific transcriptome, when we assessed the expression levels of DiSC-proximal genes in mESCs, as compared to 2i (GSE92407) (67), MEF cells (GSE113430) (68). The DiSC-proximal genes were expressed more significantly in the mESC state (Supplementary Figure S2K). Typically, SETDB1-bound regions of the genome constitute heterochromatin, i.e. inaccessible/closed chromatin. However, the DiSCs comprised of open and accessible chromatin in the wild-type state, whereas they declined in accessibility upon KO of *Setdb1*. In contrast, SETDB1 and H3K9me3-bound regions of the genome were closed in the wild-type state and showed an increase in accessibility in the KO *Setdb1* state (Supplementary Figure S2L).

Since the DiSCs were bereft of repressive histone marks – which are common associates of SETDB1, we profiled other activating histone marks across the DiSC sites, and strikingly observed a presence of H3K4me3 and H3K27Ac, H3K9Ac (which are marks enriched at promoters and enhancers) (Figure 2G, Supplementary Figure S2M). This was a strong indication of the SETDB1–Cohesin complex at DiSCs being involved in transcriptional regulation.

DiSCs mediate transcriptional control of their adjacent genes

Following on from the enrichment of activation histone marks proximal to DiSCs as well as the mESC-specific expression profile for the DiSC-proximal genes, we were curious about gene proximity to the DiSCs. Therefore, a profile was constructed based on their distance from the closest transcription start sites (TSS) (Figure 3A). Majority (51%) of the DiSCs were seen to be distal with respect to TSS, i.e. at 10 000 bp or more, whereas around 28% can be found within 1000 bp of the nearest TSS. Hyper-geometric tests confirmed that most of the DiSCs were located proximal to the promoters (∼33%) and 5′UTR (∼6%), whereas ∼8% were located downstream (1000–3000 bp) to the genes (Supplementary Figure S3A). Since significant numbers of DiSCs were proximal to the gene regulatory regions, we postulated that they were likely to engage in transcriptional activity. DiSC sites were selected and successfully tested for transcriptional activity using a luciferase assay (Supplementary Figure S3B).

To further explore gene regulatory roles of the DiSCs, we prepared RNA-Seq libraries for shControl, sh*Setdb1* and sh*Smc1a* cells. These libraries showed depleted expression of the shRNA-targeted genes (Supplementary Figure S3C). In the *Setdb1*-depleted libraries, 892 genes (proximal to the DiSCs) were upregulated while 1322 genes were downregulated. In the *Smc1a*-depleted libraries, 998 genes (proximal to the DiSCs) were upregulated and 1338 were downregulated (Supplementary Figure S3D). While a high number of differentially expressed genes (DEGs), i.e. 800, in the sh*Setdb1* RNA-Seq were seen to be DiSC-proximal genes (Supplementary Figure S3E), the majority of genes proximal to SETDB1, H3K9me3 co-bound loci were upregulated rather than downregulated as expected (Supplementary Figure S3F, Figure S3G). A substantial percentage of the dysregulated genes were seen to be impacted by both *Setdb1* and *Smc1a* depletion, i.e. 420 of the common dysregulated genes overlapped with DiSC-proximal genes (Figure 3B). Out of this number 264 genes were downregulated and 156 of the upregulated genes were common between the sh*Setdb1* and sh*Smc1a* libraries (Supplementary Figure S3H) (Supplementary Table S5) On the other hand genes commonly dysregulated between sh*Setdb1* and sh*Smc1a* did not overlap significantly with SETDB1, H3K9me3 co-bound genes (Supplementary Figure S3I). Out of the common dysregulated genes between sh*Setdb1* and sh*Smc1a*, ∼65% of the genes were commonly downregulated whereas the remaining 35% were upregulated (Figure 3C). Gene ontology (GO) analysis for the commonly downregulated DiSC-proximal genes revealed enrichment for core stem cell maintenance and survival processes – blastocyst development, stem cell population maintenance, and broadly signalling pathways regulating stem cells (Figure 3D). Similar analysis for the upregulated DiSC-proximal genes displayed functions relevant to post-embryonic development, brain development, glial cell differentiation (Figure 3E). GO for the SETDB1-H3K9me3 regulated genes was seen to be distinctly different as it was enriched with varied metabolic processes, cell maintenance functions and other differentiation related processes (Supplementary Figure S3J). The GO analysis at the DiSCs was strongly supported by the factors enriched at the DiSC-proximal upregulated genes – TAF1, ATF2, MYC, all of which have been implicated in distinct roles in neuronal development and NPC differentiation (69,70) (Supplementary Figure S3K). The factors enriched at the downregulated DiSC genes - UBTF, FOSL2,

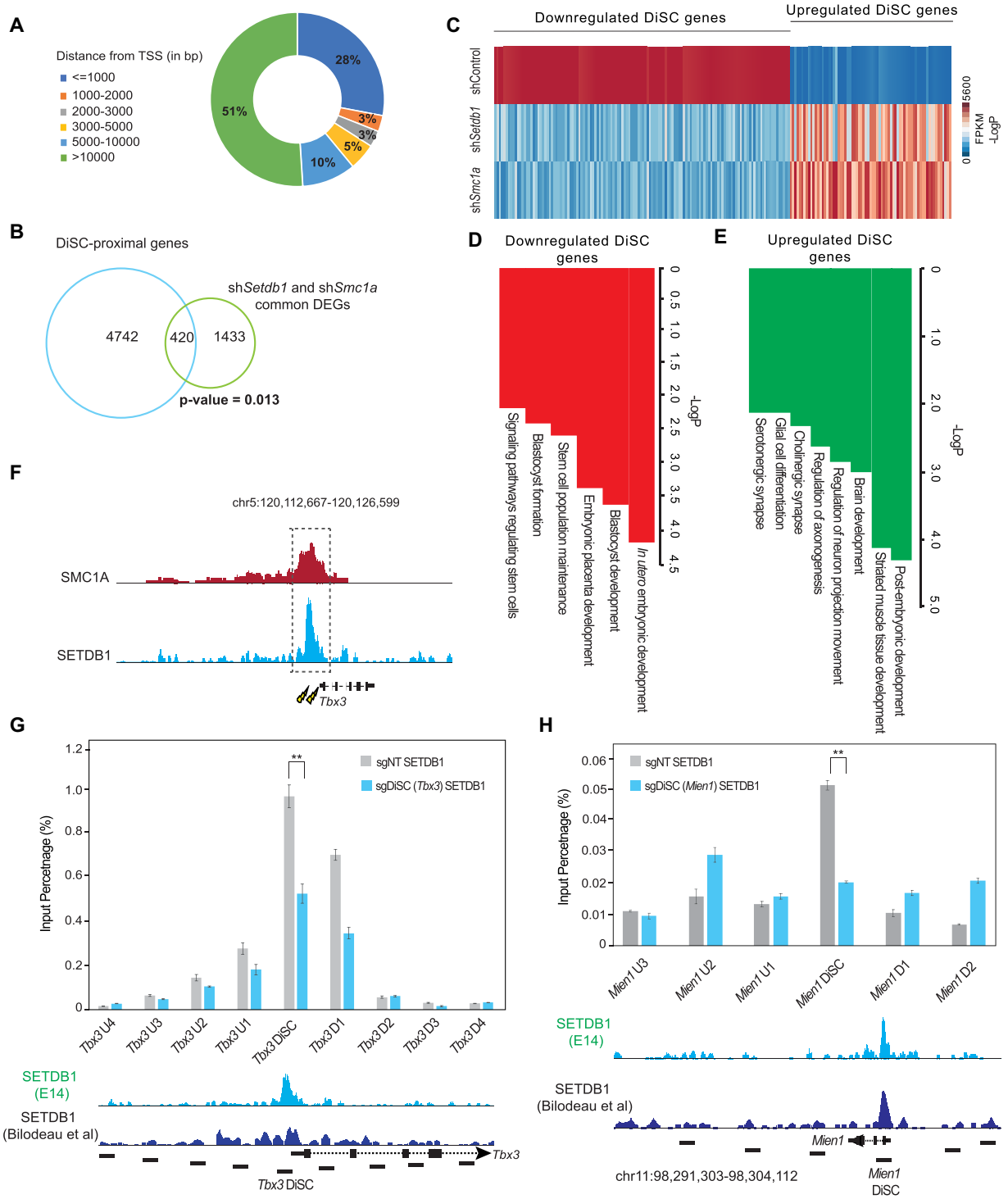

Figure 3. SETDB1 and SMC1A at DiSCs co-regulate transcriptional profiles. (**A**) Distribution of DiSCs based on their distance to the nearest transcription start site (TSS) delineating either a promoter proximal or distal profile for these sites. A percentage (28%) of the DiSCs were proximal to gene TSS and a majority were distal to promoters (51%). (**B**) Venn diagram representing the overlap of commonly dysregulated genes for *Setdb1* and *Smc1a* depletion with DiSC-proximal genes – revealing 420 such statistically significant genes (hypergeometric test, *P*-value = 0.013). (**C**) Heat map delineating the DEGs that are seen to be commonly dysregulated by *Setdb1* and *Smc1a*, at DiSCs. Red represents higher expression and blue represents lower expression. The values depicted are FPKM values from RNA-Seq data. (**D**) Gene ontology (GO) analysis for downregulated DiSC-proximal genes. Values represented are negative logarithm for *P*-values of each GO term. (**E**) Gene ontology (GO) analysis for upregulated DiSC-proximal genes. Values represented are negative logarithm for *P*-values of each GO term. (**F**) Schematic visualizing the CRISPR-Cas9 based targeting of SETDB1-SMC1A bound sites at DiSCs, proximal to the *Tbx3* gene. (**G**) Locally depleted SETDB1 enrichment at targeted *Tbx3* DiSC site, with little or no impact on SETDB1 binding around the target site ($n = 3$, error bars represent standard deviation). Two-tailed t-test was used for statistical analysis. [*P*-value < 0.05, ***P*-value < 0.01]. (**H**) Locally depleted SETDB1 enrichment at targeted *Mien1* DiSC site, with little or no impact on SETDB1 binding around the target site ($n = 3$, Error bars represent standard deviation). Two-tailed *t*-test was used for statistical analysis. [*P*-value < 0.05, ***P*-value < 0.01].

USF2, are also known to be implicated in processes regulating cell maintenance and pluripotency (65,71).

