Nucleic Acids Research, 2024, 1–17
https://doi.org/10.1093/nar/gkae724
Gene regulation, Chromatin and Epigenetics

The impact of the embryonic DNA methylation program on CTCF-mediated genome regulation

Ana Monteagudo-Sánchez [1,†], Julien Richard Albert [1,†], Margherita Scarpa [1], Daan Noordermeer [2] and Maxim V.C. Greenberg [1,*]

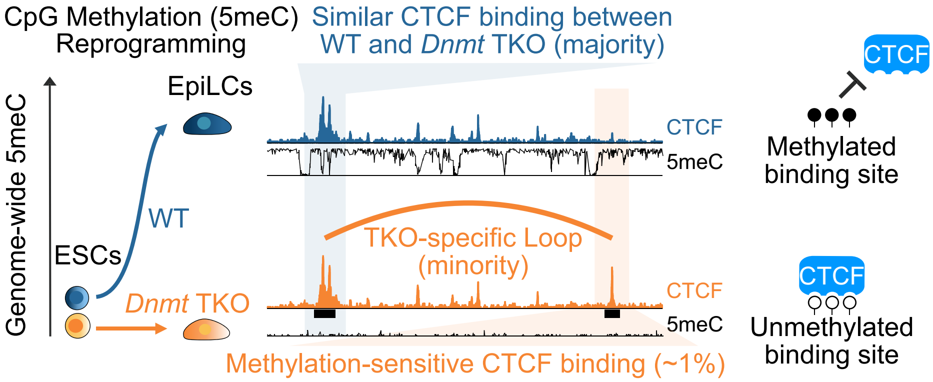
[1] Université Paris Cité, CNRS, Institut Jacques Monod, F-75013 Paris, France
[2] Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), F-91998 Gif-sur-Yvette, France

[*] To whom correspondence should be addressed. Tel: +33 1 57 27 81 20; Fax: +33 1 57 27 80 87; Email: maxim.greenberg@ijm.fr
[†] The first two authors should be regarded as Joint First Authors.
Present address: Ana Monteagudo-Sánchez, Carlos Simon Foundation - INCLIVA Health Research Institute, Valencia, Spain.

Abstract

During mammalian embryogenesis, both the 5-cytosine DNA methylation (5meC) landscape and three dimensional (3D) chromatin architecture are profoundly remodeled during a process known as 'epigenetic reprogramming.' An understudied aspect of epigenetic reprogramming is how the 5meC flux, *per se*, affects the 3D genome. This is pertinent given the 5meC-sensitivity of DNA binding for a key regulator of chromosome folding: CTCF. We profiled the CTCF binding landscape using a mouse embryonic stem cell (ESC) differentiation protocol that models embryonic 5meC dynamics. Mouse ESCs lacking DNA methylation machinery are able to exit naive pluripotency, thus allowing for dissection of subtle effects of CTCF on gene expression. We performed CTCF HiChIP in both wild-type and mutant conditions to assess gained CTCF–CTCF contacts in the absence of 5meC. We performed H3K27ac HiChIP to determine the impact that ectopic CTCF binding has on *cis*-regulatory contacts. Using 5meC epigenome editing, we demonstrated that the methyl-mark is able to impair CTCF binding at select loci. Finally, a detailed dissection of the imprinted *Zdbf2* locus showed how 5meC-antagonism of CTCF allows for proper gene regulation during differentiation. This work provides a comprehensive overview of how 5meC impacts the 3D genome in a relevant model for early embryonic events.

Graphical abstract



Introduction

5-Cytosine DNA methylation (5meC) is a highly conserved epigenetic mark, generally associated with gene repression. In mammals, DNA methylation is typically found in the CpG dinucleotide context and approximately 80% of CpGs are methylated in somatic tissues. During the early stages of mammalian development following fertilization, most of the gametic 5meC is erased; subsequently, during implantation, the *de novo* methyltransferases DNA METHYLTRANSFERASE 3A and 3B (DNMT3A and DNMT3B) rapidly establish the embryonic DNA methylation landscape (1). This period is called the naïve-to-primed pluripotency transition, and occurs just prior to germ layer specification (2). Changes in the histone modification patterns and in the transcriptional landscape are also substantial during this period (3). Thus, it is presumed that the epigenome plays an integral role in preparing the cells within the embryo for lineage commitment.