In order to show a direct correlation between DiSCs and their target gene expression, two DiSCs proximal to the genes, *Tbx3* (Figure 3F) and *Mien1*(Supplementary Figure S3L), were targeted for knock-out (KO) using CRISPR-Cas9. After disrupting the SETDB1-binding site for the targeted DiSC, the expression of the proximal gene (*Tbx3*, *Mien1*) was significantly downregulated in accordance with the trends previously observed in the KD RNA-Seq libraries (Supplementary Figure S3L, M). Moreover, the disruption of the proximal DiSC site contributed to gene dysregulation via a localized decline in enrichment of SETDB1 at the targeted site (Figure 3G, H), thereby indicating DiSCs to be gene expression regulatory elements functioning via the presence of SETDB1 and Cohesin. Our gene annotation proximal to the DiSCs had also discovered several key pluripotency regulators (*Esrrb*, *Fgf4*, *Klf4*, *Prdm1*, *Tbx3*) as DiSC-proximal genes. To ascertain the impact of these DiSCs, we performed CRISPR-Cas9 targeting the SETDB1, SMC1A binding sites for the pluripotency-specific DiSC sites. A measurement of a panel of prominent pluripotency genes upon disruption of these DiSCs revealed an impact on the expression of several key genes such as *Sall4*, *Oct4*, *Klf4*, *Esrrb*, *Prdm1*, thereby indicating that these DiSCs and their modes of transcriptional regulation can have larger consequences on the pluripotent cell state (Supplementary Figure S3N).

SETDB1 and Cohesin binding are mutually interdependent and critical for genome topology at DiSCs

Having ascertained the interactions between SETDB1 and Cohesin at the DiSCs and their combined impact on proximal gene dysregulation, we next wished to explore the nature of dependency between these two 'unlikely' partners. Since SMC1A binding was seen to be stronger at SETDB1-bound genomic regions, we performed a genome-wide analysis of SMC1A binding after depletion of *Setdb1*. Interestingly, we observed a drastic decrease of Cohesin binding at the DiSCs after the knock-down of *Setdb1* (Figure 4A). The reduction in SMC1A binding upon *Setdb1* depletion was confined to the non-H3K9me3 cluster of SETDB1-bound sites – which constituted the DiSCs (Supplementary Figure S4A). The decrease in SMC1A in sh*Setdb1*-treated cells was also largely independent of CTCF co-binding, as DiSCs with and without CTCF both showed a decline in SMC1A levels (Supplementary Figure S4B). At the impacted DiSCs, the decrease in SMC1A enrichment was seen to be at least 2-fold for sites that were regarded as significant (Figure 4B, C). An observable decrease in SMC1A levels after *Setdb1* KD was also observed across other tested sites such as SMC1A-H3K9me3 and SMC1A-CTCF co-bound sites, however the decreased SMC1A enrichment was most pronounced across the DiSCs (Supplementary Figure S4C). Despite the decreased SMC1A enrichment upon *Setdb1* depletion, there was no change in the levels of gene expression for the Cohesin subunits *Smc1a*, *Smc1b*, *Smc3* (Supplementary Figure S4D). ChIP-qPCR based validation experiments further showed both Cohesin complex sub-units, SMC1A and RAD21 as having reduced enrichment at the tested DiSCs after *Setdb1* KD (Figure 4D and Supplementary Figure S4E). At several other control sites (SMC1A positive controls, ERVs, SETDB1 positive controls), no significant changes in SMC1A enrichment were seen in sh*Setdb1*-treated cells as compared to the shControl cells (Supplementary Figure S4F).

SETDB1 activity is stabilized in cells by the protein ATF7IP—which sequesters SETDB1 and protects it from proteosomal degradation (6,7). Therefore, SETDB1 and ATF7IP co-exist and function together at a molecular level. To associate this with the SETDB1-SMC1A relationship, we performed KD for *Atf7ip*, and observed a decline in the enrichment levels of SETDB1, SMC1A and RAD21 at tested DiSC sites (Figure 4E, Supplementary Figure S4G, H). This clearly implied that the ATF7IP–SETDB1–Cohesin cascade existed as a determinant of Cohesin binding across the genome and thereby could be essential for gene regulation and genome topology.

Cohesin is specifically enriched at TADs (topologically associating domains) and other topological structures across the genome (72). Since the enrichment of SMC1A at DiSCs is a function of SETDB1 levels, we next explored whether the presence or absence of SETDB1 and SMC1A was influential in regulating topological interactions at the DiSCs, by performing a 3C-qPCR assay. We identified two DiSC-proximal genes *Hoxc13* and *Ppp1r1b* that were seen to be dysregulated in sh*Setdb1* and sh*Smc1a* RNA-Seq samples (Supplementary Figure S4I), and used them to probe the topological interactions in their vicinity. Upon the depletion of *Setdb1* or *Smc1a*, we observed a drastic reduction in the prevalence of genomic interactions around the vicinity of these tested sites (Figure 4F, Supplementary Figure S4J, K). The presence of a proximal DiSC (indicated by SETDB1, SMC1A binding profiles) could be clearly seen for the tested sites, thereby asserting that the changes in localized expression were accompanied by topological changes in the vicinity (Figure 4G, Supplementary Figure 4L). Hi-C data clearly displayed interactions around these two tested gene loci and showed presence of loops dictated by Cohesin (as per public Cohesin Hi-ChIP data (49)) (Supplementary Figure S4M, Figure S4N). These findings cumulatively indicated that the SETDB1–Cohesin partnership was crucial for dictating genome topology at the DiSCs, and thereby was also critical for gene expression regulation. We also carried out CRISPR-Cas9 based targeted disruption of a DiSC proximal to the *Tbx3* gene (Figure 4H), which is critical for pluripotency, by targeting the SETDB1, SMC1A binding sites. Following the disruption of the DiSC, 3C was performed to survey the interactions in the vicinity of the targeted site. Interaction frequencies around fragment 1 (F1) and F4 (in the vicinity of the *Tbx5* gene), as well as F7, F9 (proximal to the *Rbm19* gene and close to another DiSC) were seen to be significantly reduced after disruption of the *Tbx3* DiSC (Figure 4I).

DiSCs are enriched exclusively with SETDB1 independent of its typical methyltransferase activity

Along with SETDB1, several other HMTs such as the HP1, G9A and SUV39H1 are often seen to be associated with the H3K9me2/H3K9me3 marks and therefore possess overlaps

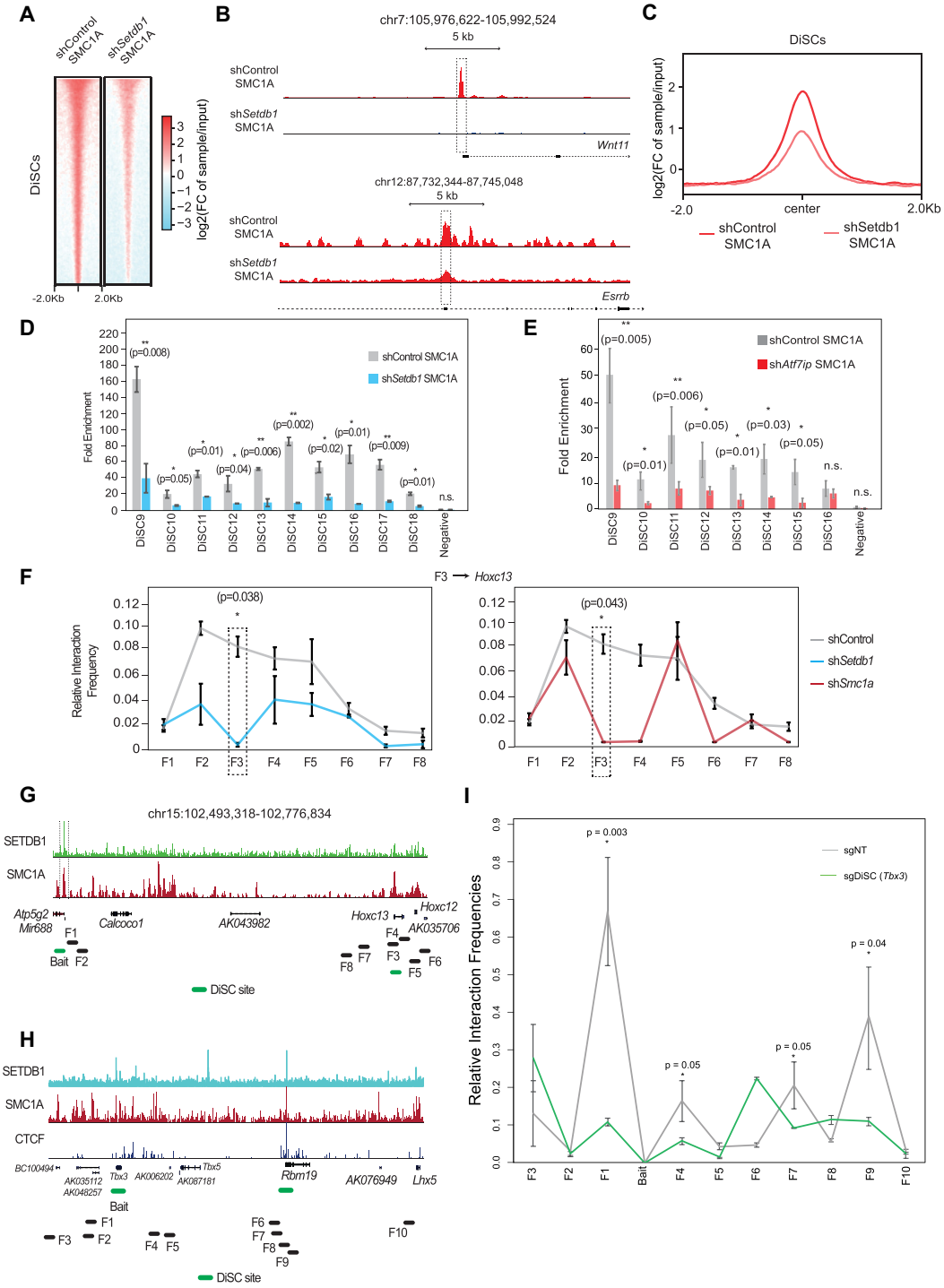

Figure 4. SETDB1 dictates Cohesin levels at the DiSCs. (**A**) Heat map delineating the decreased levels of SMC1A upon *Setdb1* KD across the DiSCs. (**B**) UCSC view of the drastic decline of SMC1A (Cohesin) binding after KD of *Setdb1* at the visualized DiSCs. (**C**) Average enrichment profile for SMC1A enrichment in shControl and sh*Setdb1* cells representing ∼2-fold decrease in Cohesin levels across DiSCs upon SETDB1 KD. (**D**) ChIP-qPCR based validation of the decrease in SMC1A binding at the DiSCs after KD of *Setdb1* ($n = 3$, Error bars represent standard deviation). Two-tailed t-test was used for statistical analysis. [P-value < 0.05, $^{**}P$-value < 0.01]. (**E**) ChIP-qPCR based validation of the decrease in SMC1A binding at the DiSCs after KD of *Atf7ip* ($n = 3$, Error bars represent standard deviation). Two-tailed t-test was used for statistical analysis. [P-value < 0.05, $^{**}P$-value < 0.01]. (**F**) 3C-qPCR interaction profiles around *Hoxc13* genomic locus which is located proximal to a DiSC, upon KD of *Setdb1* (left) and *Smc1a* (right). ($n = 3$, Error bars represent standard deviation). Two-tailed t-test was used for statistical analysis. [$^{*}P$-value < 0.05, $^{**}P$-value < 0.01]. (**G**) UCSC view of the DiSC site (proximal to the *Mir688* gene) and 3C-qPCR amplified fragments in the vicinity of the *Hoxc13* gene. (**H**) UCSC view of the DiSC sites (proximal to the *Tbx3* and *Rbm19* genes) and 3C-qPCR amplified fragments in the vicinity of these DiSCs. (**I**) 3C-qPCR interaction profiles around the *Tbx3* DiSC site, after CRISPR-Cas9 based targeting of the SETDB1, Cohesin binding site of the DiSC. sgNT was used as a baseline control using a non-targeting sgRNA construct ($n = 3$, Error bars represent standard deviation). Fragments F1, F4, F7 and F9 showed a decline in interaction frequencies upon disruption of the DiSC. Two-tailed t-test was used for statistical analysis. [$^{*}P$-value < 0.05, $^{**}P$-value < 0.01].

with SETDB1 in terms of their canonical functions (35,73–76). We queried public datasets to find out whether these alternative HMTs were also existing at the DiSCs. We found that neither SUV39H1 nor G9A (35,36) were found to be enriched across the DiSCs (Supplementary Figure S5A). We also tested for the levels of the alternative repressive H3K9me2 mark at the DiSCs and found it to be absent. Moreover, H3K9me2 levels at the DiSCs ($n = 62$) were unperturbed even upon sh*Setdb1* or sh*Smc1a* treatment of the cells (Figure 5A, Supplementary Figure S5B). Similar testing of the G9A levels at the DiSCs after sh*Setdb1* or sh*Smc1a* treatment also revealed no changes at the DiSCs (Figure 5B, Supplementary Figure S5C), clearly suggesting that the function of SETDB1 at DiSCs was highly specific and could not be replaced by other HMTs. At certain control regions that were dominated by SETDB1 and H3K9me3 (*Polrmt*), we could see a slight elevation in both H3K9me2 and G9A levels especially upon sh*Setdb1* treatment to the cells (Supplementary Figure S5D). SETDB1 often partners with and is recruited to the genome by the HP1 cluster of proteins in order to mark heterochromatin with its typical repressive mark (3,60). However, we also found the HP1 proteins (HP1a, HP1b, HP1g) (38) to be absent at the DiSCs (Figure 5C), yet again reiterating the unique roles of SETDB1 at the DiSCs. SETDB1 has also often been closely associated to the repressive PRC2 complex that deposits the H3K27me3 repressive mark (10,77). Although we did see some minimal binding for PRC2 components such as EZH2 and SUZ12 (37) across DiSCs, this was not comparable to the levels of SETDB1 at the DiSCs (Supplementary Figure S5E). Moreover, typical annotated H3K27me3 repressive peaks did not significantly overlap with our DiSCs and the overlap of the genes constituted within H3K27me3 domains with DiSC-proximal dysregulated genes was also highly insignificant (Supplementary Figure S5F, G). Therefore at large, the DiSCs were independent of other HMTs, conferring SETDB1 as their primary determinant.

Since SETDB1 fundamentally is a methyltransferase, we next wanted to assess whether its binding and functions at the DiSCs were potentially linked to its methyltransferase activity. To assess this, we first generated a deletion mutant for SETDB1, that was catalytically inactive and did not possess the SET domain (78,79) (*mutSET*). As a control, the wild-type SETDB1 ORF was also cloned out into the same backbone (pCAG-HA-hygromycin). We next checked for the nuclear localization of these cloned constructs and could indeed see that both the SETDB1 and mutSET proteins localized to the nucleus (Supplementary Figure S5H). Following over-expression of these constructs in E14 mESCs, we performed HA ChIP to assess if both the SETDB1 and mutSET proteins could localize to the DiSCs. We noticed that both SETDB1 and mutSET were strongly able to retain their binding across DiSCs (Figure 5D, E). Although at sites such as ERVs where SETDB1 performs its typical repressive methyltransferase functions, interestingly only the exogenous SETDB1 protein retained its binding while mutSET binding was lost (Figure 5F, Supplementary Figure S5I). This drastic difference indicated that the lack of the SET domain impaired binding of SETDB1 to the genome at typical heterochromatin regions, while there was no impact in binding at the DiSCs. To assess the linkage of SETDB1 and SMC1A binding broadly with methyltransferase activity, we used the broad HMT inhibitor—3-deazaneplanocin dihydrochloride (DZNep) (80–82) on mESCs to inhibit methylation by all HMTs (particularly targeting H3K27me3 and H3K9me3 depositing enzymes). Post-DZNep treatment, we performed SETDB1 and SMC1A ChIP across the DiSCs and did not see any substantial differences (Figure 5G, H). H3K9me3 levels across DZNep-treated and control (DMSO-treated) cells also did not show any significant differences and sustained low levels of H3K9me3 across DiSCs (Supplementary Figure S5J).

We then decided to test the functions of the SETDB1 and mutSET ORFs independently by using a *Setdb1*-depletion system followed by rescue using the ORFs. For this purpose, cells were treated with the SETDB1-inhibitor mithramycin A (83,84). After inducing *Setdb1* depletion, we transfected the cells with the SETDB1, mutSET ORFs as well as an empty vector control (pCAG-HA). When *Setdb1* expression levels were assayed post-harvest for these rescue experiments, we could see that the pCAG-HA control sustained lower levels of *Setdb1*, whereas both SETDB1, mutSET rescued the *Setdb1* transcript levels (Supplementary Figure S5K). Next, we profiled the expression of the downregulated and upregulated DiSC-proximal genes that were seen to be originally dysregulated upon perturbation of SETDB1 and SMC1A binding at the DiSCs. We observed that *Setdb1*-depleted cells that were rescued with both SETDB1 and mutSET ORFs, were able to rescue the expression of majority of the downregulated DiSC-proximal genes (Figure 5I, Supplementary Figure S5L). Similarly, the upregulated DiSC-proximal gene levels were lowered in both the SETDB1, mutSET ORF rescue systems, with only a small cluster of genes not being rescued in terms of their expression (cluster 2) (Figure 5J, Supplementary Figure S5M). Interestingly, when we assessed the same rescue system for ERV transcripts, it was clearly seen that only SETDB1 repressed the levels of these transcript, while mutSET was ineffective in doing so (Supplementary Figure S5N). Using the same rescue system, we also performed ChIP-qPCR for SMC1A across the DiSCs ($n = 62$). In line with the gene expression rescue, we were able to observe the binding of SMC1A being retained at the DiSCs in both the SETDB1 and mutSET rescue systems (Figure 5K), while no changes in SMC1A binding were observed at control regions (ERVs, SMC1A positive controls, SETDB1 positive controls) (Supplementary Figure S5O). Lastly, we also designed an shRNA construct targeting the 3′UTR of the *Setdb1* gene (sh*Setdb1**), so as to exclusively deplete only the endogenous *Setdb1* levels. Using this construct, we repeated the rescue experiment, where cells were first treated with sh*Setdb1** followed by rescue with the SETDB1, mutSET ORFs, as well as the pCAG-HA control. In conjunction with our previous rescue trials, we saw both SETDB1, mutSET rescuing the downregulated and upregulated DiSC-proximal genes (Figure 5L). There was also seen to be rescue of ERV and lineage-specific gene expression with the SETDB1 ORF, whereas the mutSET ORF failed to do so (Supplementary Figure S5P, Figure S5Q), in line with the first set of rescue experiments. Overall these rescue experiments implicated that SETDB1 was not necessarily

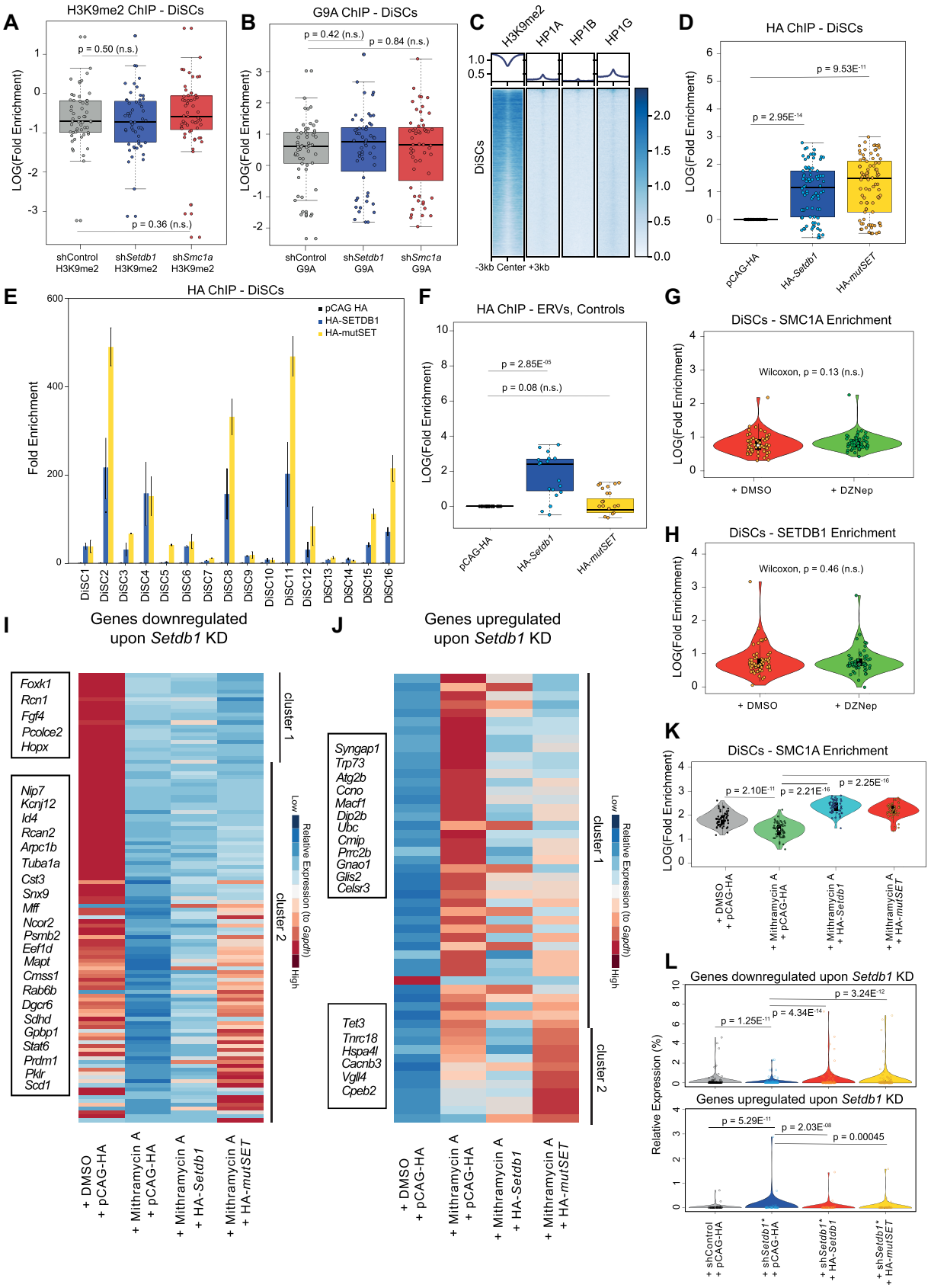

Figure 5. SETDB1 function at the DiSCs is unique and independent of its methyltransferase activity. (**A**) Boxplot depicting non-significant changes in H3K9me2 levels across DiSCs ($n = 62$) in sh*Setdb1* (Wilcoxon test, *P*-value = 0.50), sh*Smc1a*-treated cells (Wilcoxon test, *P*-value = 0.36) as compared

in need of its methyltransferase activity for associating with Cohesin and binding to the DiSCs.