In mammalian cell nuclei, the chromatin is organized in hierarchical structures that range from multi-megabase chromosome territories to more local *cis*-regulatory contacts (4–7). Inside the territories, different chromatin compartments are defined by their transcriptional activity: the euchromatic 'A' compartments that are typically transcriptionally active and the 'B' compartments that are relatively transcriptionally repressed (4,8). The compartments themselves are organized into Topologically Associating Domains (TADs)—'regulatory

Received: November 20, 2023. **Revised:** July 23, 2024. **Editorial Decision:** July 27, 2024. **Accepted:** August 21, 2024

neighborhoods' that facilitate gene expression programs (9–11). Within TADs, DNA loops can be formed, which are the smallest degree of organization, and can enable or insulate interactions between gene promoters and *cis*-regulatory elements such as enhancers (6,12). The *cis* regulatory contacts differ substantially between cell types and they are crucial for determining proper cell identity (13,14). This hierarchical chromosome organization is dynamic and very important for several genomic processes, including transcription, gene regulation, replication and cell division (15).

There are several architectural proteins involved in chromatin organization, CCCTC-BINDING FACTOR (CTCF) being one of the most well-characterized. CTCF is a zinc finger (ZF) protein that is highly conserved in mammals and that binds pervasively throughout the genome. It is known that CTCF plays a role together with the cohesin complex in the demarcation of TADs boundaries (16–18) and also has a role in transcription as it regulates loops between enhancers and promoters (19,20). The absence of CTCF in mice is lethal in the early embryo, whereas heterozygous deletions of the protein present predisposition to cancer (21,22), indicating that CTCF plays an essential role in development and cell identity. Genome profiling analysis of CTCF occupancy in human cells, obtained from different tissues, reveals cell-type-specific signatures (23).

What are the mechanisms that dictate cell-specific CTCF binding patterns? Certainly, transcription factors play a role (24,25), as well as chromatin modifying complexes (26). However, a compelling mechanism is DNA methylation itself, given that roughly 40% of variability of CTCF binding patterns between cell types can be linked to 5meC status in the binding site (27). Biochemical analyses have confirmed that the presence of 5meC at certain cytosines within the CTCF binding motif can significantly impair CTCF-DNA interactions (28). While DNA methylation does not appear to play a significant role in either TAD or compartment establishment (29–31)—with the notable exception of some cancers (32,33)—5meC indeed has an effect on relatively short *cis*-regulatory contacts in some contexts. *In vivo* the antagonistic relationship between CTCF binding and 5meC has been well-documented at genomic imprints. For example, the paternally methylated *H19-Igf2* imprinting control region (ICR) repels CTCF binding, allowing for interactions between enhancers at this locus and the *Igf2* promoter leading to expression. Conversely, CTCF binds the unmethylated maternal allele, insulating the activation of the *Igf2* promoter from its enhancers, which in turn allows the expression of the *H19* long non-coding RNA (lncRNA) (34–36). The antagonism between CTCF binding and 5meC has also been observed in tumors (37). To wit, in IDH mutant gliomas the hypermethylation of a CTCF binding site causes a reduction in CTCF binding that results in the expression of a glioma oncogene (32,38).

In this study, we set out to determine how the dramatic embryonic DNA methylation program impacts three-dimensional chromatin architecture and underlying gene regulation in a dynamic system. We employed a mouse embryonic stem cell (ESC) differentiation approach that recapitulates the embryonic *de novo* DNA methylation dynamics: naïve ESCs cultured in serum-free media, which are characterized by low levels of DNA methylation (39), were differentiated to Epiblast-like cells (EpiLCs) inducing the transition towards primed pluripotency (40). In parallel, we employed a *Dnmt1; Dnmt3a; Dnmt3b* triple knockout (TKO) cell line, which despite completely lacking DNA methylation (41), is able to adopt a primed-like state during EpiLC differentiation (42–45). We reasoned that our strategy using defined media could provide developmentally relevant insights that were masked in prior studies using solely serum-grown ESCs (46,47), which exhibit relatively high heterogeneity (48). We profiled CTCF binding changes in ESCs and EpiLCs in the presence and absence of 5meC, showing that ∼1% CTCF binding sites are enriched in TKO EpiLCs, relative to WT. Previous chromosome conformation studies using DNA methylation mutants were not able to detect architectural differences at finer scales, therefore likely missed many short-range *cis*-regulatory interactions (30,31). Hence, we utilized HiChIP (49), which captures both short and long-range interactions and allowed us to assess either chromatin loops by enriching for CTCF-bound loci, or enhancer-promoter contacts by enriching for histone H3 lysine 27 acetylation (H3K27ac)-marked regions. We could then determine how differential CTCF binding affects the *cis*-regulatory landscape. We functionally demonstrated that 5meC negatively impacts CTCF binding at multiple loci by implementing epigenome editing. Finally, we carried out fine grained genetic experiments to show how 5meC influences CTCF-mediated gene regulation at the imprinted *Zdbf2* locus. In sum, our study provides a comprehensive view of how the embryonic DNA methylation program contributes to chromatin folding as a means to control gene expression.