DiSCs comprise critical topological structures in the genome dictated by SETDB1

Overlaying DiSCs upon Cohesin Hi-ChIP data for the E14 mouse ES cell line implicated the DiSCs in TAD and loop architecture (Figure 6A, Supplementary Figure S6A). To further probe the roles of SETDB1 and Cohesin at the DiSCs, we prepared Hi-C libraries after KD of *Setdb1* and *Smc1a*, with the shControl library as a wild-type control. Strikingly, it was seen that the depletion of SETDB1 impacted TADs across the genome adversely, while expectedly SMC1A (Cohesin) loss depleted topological domains drastically (Supplementary Figure S6B). Both SETDB1 and SMC1A loss also depleted topological interactions at overall DNA loops and adversely impacted the loops located at DiSCs (Figure 6B, Supplementary Figure S6C) (Supplementary Table S6).

The loop anchors that are located at DiSCs concurrently show depletion of SMC1A enrichment after *Setdb1* depletion, thereby connecting the loss of topology at DiSCs with the SMC1A depletion upon *Setdb1* KD (Figure 6C). DiSC-proximal genes were seen to be upregulated as well as downregulated upon depletion of *Setdb1* and *Smc1a*. Typically, *Setdb1* depletion is known to cause widespread upregulation of genes as a result of the ablation of its deposited H3K9me3 mark. However, the duality of gene expression at DiSCs via SETDB1 and Cohesin was a unique defining trait for the DiSCs. In order to study this further, we profiled the active (A) and inactive (B) topological compartment profiles in our Hi-C datasets, and observed A-B and B-A compartment switches between shControl and sh*Setdb1* Hi-C datasets (Figure 6D, Supplementary Figure S6D). As a comparison, sh*Smc1a* Hi-C data also demonstrated widespread compartment switches, as has been known to occur during Cohesin depletion (85,86) (Supplementary Figure S6E). The topological compartment switches expectedly led to changes in gene expression with upregulation and downregulation of several genes being observed across the sh*Setdb1*-treated cells (Figure 6E) (Supplementary Table S7). We further correlated the compartment switches and gene expression changes with changes in insulating neighbourhoods around genomic loops and indeed we could observe gain and loss of insulating boundaries in both the sh*Setdb1* and sh*Smc1a* treated cells (Figure 6F, G). These insulating boundaries which are typical features of topologically interacting regions, were also dominantly present across the DiSCs as compared to typical SETDB1, H3K9me3-bound regions (Figure 6H), thus reiterating the topological significance of the DiSCs. It was also noted that the overall SMC1A peak number remained consistent between shControl and sh*Setdb1* treated cells across the compartment switches in the genome, indicating that the strong peak number correlation was maintained despite Cohesin loss upon *Setdb1* depletion (Supplementary Figure S6F).

We observed, that there was a substantial overlap between the interactions (dots) lost in the sh*Setdb1* and sh*Smc1a* treated cells, strongly suggesting co-regulation of topology by SETDB1, Cohesin (Figure 6I). Since we had previously tested the vicinity of the DiSC proximal to the *Tbx3* gene, representing localized changes in topology upon disruption of the DiSC, we looked at the same site in terms of our Hi-C datasets and found extensive changes in chromatin loops between the shControl and

←

to shControl cells. (**B**) Boxplot depicting non-significant changes in HMT G9A levels across DiSCs ($n = 62$) in sh*Setdb1* (Wilcoxon test, *P*-value = 0.42), sh*Smc1a*-treated cells (Wilcoxon test, *P*-value = 0.84) as compared to shControl cells. (**C**) Heat map representation of the lack of H3K9me2 and HP1 (HP1A, HP1B, HP1G) binding across the DiSCs. (**D**) HA ChIP-qPCR across DiSCs ($n = 62$) for cells transfected with pCAG-HA (control) and pCAG-HA-*Setdb1*, pCAG-HA-*mutSET* ORFs, represented as a boxplot. Values are double normalized to a negative control site as well as to the pCAG-HA control sample. Enrichment is retained across DiSCs for both the HA-SETDB1 (Wilcoxon test, *P*-value = $2.95e^{-14}$) and HA-mutSET (Wilcoxon test, *P*-value = $9.53e^{-11}$) proteins, with the latter showing higher enrichment. (**E**) Bar graph representation of HA ChIP qPCR for DiSCs showing retained binding of HA-SETDB1 and HA-mutSET across the DiSCs as compared to the pCAG-HA control ($n = 3$, Error bars represent standard deviation). (**F**) HA ChIP-qPCR across control primers (ERVs, SETDB1 sites) for cells transfected with pCAG-HA (control) and pCAG-HA-*Setdb1*, pCAG-HA-*mutSET* ORFs, represented as a boxplot. Values are double normalized to a negative control site as well as to the pCAG-HA control sample. Enrichment is retained across DiSCs for only the HA-SETDB1 (Wilcoxon test, *P*-value = $2.85e^{-05}$) whereas HA-mutSET binding is ablated (Wilcoxon test, *P*-value = 0.08, n.s.). (**G**) Violin plot visualization of Cohesin (SMC1A) enrichment levels across DiSCs ($n = 62$) in DMSO-treated versus HMT inhibitor DZNep-treated E14 mESCs, depicting no significant changes in enrichment (Wilcoxon test, *P*-value = 0.13). (**H**) Violin plot visualization of SETDB1 enrichment levels across DiSCs ($n = 62$) in DMSO-treated versus HMT inhibitor DZNep-treated E14 mESCs, depicting no significant changes in enrichment (Wilcoxon test, *P*-value = 0.46). (**I**) Heat maps representing the qPCR-based rescue profile for DiSC-proximal downregulated genes, comparing DMSO, pCAG-HA treated control cells with mithramycin A-treated cells rescued for *Setdb1* expression with pCAG-HA, pCAG-HA-*Setdb1* and pCAG-HA-*mutSET*. Expression of both *Setdb1* and *mutSET* rescues majority of the downregulated DiSC genes (cluster 1), with a minority not being rescued effectively by *mutSET* expression (cluster 2). Red represents higher gene expression and blue represents low gene expression. All expression is normalized to *Gapdh* expression. (**J**) Heat maps representing the qPCR-based rescue profile for DiSC-proximal upregulated genes, comparing DMSO, pCAG-HA treated control cells with mithramycin A-treated cells rescued for *Setdb1* expression with pCAG-HA, pCAG-HA-*Setdb1* and pCAG-HA-*mutSET*. Expression of both *Setdb1* and *mutSET* rescues majority of the upregulated DiSC genes (cluster 1), with a minority not being rescued effectively by *mutSET* expression (cluster 2). Red represents higher gene expression and blue represents low gene expression. All expression is normalized to *Gapdh* expression. (**K**) Comparison of SMC1A levels across DiSCs ($n = 62$) across DMSO, pCAG-HA treated control cells with mithramycin A-treated cells rescued for *Setdb1* expression with pCAG-HA, pCAG-HA-*Setdb1* and pCAG-HA-*mutSET*. Mithramycin A-treated, pCAG-HA control cells showed a sustained reduction in SMC1A levels (Wilcoxon test, *P*-value = $2.10e^{-11}$) as compared to the DMSO-treated control. Post-mithramycin A treatment, expression of *Setdb1* (Wilcoxon test, *P*-value = $2.21e^{-16}$) and *mutSET* (Wilcoxon test, *P*-value = $2.25e^{-16}$) both showed a significant recovery and rescue of Cohesin enrichment. (**L**) Violin plots depicting the rescue of DiSC-proximal downregulated (top) and upregulated (bottom) genes using an alternative *Setdb1* depletion and rescue system. Cells depleted for endogenous *Setdb1* (sh*Setdb1**-treated cells) are rescued with over-expression of pCAG-HA, pCAG-HA-*Setdb1* and pCAG-HA-*mutSET*. Cells treated with shControl, pCAG-HA are used as a baseline control. Downregulation of DiSC genes is seen in sh*Setdb1**, pCAG-HA-treated cells (Wilcoxon test, *P*-value = $1.25e^{-11}$) and expression is elevated and rescued by the HA-SETDB1 (Wilcoxon test, *P*-value = $4.34e^{-14}$) and HA-mutSET (Wilcoxon test, *P*-value = $3.24e^{-12}$) proteins. Similarly, upregulation of DiSC genes is seen in sh*Setdb1**, pCAG-HA-treated cells (Wilcoxon test, *P*-value = $5.29e^{-11}$) and expression is lowered and rescued by the HA-SETDB1 (Wilcoxon test, *P*-value = $2.03e^{-08}$) and HA-mutSET (Wilcoxon test, *P*-value = 0.00045) proteins.

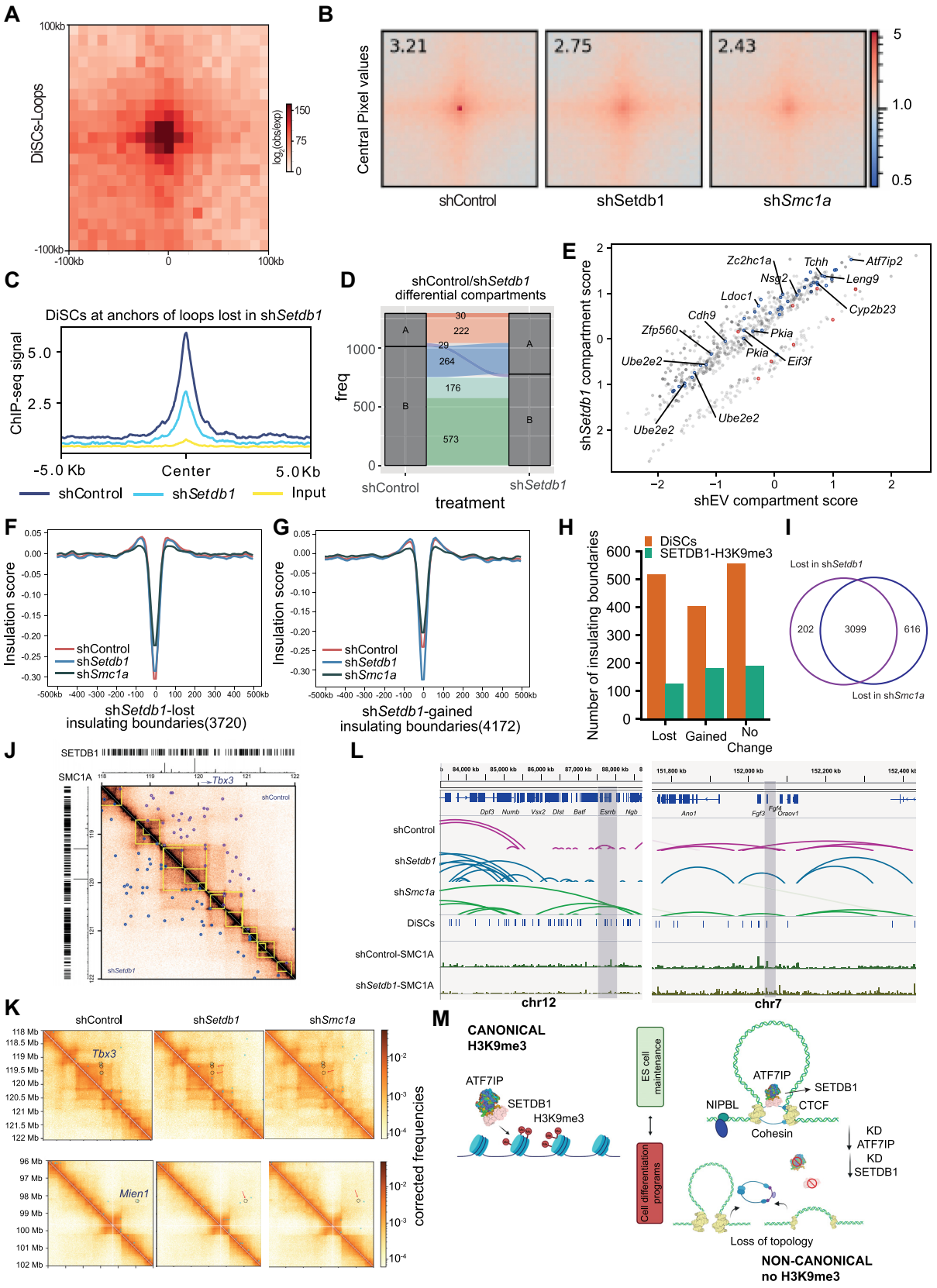

Figure 6. SETDB1–Cohesin constitute a topological axis that regulates gene expression. (**A**) Overlap of significant loops detected in Cohesin HiChIP with DiSC sites at a 10 kb resolution. (**B**) Pile-up analysis showing depletion of DiSC-associated chromatin loops upon KD of *Setdb1* and *Smc1a* using Hi-C

sh*Setdb1*-treated cells (Figure 6J, Supplementary Figure S6G). We could also see certain common patterns in the loop domains that suffered losses in the vicinity of the *Tbx3* and *Mien1* DiSCs (Figure 6K, Supplementary Figure S6H). We also recorded additional localized changes in topology at DiSCs proximal to the *Fgf4* and *Esrrb* genes, with decreased localized chromatin loops being seen especially in the sh*Setdb1*-treated cells (Figure 6L). Therefore the overall impact of *Setdb1* depletion on genome topology at the DiSCs included the disruption of insulating boundaries, switches between topological compartments and accompanying changes in gene expression—all of which substantiated our hypothesis about the DiSCs being critical topological domains that are chiefly dictated by SETDB1.

DISCUSSION

We have identified through our extensive analysis of SETDB1 ChIP-Seq data—novel regions known as DiSCs which are bound by SETDB1 and Cohesin. These unique regions were seen to be proximal to genes and bereft of all known repressive histone marks, implying their significantly different functions with respect to SETDB1 and its known roles. The strong co-occupancy of SETDB1 and Cohesin at these sites suggested their inter-dependency and indeed it was strikingly seen that Cohesin enrichment at these sites was heavily reliant on SETDB1 binding. Furthermore, depletion of ATF7IP, which is an upstream regulator and stabilizer for SETDB1, also cascaded on to depletion of SMC1A and the Cohesin complex at large. Therefore, we postulated the existence of a SETDB1–Cohesin regulatory system, wherein SETDB1 functioned as an upstream regulator of Cohesin at the DiSCs. We also set up extensive rescue systems using a wild-type and catalytically inactive mutant for SETDB1, post-endogenous cellular depletion of the same. DiSC-proximal gene expression as well as Cohesin binding at the DiSCs was seen to be rescued by both the wild-type as well as catalytically inactive mutant SETDB1. This was indicative of the fact that the association of SETDB1 at the DiSCs and with Cohesin was independent of its methyltransferase activity, and the removal of the catalytic SET domain did not have adverse effects on the DiSCs. In contrast, the inactive mutant SETDB1 displayed vastly different activities at its canonical repressive domains (ERVs, lineage-specific repressed genes) further reinforcing the difference in the mechanistic nature of SETDB1 at the DiSCs. The inactive mutant also sustained its binding at the DiSCs and as a result also maintained Cohesin levels at these sites. In totality, these findings suggest a methyltransferase-independent mechanism by which SETDB1 tethers Cohesin to the DiSCs. There is a likelihood of other partner proteins involved in this function along with SETDB1 as one of the chief architects for the same. Further studies would be needed to divulge the protein regulome that maintains and functions at the DiSCs.

This widespread impact on DiSC-proximal gene expression as well as localized genome topology after *Setdb1* depletion in the cells, which are likely effects of the decline in localized Cohesin levels, further strengthened the influence of the SETDB1–Cohesin pairing at the DiSCs. KD *Setdb1* Hi-C data revealed the functional significance of this SETDB1–Cohesin regulatory axis by establishing DiSCs as significant topological structures that are impacted upon *Setdb1* loss, leading to gene expression regulation. Overall, we have been able to profile a unique regulatory relationship between SETDB1 and Cohesin, thereby leading to the understanding that SETDB1 is a critical regulator of cell fates and cell lineage, not just via its deposited repressive H3K9me3 marks, but also by virtue of exerting control on Cohesin association to the genome. By combining with Cohesin, SETDB1 establishes the DiSCs as unique topological and functional domains that are crucial on the one hand in maintaining the stem cell state, but can also serve as triggers for differentiation and cellular transformation to alternative lineages (neurons, muscles etc.).

SETDB1 has not been extensively studied, outside of its conventional histone methyltransferase function. A recent report indicated that SETDB1 could be located proximally to a large neuron-specific topological domain and be potentially involved in its regulation (13), however this too was linked to its H3K9me3-dependent role. The unique nature of the DiSCs stems from the fact that they are free of SETDB1-associated repressive histone marks such as H3K9me3 and H3K27me3, and are rather co-occupied by

←

datasets. Value of central pixel was used for display. The enrichment of interactions represents a 5kb resolution for the Hi-C data. (**C**) Average profile representing decrease in SMC1A ChIP-seq binding signal DiSCs which are associated with lost chromatin interactions after *Setdb1* KD. (**D**) Alluvial plots delineating compartment switches between shControl and sh*Setdb1* cell states. The numbers of differential compartments between cell states were represented. (**E**) Scatter plot showing the presence of upregulated genes (blue) and downregulated genes (red) in differential compartments that arise because of sh*Setdb1* treatment, contrasted with shControl based on Hi-C data inferred compartmental switches. (**F**) Average profile of insulation score in high confidence lost insulating boundaries upon KD of *Setdb1* identified by cooltools. (**G**) Average profile of insulation score in high confidence gained insulating boundaries upon KD of *Setdb1* identified by cooltools. (**H**) Barplot for number of differential insulating boundaries compared between DiSCs and SETDB1-H3K9me3 regions for each category of insulating boundaries (gained/lost/unchanged) upon KD of *Setdb1*. (**I**) Venn diagram representing the large overlap between interactions (dots) lost in sh*Setdb1* and shSmc1a-treated cells. sh*Setdb1* has 3301 lost dots, sh*Smc1a* has 3715 lost dots, and the number of common lost dots is 3099. (**J**) GENOVA-based visualization of changes in overall chromatin looping (purple loops for shControl cells, blue loops for sh*Setdb1* cells), as well as TAD structure (yellow boxes) across the *Tbx3* gene proximal DiSC site. Tracks on the right side represent SMC1A bound peaks and tracks on top represent the SETDB1-bound peaks. (**K**) Contact matrix heatmaps around example genes *Tbx3* and *Mien1*. Dots called by cooltools - representing chromatin loops, are visualized using cyan-coloured rectangles. Dashed circles depict the topological sites involved in chromatin loops around the DiSCs in the vicinity of the represented genes. (**L**) IGV browser showing the chromatin looping change at DiSCs proximal to *Esrrb* and *Fgf4* separately, upon KD of *Setdb1* and *Smc1a*. The DiSCs, SMC1A ChIP-seq before and after *Setdb1* KD are also represented as tracks. (**M**) Schematic representation of bi-modal regulation of the genome by SETDB1. This non-canonical model (right) depicts SETDB1 (stabilized by ATF7IP) associating with the Cohesin complex at topological structures (DNA loops, TADs) across the genome and exerting control over them by regulating Cohesin binding to DNA (right). In totality, these two modes of SETDB1 activity control genome topology and gene expression, striking a fine balance in the ES state between embryonic stem cell maintenance and differentiation to other lineages (neuronal, muscle etc.).

Cohesin. The co-existence of SETDB1 and SMC1A at these domains is extensive and highly functional, suggesting that they are acting together as a complex.

Many of the DiSCs have unique broad binding profiles which we suspect represents the dynamic binding of SETDB1 and Cohesin to chromatin. Cohesin complex is known to slide and translocate across the length of chromatin (87–90). As such, they typically exhibit broader binding profiles. We have been able to validate this observation for RAD21—another Cohesin sub-unit, whereas this does not seem to be the case for a transcriptional regulator such as CTCF, which binds to a specific motif. It is feasible that the broad binding profile of SETDB1, as an enzyme, might be a consequence of its transient and rapidly changing association to its substrate proteins that are bound across the genome. The broad binding profile of SETDB1 was reproducible across different antibody reagents and a FLAG-*Setdb1* ES cell line we had engineered. For Cohesin, in accordance to the model of loop extrusion, the broad peak profiles could also arise from the transitional nature of the CTCF-Cohesin associations (91). DiSCs are enriched with activating histone marks such as H3K4me3 and hence comprise of transcriptionally active regions, unlike the regular association of SETDB1 with repressive marks (92,93). Therefore, our discovery of the DiSCs demarcates them as unique, functionally relevant regions.

Recent studies have suggested that Cohesin has a prominent role in terms of regulating the 3D genome , and have also implicated genome topology in tissue-specific transcriptional regulatory roles (94). Our analysis of Hi-C, KD Hi-C and HiChIP datasets implied that the DiSCs are indeed topological structures that regulate chromatin architecture. Interestingly, a significant proportion of the DiSCs with disrupted Cohesin binding after *Setdb1* ablation were accompanied by a loss of genome topology, subsequently leading on to proximal gene dysregulation. Out of these dysregulated genes a substantial percentage were commonly impacted by the ablation of *Setdb1* and *Smc1a*. The functions of the commonly downregulated functional DiSC genes were seen to be related to stem cell maintenance and embryonic development as well as other metabolic and cellular processes such as RNA and DNA metabolism, translation. This aligns with previous findings which suggested that SETDB1 was essential in the maintenance of pluripotency (95–97). Previous findings discussed SETDB1-regulated repression of lineage-specific or developmental genes by deposition of the H3K9me3 mark (98). The regulation of pluripotency specific genes, pathways and processes via topological domains such as DiSCs adds a new dimension to its role in the regulation chromatin architecture. Upregulated DiSC proximal genes, on the other hand had lineage specific functions such as muscle and neuronal lineage development. On a functional basis, the discovery of DiSCs in this study, has thus paved a new mechanistic and topological aspect to gene expression regulation for differentiated cell types (such as neurons) as well as for the maintenance of pluripotency by SETDB1 through its control over Cohesin binding at these domains. The topological link of SETDB1 at the DiSCs was further strengthened by observing a strong loss of topological interactions at DiSC loop anchors upon ablation of SETDB1. The topological impact exerted by SETDB1 via Cohesin also contributed to compartment switches and gene dysregulation as a result.

Together, DiSCs exhibit a non-canonical model for the binding of SETDB1 (Figure 6M) on the genome, without its classical association to repressive histone marks. The regulation of Cohesin binding by SETDB1 at these sites as well as the subsequent impact on gene dysregulation are tightly linked to the local genomic 3D topology and architecture (99). DiSCs also possess lineage specific gene regulation that are chiefly controlled by SETDB1 and Cohesin. Recent studies have shown that histone-specific methyltransferases have been found to methylate non-histone proteins in some instances (100,101). SETDB1 too, has recently been shown to act on P53 as well as AKT as a non-histone protein methyltransferase (14,102,103). We have however indicated that SETDB1 and its association with Cohesin transcends even this atypical methyltransferase activity, since catalytic inactivation of SETDB1 does manage to maintain Cohesin levels at the DiSCs. In the wider context, this study portrays SETDB1 as an articulate and versatile regulator of the genome—with its functions now branching out beyond its typical methyltransferase roles. Moreover, we also identify an upstream regulator of Cohesin in the form of SETDB1, posing a bigger question about the other proteins regulating the maintenance of the DiSCs. However, this biochemical validation of the SETDB1-Cohesin axis needs further experimental validation and probing, and was not within the current functional and topological scope of our exploration of the DiSCs.

Therefore, this study reports the discovery of DiSCs, revealing these previously undescribed domains of the genome that act as localized transcriptional factories, regulating specific lineages and functions via topological structures cumulatively controlled by SETDB1 and Cohesin. The impact of DiSC perturbation on gene expression linked to neural and muscle system development is particularly interesting since it provides an avenue to explore facilitation of differentiation processes by targeting of specific DiSCs. These findings also illustrate how regions like DiSCs could prove to be critical regulators of cell fate and lineage. The broad binding profiles of Cohesin and SETDB1 identified from this study, suggests that transient and dynamic chromatin binders may exhibit such a hallmark in its binding to the genome. Our study also delineates the versatility of SETDB1 beyond its traditional repressive role, suggesting that it dictates chromatin topology and exerts gene expression control across the genome independent of its histone H3K9-methylation capabilities.

DATA AVAILABILITY

All the next-generation sequencing data have been submitted to GEO under the accession GSE123245. All other data can be made available by the corresponding authors upon reasonable request.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Don Loi Xu for his sincere contributions in discussing the data and experiments for this study. We also thank Kelly Yu Sing Tan for her assistance in allowing us to access and work with the mass spectrometry equipment. We sincerely thank Dr Fang Haitong for technical assistance in the study. We also thank Dr Hao Fei Wang, Dr Qiaorui Xing and Dr. Nareshwaran Gnanasegaran and Aloysius Quek for their helpful suggestions as well as technical assistance. We also thank Bobby Tan from the Genome Institute of Singapore for providing the SETDB1 antibody.
Author contributions: T.W., C.E.F., Y.Y.Z. designed and performed the research, analyzed the data and wrote the paper. B.S.Q.H., Q.B., Z.H.Z., X.B. designed and conducted research. H.H.N., D.S.T.O., J.J.H.C., A.S., M.J.F., J.J.C., H.L. and J.X. analyzed data. Y-H.L. designed research, analyzed data and wrote the paper.

FUNDING

H.L. is supported by grants from National Institutes of Health (NIH; R01AG056318); the Glenn Foundation for Medical Research, Mayo Clinic Center for Biomedical Discovery, Center for Individualized Medicine, Mayo Clinic Comprehensive Cancer Center (NIH; P30CA015083), and the David F. and Margaret T. Grohne Cancer Immunology and Immunotherapy Program; Y-H.L. is supported by the NRF Investigatorship award [NRFI2018-02]; JCO Development Programme [1534n00153]; Singapore National Research Foundation under its Cooperative Basic Research Grant administered by the Singapore Ministry of Health's National Medical Research Council [NMRC/CBRG/0092/2015]; A*STAR Biomedical Research Council, Central Research Fund, Use-Inspired Basic Research (CRF UIBR); X.J. was supported by the Department of Biological Sciences, National University of Singapore, during the period that this research was conducted; Biomedical Research Council, Agency for Science, Technology and Research, Singapore. Funding for open access charge: A*STAR Biomedical Research Council, Central Research Fund, Use-Inspired Basic Research (CRF UIBR).
Conflict of interest statement. None declared.

REFERENCES

1. Harte,P.J., Wu,W., Carrasquillo,M.M. and Matera,A.G. (1999) Assignment of a novel bifurcated SET domain gene, SETDB1, to human chromosome band 1q21 by in situ hybridization and radiation hybrids. *Cytogenet. Genome Res.*, **84**, 83–86.
2. Schultz,D.C., Ayyanathan,K., Negorev,D., Maul,G.G. and Rauscher,F.J. (2002) SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.*, **16**, 919–932.
3. Ayyanathan,K., Lechner,M.S., Bell,P., Maul,G.G., Schultz,D.C., Yamada,Y., Tanaka,K., Torigoe,K. and Rauscher,F.J III (2003) Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. *Genes Dev.*, **17**, 1855–1869.
4. Xing,Q.R., El Farran,C.A., Gautam,P., Chuah,Y.S., Warrier,T., Toh,C.X.D., Kang,N.Y., Sugii,S., Chang,Y.T., Xu,J. *et al.* (2020) Diversification of reprogramming trajectories revealed by parallel single-cell transcriptome and chromatin accessibility sequencing. *Sci. Adv.*, **6**, 18.
5. Loh,Y.H., Zhang,W., Chen,X., George,J. and Ng,H.H. (2007) Jmjd1a and jmjd2c histone H3 lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev.*, **21**, 2545–2557.
6. Timms,R.T., Tchasovnikarova,I.A., Antrobus,R., Dougan,G. and Lehner,P.J. (2016) ATF7IP-mediated stabilization of the histone methyltransferase SETDB1 is essential for heterochromatin formation by the HUSH complex. *Cell Rep.*, **17**, 653–659.
7. Basavapathruni,A., Gureasko,J., Porter Scott,M., Hermans,W., Godbole,A., Leland,P.A., Boriack- Sjodin,P.A., Wigle,T.J., Copeland,R.A. and Riera,T.V. (2016) Characterization of the enzymatic activity of SETDB1 and its 1:1 complex with ATF7IP. *Biochemistry*, **55**, 1645–1651.
8. Rowe,H.M., Jakobsson,J., Mesnard,D., Rougemont,J., Reynard,S., Aktas,T., Maillard,P.V., Layard- Liesching,H., Verp,S., Marquis,J. *et al.* (2010) KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*, **463**, 237–240.
9. Yuan,P., Han,J., Guo,G., Orlov,Y.L., Huss,M., Loh,Y.-H., Yaw,L.-P., Robson,P., Lim,B. and Ng,H.-H. (2009) Eset partners with oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. *Genes Dev.*, **23**, 2507–2520.
10. Fei,Q., Yang,X., Jiang,H., Wang,Q., Yu,Y., Yu,Y., Yi,W., Zhou,S., Chen,T., Lu,C. *et al.* (2015) SETDB1 modulates PRC2 activity at developmental genes independently of H3K9 trimethylation in mouse ES cells. *Genome Res.*, **25**, 1325–1335.
11. Gautam,P., Yu,T. and Loh,Y.H. (2017) Regulation of ERVs in pluripotent stem cells and reprogramming. *Curr. Opin. Genet. Dev.*, **46**, 194–201.
12. Toh,C.X.D., Chan,J.W., Chong,Z.S., Wang,H.F., Guo,H.C., Satapathy,S., Ma,D., Goh,G.Y.L., Khattar,E., Yang,L. *et al.* (2016) RNAi reveals phase-specific global regulators of human somatic cell reprogramming. *Cell Rep.*, **15**, 2597–2607.
13. Jiang,Y., Loh,Y.-H.E., Rajarajan,P., Hirayama,T., Liao,W., Kassim,B.S., Javidfar,B., Hartley,B.J., Kleofas,L., Park,R.B. *et al.* (2017) The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain. *Nat. Genet.*, **49**, 1239–1250.
14. Guo,J., Dai,X., Laurent,B., Zheng,N., Gan,W., Zhang,J., Guo,A., Yuan,M., Liu,P., Asara,J.M. *et al.* (2019) AKT methylation by SETDB1 promotes AKT kinase activity and oncogenic functions. *Nat. Cell Biol.*, **21**, 226–237.
15. Merkenschlager,M. and Nora,E.P. (2016) CTCF and cohesin in genome folding and transcriptional gene regulation. *Annu. Rev. Genomics Hum. Genet.*, **17**, 17–43.
16. Kim,Y.J., Cecchini,K.R. and Kim,T.H. (2011) Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene a locus. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 7391–7396.
17. Chernukhin,I., Shamsuddin,S., Kang,S.Y., Bergstrom,R., Kwon,Y.-W., Yu,W., Whitehead,J., Mukhopadhyay,R., Docquier,F., Farrar,D. *et al.* (2007) CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. *Mol. Cell. Biol.*, **27**, 1631–1648.
18. Lutz,M., Burke,L.J., Barreto,G., Goeman,F., Greb,H., Arnold,R., Schultheiss,H., Brehm,A., Kouzarides,T., Lobanenkov,V. *et al.* (2000) Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic. Acids. Res.*, **28**, 1707–1713.
19. Dixon,J.R., Selvaraj,S., Yue,F., Kim,A., Li,Y., Shen,Y., Hu,M., Liu,J.S. and Ren,B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.
20. Handoko,L., Xu,H., Li,G., Ngan,C.Y., Chew,E., Schnapp,M., Lee,C.W.H., Ye,C., Ping,J.L.H., Mulawadi,F. *et al.* (2011) CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.*, **43**, 630–638.
21. Lengronne,A., Katou,Y., Mori,S., Yokabayashi,S., Kelly,G.P., Ito,T., Watanabe,Y., Shirahige,K. and Uhlmann,F. (2004) Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature*, **430**, 573–578.
22. Wang,H.F., Warrier,T., Farran,C.A., Zheng,Z.H., Xing,Q.R., Fullwood,M.J., Zhang,L.F., Li,H., Xu,J., Lim,T.M. *et al.* (2020) Defining essential enhancers for pluripotent stem cells using a features-oriented CRISPR-Cas9 screen. *Cell Rep.*, **33**, 108309.
23. Wendt,K.S., Yoshida,K., Itoh,T., Bando,M., Koch,B., Schirghuber,E., Tsutsumi,S., Nagae,G., Ishihara,K., Mishiro,T.

et al. (2008) Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*, **451**, 796–801.
24. Rao,S.S.P., Huang,S.-C., Glenn,B., Hilaire,S., Casellas,R., Lander,E.S., Lieberman,E. and Correspondence,A. (2017) Cohesin loss eliminates all loop domains.*Cell*, **171**, 305–320.
25. Johannes Stigler,A., Koshland,D.E., Greene,E.C. and Stigler,J. (2016) Single-Molecule imaging reveals a collapsed conformational state for DNA-Bound cohesin article single-molecule imaging reveals a collapsed conformational state for DNA-Bound cohesin. *Cell Rep.*, **15**, 988–998.
26. Bilodeau,S., Kagey,M.H., Frampton,G.M., Rahl,P.B. and Young,R.A. (2009) SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev.*, **23**, 2484–2489.
27. Loh,Y.-H., Wu,Q., Chew,J.-L., Vega,V.B., Zhang,W., Chen,X., Bourque,G., George,J., Leong,B., Liu,J. *et al.* (2006) The oct4 and nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.*, **38**, 431–440.
28. Edgar,R., Domrachev,M. and Lash,A.E. (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic. Acids. Res.*, **30**, 207–210.
29. Noordermeer,D., Leleu,M., Schorderet,P., Joye,E., Chabaud,F. and Duboule,D. (2014) Temporal dynamics and developmental memory of 3D chromatin architecture at hox gene loci. *Elife*, **3**, e02557.
30. Yu,P., Xiao,S., Xin,X., Song,C.-X., Huang,W., McDee,D., Tanaka,T., Wang,T., He,C. and Zhong,S. (2013) Spatiotemporal clustering of the epigenome reveals rules of dynamic gene regulation. *Genome Res.*, **23**, 352–364.
31. Yue,F., Cheng,Y., Breschi,A., Vierstra,J., Wu,W., Ryba,T., Sandstrom,R., Ma,Z., Davis,C., Pope,B.D. *et al.* (2014) A comparative encyclopedia of DNA elements in the mouse genome. *Nature*, **515**, 355–364.
32. Kagey,M.H., Newman,J.J., Bilodeau,S., Zhan,Y., Orlando,D.A., van Berkum,N.L., Ebmeier,C.C., Goossens,J., Rahl,P.B., Levine,S.S. *et al.* (2010) Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, **467**, 430–435.
33. Chen,X., Xu,H., Yuan,P., Fang,F., Huss,M., Vega,V.B., Wong,E., Orlov,Y.L., Zhang,W., Jiang,J. *et al.* (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, **133**, 1106–1117.
34. Yang,B.X., El Farran,C.A., Guo,H.C., Yu,T., Fang,H.T., Wang,H.F., Schlesinger,S., Seah,Y.F.S., Goh,G.Y.L., Neo,S.P. *et al.* (2015) Systematic identification of factors for provirus silencing in embryonic stem cells. *Cell*, 163, 230–245.
35. Bulut- Karslioglu,A., De,La, Rosa- Velázquez,I.A., Ramirez,F., Barenboim,M., Onishi- Seebacher,M., Arand,J., Galán,C., Winter,G.E., Engist,B. *et al.* (2014) Suv39h-Dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol. Cell*, **55**, 277–290.
36. Mozzetta,C., Pontis,J., Fritsch,L., Robin,P., Portoso,M., Proux,C., Margueron,R. and Ait-Si- Ali,S. (2014) The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-Mediated gene silencing. *Mol. Cell*, **53**, 277–289.
37. Peng,J.C., Valouev,A., Swigut,T., Zhang,J., Zhao,Y., Sidow,A. and Wysocka,J. (2009) Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell*, **139**, 1290–1302.
38. Ostapcuk,V., Mohn,F., Carl,S.H., Basters,A., Hess,D., Iesmantavicius,V., Lampersberger,L., Flemr,M., Pandey,A., Thomä,N.H. *et al.* (2018) Activity-dependent neuroprotective protein recruits HP1 and CHD4 to control lineage-specifying genes. *Nature*, **557**, 739–743.
39. Wang,L., Wang,S. and Li,W. (2012) RSeQC: quality control of RNA-seq experiments. *Bioinformatics*, **28**, 2184–2185.
40. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
41. Trapnell,C., Roberts,A., Goff,L., Pertea,G., Kim,D., Kelley,D.R., Pimentel,H., Salzberg,S.L., Rinn,J.L. and Pachter,L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with tophat and cufflinks. *Nat. Protoc.*, **7**, 562–578.
42. Hagège,H., Klous,P., Braem,C., Splinter,E., Dekker,J., Cathala,G., de Laat,W. and Forné,T. (2007) Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat. Protoc.*, **2**, 1722–1733.
43. Doench,J.G., Fusi,N., Sullender,M., Hegde,M., Vaimberg,E.W., Donovan,K.F., Smith,I., Tothova,Z., Wilen,C., Orchard,R. *et al.* (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.*, **34**, 184–191.
44. Rao,S.S.P., Huntley,M.H., Durand,N.C., Stamenova,E.K., Bochkov,I.D., Robinson,J.T., Sanborn,A.L., Machol,I., Omer,A.D., Lander,E.S. *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665–1680.
45. Wang,J., Chakraborty,A. and Ay,F. (2021) dcHiC: differential compartment analysis of Hi-C datasets. bioRxiv doi: https://doi.org/10.1101/2021.02.02.429297, 01 April 2022, preprint: not peer reviewed.
46. Abdennur,N. and Mirny,L.A. (2020) Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics*, **36**, 311–316.
47. Van Der Weide,R.H., Van Den Brand,T., Haarhuis,J.H.I., Teunissen,H., Rowland,B.D. and De Wit,E. (2021) Hi-C analyses with GENOVA: a case study with cohesin variants. *NAR Genomics Bioinforma.*, **3**, lqab040.
48. Robinson,J.T., Thorvaldsdóttir,H., Winckler,W., Guttman,M., Lander,E.S., Getz,G. and Mesirov,J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24.
49. Mumbach,M.R., Rubin,A.J., Flynn,R.A., Dai,C., Khavari,P.A., Greenleaf,W.J. and Chang,H.Y. (2016) HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods*, **13**, 919–922.
50. Ji,X., Dadon,D.B., Powell,B.E., Fan,Z.P., Borges- Rivera,D., Shachar,S., Weintraub,A.S., Hnisz,D., Pegoraro,G., Lee,T.I. *et al.* (2016) 3D Chromosome regulatory landscape of human pluripotent cells. *Cell Stem Cell*, **18**, 262–275.
51. Durand,N.C., Shamim,M.S., Machol,I., Rao,S.S.P., Huntley,M.H., Lander,E.S. and Aiden,E.L. (2016) Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.*, **3**, 95–98.
52. Knight,P.A. and Ruiz,D. (2013) A fast algorithm for matrix balancing. *IMA J. Numer. Anal.*, **33**, 1029–1047.
53. Robinson,J.T., Turner,D., Durand,N.C., Thorvaldsdóttir,H., Mesirov,J.P. and Aiden,E.L. (2018) Juicebox.js provides a cloud-based visualization system for Hi-C data. *Cell Syst.*, **6**, 256–258.
54. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
55. Tripathi,S., Pohl,M.O., Zhou,Y., Rodriguez- Frandsen,A., Wang,G., Stein,D.A., Moulton,H.M., DeJesus,P., Che,J., Mulder,L.C.F. *et al.* (2015) Meta- and Orthogonal integration of influenza "OMICs" data defines a role for UBR4 in virus budding. *Cell Host Microbe*, **18**, 723–735.
56. Szklarczyk,D., Morris,J.H., Cook,H., Kuhn,M., Wyder,S., Simonovic,M., Santos,A., Doncheva,N.T., Roth,A., Bork,P. *et al.* (2017) The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic. Acids. Res.*, **45**, D362–D368.
57. Bader,G.D. and Hogue,C.W.V (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinf.*, **4**, 2.
58. Cline,M.S., Smoot,M., Cerami,E., Kuchinsky,A., Landys,N., Workman,C., Christmas,R., Avila- Campilo,I., Creech,M., Gross,B. *et al.* (2007) Integration of biological networks and gene expression data using cytoscape. *Nat. Protoc.*, **2**, 2366–2382.
59. Collins,P.L., Kyle,K.E., Egawa,T., Shinkai,Y., Oltz,E.M. and Moran,J.V (2015) The histone methyltransferase SETDB1 represses endogenous and exogenous retroviruses in B lymphocytes. *Proc. Natl. Acad. Sci.*, **112**, 8367–8372.
60. Matsui,T., Leung,D., Miyashita,H., Maksakova,I.A., Miyachi,H., Kimura,H., Tachibana,M., Lorincz,M.C. and Shinkai,Y. (2010) Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature*, **464**, 927–931.
61. Ong,C.-T. and Corces,V.G. (2014) CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.*, **15**, 234–246.

62. Dowen,J.M., Fan,Z.P., Hnisz,D., Ren,G., Abraham,B.J., Zhang,L.N., Weintraub,A.S., Schuijers,J., Lee,T.I., Zhao,K. *et al.* (2014) Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*, **159**, 374–387.
63. Rahman,S.A., Cuesta,S.M., Furnham,N., Holliday,G.L. and Thornton,J.M. (2014) EC-BLAST: a tool to automatically search and compare enzyme reactions. *Nat. Methods*, **11**, 171–174.
64. Sanij,E., Diesch,J., Lesmana,A., Poortinga,G., Hein,N., Lidgerwood,G., Cameron,D.P., Ellul,J., Goodall,G.J., Wong,L.H. *et al.* (2015) A novel role for the pol i transcription factor ubtf in maintaining genome stability through the regulation of highly transcribed pol II genes. *Genome Res.*, **25**, 201–212.
65. Woolnough,J.L., Atwood,B.L., Liu,Z., Zhao,R. and Giles,K.E. (2016) The regulation of rRNA gene transcription during directed differentiation of human embryonic stem cells. *PLoS One*, **11**, e0157276.
66. Fang,H.T., El Farran,C.A., Xing,Q.R., Zhang,L.F., Li,H., Lim,B. and Loh,Y.H. (2018) Global H3.3 dynamic deposition defines its bimodal role in cell fate transition. *Nat. Commun.*, **9**, 1537.
67. Atlasi,Y., Megchelenbrink,W., Peng,T., Habibi,E., Joshi,O., Wang,S.Y., Wang,C., Logie,C., Poser,I., Marks,H. *et al.* (2019) Epigenetic modulation of a hardwired 3D chromatin landscape in two naive states of pluripotency. *Nat. Cell Biol.*, **21**, 568–578.
68. Di Giammartino,D.C., Kloetgen,A., Polyzos,A., Liu,Y., Kim,D., Murphy,D., Abuhashem,A., Cavaliere,P., Aronson,B., Shah,V. *et al.* (2019) KLF4 is involved in the organization and regulation of pluripotency-associated three-dimensional enhancer networks. *Nat. Cell Biol.***21**, 1179–1190.
69. Jambaldorj,J., Makino,S., Munkhbat,B. and Tamiya,G. (2012) Sustained expression of a neuron-specific isoform of the taf1 gene in development stages and aging in mice. *Biochem. Biophys. Res. Commun.*, **425**, 273–277.
70. Ackermann,J., Ashton,G., Lyons,S., James,D., Hornung,J.P., Jones,N. and Breitwieser,W. (2011) Loss of ATF2 function leads to cranial motoneuron degeneration during embryonic mouse development. *PLoS One*, **6**, e19090.
71. Wang,C., Lee,G., Hsu,W., Yeh,C.H., Ho,M.L. and Wang,G.J. (2006) Identification of USF2 as a key regulator of runx2 expression in mouse pluripotent mesenchymal D1 cells. *Mol. Cell. Biochem.*, **292**, 79–88.
72. Srinivasan,M., Scheinost,J.C., Petela,N.J., Gligoris,T.G., Wissler,M., Ogushi,S., Collier,J.E., Voulgaris,M., Kurze,A., Chan,K.L. *et al.* (2018) The cohesin ring uses its hinge to organize DNA using Non-topological as well as topological mechanisms. *Cell*, **173**, 1508–1519.
73. Fritsch,L., Robin,P., Mathieu,J.R.R., Souidi,M., Hinaux,H., Rougeulle,C., Harel- Bellan,A., Ameyar- Zazoua,M. and Ait-Si-Ali,S. (2010) A subset of the histone H3 lysine 9 methyltransferases suv39h1, G9a, GLP, and SETDB1 participate in a multimeric complex. *Mol. Cell*, **37**, 46–56.
74. Rice,J.C., Briggs,S.D., Ueberheide,B., Barber,C.M., Shabanowitz,J., Hunt,D.F., Shinkai,Y. and Allis,C.D. (2003) Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. *Mol. Cell*, **12**, 1591–1598.
75. Liu,N., Zhang,Z., Wu,H., Jiang,Y., Meng,L., Xiong,J., Zhao,Z., Zhou,X., Li,J., Li,H. *et al.* (2015) Recognition of H3K9 methylation by GLP is required for efficient establishment of H3K9 methylation, rapid target gene repression, and mouse viability. *Genes Dev.*, **29**, 379–393.
76. Lehnertz,B., Ueda,Y., Derijck,A.A.H.A., Braunschweig,U., Perez-Burgos,L., Kubicek,S., Chen,T., Li,E., Jenuwein,T. and Peters,A.H.F.M. (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.*, **13**, 1192–1200.
77. Zhu,Y., Sun,D., Jakovcevski,M. and Jiang,Y. (2020) Epigenetic mechanism of SETDB1 in brain: implications for neuropsychiatric disorders. *Transl. Psychiatry*, **10**, 115.
78. Markouli,M., Strepkos,D. and Piperi,C. (2021)Structure, activity and function of the SETDB1 protein methyltransferase. *Life*, **11**, 817.
79. Blackburn,M.L., Chansky,H.A., Zielinska- Kwiatkowska,A., Matsui,Y. and Yang,L. (2003) Genomic structure and expression of the mouse ESET gene encoding an ERG-associated histone methyltransferase with a SET domain. *Biochim. Biophys. Acta - Gene Struct. Expr.*, **1629**, 8–14.
80. Zheng,S., Xiao,L., Liu,Y., Wang,Y., Cheng,L., Zhang,J., Yan,N. and Chen,D. (2018) DZNep inhibits H3K27me3 deposition and delays retinal degeneration in the rd1 mice. *Cell Death. Dis.*, **9**, 310.
81. Miranda,T.B., Cortez,C.C., Yoo,C.B., Liang,G., Abe,M., Kelly,T.K., Marquez,V.E. and Jones,P.A. (2009) DZNep is a global histone methylation inhibitor that reactivates developmental genes not silenced by DNA methylation. *Mol. Cancer Ther.*, **8**, 1579–1588.
82. Monaghan,L., Massett,M.E., Bunschoten,R.P., Hoose,A., Pirvan,P.A., Liskamp,R.M.J., Jørgensen,H.G. and Huang,X. (2019) The emerging role of H3K9me3 as a potential therapeutic target in acute myeloid leukemia. *Front. Oncol.*, **9**, 705.
83. Karanth,A.V., Maniswami,R.R., Prashanth,S., Govindaraj,H., Padmavathy,R., Jegatheesan,S.K., Mullangi,R. and Rajagopal,S. (2017) Emerging role of SETDB1 as a therapeutic target. *Expert Opin. Ther. Targets*, **21**, 319–331.
84. Federico,A., Steinfass,T., Larribère,L., Novak,D., Morís,F., Núñez,L.E., Umansky,V. and Utikal,J. (2020) Mithramycin a and mithralog EC-8042 inhibit SETDB1 expression and its oncogenic activity in malignant melanoma. *Mol. Ther. - Oncolytics*, **18**, 83–99.
85. Schwarzer,W., Abdennur,N., Goloborodko,A., Pekowska,A., Fudenberg,G., Loe- Mie,Y., Fonseca,N.A., Huber,W., Haering,C., Mirny,L. *et al.* (2017) Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, **551**, 51.
86. Vara,C., Paytuví- Gallart,A., Cuartero,Y., Waters,P.D., Marti-Renom,M.A. and Ruiz- Herrera,A. (2019) Three-dimensional genomic structure and cohesin occupancy correlate with transcriptional activity during spermatogenesis. *Cell Rep.*, **28**, 352–367.
87. Ocampo- Hafalla,M.T. and Uhlmann,F. (2011) Cohesin loading and sliding. *J. Cell Sci.*, **124**, 685–691.
88. Davidson,I.F., Goetz,D., Zaczek,M.P., Molodtsov,M.I., Huis In 't Veld,P.J., Weissmann,F., Litos,G., Cisneros,D.A., Ocampo-Hafalla,M., Ladurner,R. *et al.* (2016) Rapid movement and transcriptional re-localization of human cohesin on DNA. *EMBO J.*, **35**, 2671–2685.
89. Borrie,M.S., Campor,J.S., Joshi,H. and Gartenberg,M.R. (2017) Binding, sliding, and function of cohesin during transcriptional activation. *Proc. Natl. Acad. Sci.*, **114**, E1062–E1071.
90. Ross,K.E. and Cohen- Fix,O. (2004) Molecular biology: cohesins slip sliding away. *Nature*, **430**, 520–521.
91. Hansen,A.S., Pustova,I., Cattoglio,C., Tjian,R. and Darzacq,X. (2017) CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife*, **6**, e25776.
92. Liu,S., Brind' Amour,J., Karimi,M.M., Shirane,K., Bogutz,A., Lefebvre,L., Sasaki,H., Shinkai,Y. and Lorincz,M.C. (2014) Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes Dev.*, **28**, 2041–2055.
93. He,J., Fu,X., Zhang,M., He,F., Li,W., Abdul,M.M., Zhou,J., Sun,L., Chang,C., Li,Y. *et al.* (2019) Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. *Nat. Commun.*, **10**, 34.
94. Schmidt,D., Schwalie,P.C., Ross- Innes,C.S., Hurtado,A., Brown,G.D., Carroll,J.S., Flicek,P. and Odom,D.T. (2010) A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.*, **20**, 578–588.
95. Lohmann,F., Loureiro,J., Su,H., Fang,Q., Lei,H., Lewis,T., Yang,Y., Labow,M., Li,E., Chen,T. *et al.* (2009) KMT1E mediated H3K9 methylation is required for the maintenance of embryonic stem cells by repressing trophectoderm differentiation. *Stem Cells*, **28**, 201–212.
96. Wolf,D. and Goff,S.P. (2007) TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell*, **131**, 46–57.
97. Kang,Y.-K. (2015) SETDB1 in early embryos and embryonic stem cells. *Curr. Issues Mol. Biol.*, **17**, 1–10.
98. Ishimoto,K., Kawamata,N., Uchihara,Y., Okubo,M., Fujimoto,R., Gotoh,E., Kakinouchi,K., Mizohata,E., Hino,N., Okada,Y. *et al.* (2016) Ubiquitination of lysine 867 of the human SETDB1 protein upregulates its histone H3 lysine 9 (H3K9) methyltransferase activity. *PLoS One*, **11**, e0165766.

99. Gorkin,D.U., Leung,D. and Ren,B. (2014) The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell*, **14**, 762–775.
100. Rathert,P., Dhayalan,A., Murakami,M., Zhang,X., Tamas,R., Jurkowska,R., Komatsu,Y., Shinkai,Y., Cheng,X., Jeltsch,A. *et al.* (2008) Protein lysine methyltransferase G9a acts on non-histone targets. *Nat. Chem. Biol.*, **4**, 344–346.
101. Binda,O. (2016) Lysine methyltransferase SETD6 modifies histones and non-histone proteins. bioRxiv doi: https://doi.org/10.1101/092031, 08 December 2016, preprint: not peer reviewed.
102. Wang,G., Long,J., Gao,Y., Zhang,W., Han,F., Xu,C., Sun,L., Yang,S.-C., Lan,J., Hou,Z. *et al.* (2019) SETDB1-mediated methylation of akt promotes its K63-linked ubiquitination and activation leading to tumorigenesis. *Nat. Cell Biol.*, **21**, 214–225.
103. Fei,Q., Shang,K., Zhang,J., Chuai,S., Kong,D., Zhou,T., Fu,S., Liang,Y., Li,C., Chen,Z. *et al.* (2015) Histone methyltransferase SETDB1 regulates liver cancer cell growth through methylation of p53. *Nat. Commun.*, **6**, 8651.