Materials and methods

ESC cell lines

E14Tg2a (E14) mouse ESCs was the parental line used for all experiments in this study, as well as serving as the background for all transgenic lines. The TKO was previously generated in-house (50).

Cell culture and differentiation

For the cells grown in serum culture conditions we used Glasgow medium (Gibco) supplemented with 15% Fetal bovine Serum (FBS), 0.1 mM MEM non-essential amino acid, 1 mM sodium pyruvate, 2 mM L-glutamine, penicillin, streptomycin, 0.1 mM β-mercaptoethanol and 1000 U/ml leukemia inhibitory factor (LIF). To pass, the cells were washed with 1× PBS, then trypsin was added to detach and disaggregate the cells for 5 min at 37°C. The desired number of cells were then transferred to the new flask.

For the 2i + vitC culture conditions we used N2B27 medium (50% neurobasal medium, 50% DMEM) supplemented with N2 (Gibco), B27 (Gibco), 2 mM L-glutamine, 0,1 mM β-mercaptoethanol, penicillin, streptomycin, LIF and 2i (3 μM Gsk3 inhibitor CT-99021, 1 μM MEK inhibitor PD0325901) and Vitamin C (Sigma) at a final concentration of 100 μg/ml. To pass the cells, the media was removed, then Accutase (Gibco) was added to detach and disaggregate the cells and incubated for 5 min at room temperature. The desired number of cells were then transferred to the new plate. The ESCs in both conditions were grown on 0.1% gelatin-coated flask in an incubator at 37°C and 5% CO_2.

To induce EpiLC differentiation, cells were gently washed with PBS, dissociated, and replated at a density of 2×10^5 cells/cm^2 on Fibronectin (10 μg/ml, Sigma))-coated plates in

N2B27 medium supplemented with 12 ng/ml FGF2 (R&D) and 20 ng/ml Activin A (R&D). EpiLCs were passed with Accutase at day 3 of differentiation when the differentiation time was 7 days.

To inhibit PRC2 activity, cells were treated with 2 µM UNC1999 (Tocris) for at least one week. Control cells were treated with UNC2400 (Tocris)—an analog with >1000-fold lower potency—at the same concentration and in parallel.

Generation of sgRNA constructs for epigenome editing

A piggyBac transposition compatible vector (51) was modified by removing the sequences between the inverted terminal repeats by restriction digest, and incorporating a Hygromycin B resistance gene and U6-TRACR sequence by Gibson assembly. For *Nrp2*, a one guide RNA sequence was inserted by digesting the vector with BbsI, and ligating a double stranded DNA sequence containing compatible overhangs. For the other targets, dual-guide constructs were generated by linearizing the plasmid with BbsI and inserting a PCR product containing one gRNA sequence, the invariant sgRNA scaffold sequence, a modified murine U6 promoter and a second gRNA sequence, using the pLKO.1-blast-U6-sgRNA-BfuA1-stuffer plasmid as a template for amplification (52). Guide RNA sequences were designed using the CRISPOR online program (crispor.tefor.net). Non-specific gRNA sequences were used as controls. Oligo sequences can be found in Supplementary Table S1.

Generation of transgenic ESCs

All transgenesis experiments were performed with ESCs cultured in serum-containing media. Briefly, in each transfection ~5 million cells were transfected with a mix containing 2.5 µg of each plasmid and plated at different concentrations to allow clone selection. We then performed electroporation using the Amaxa® Nucleofector® II Device from Lonza with the mouse ESC (A-013) program according to the manufacturer's instructions. The transfected cells were cultured for a day in antibiotic-free media and then were placed under antibiotic selection. The SunTag/TET epigenome editing construct was obtained from Addgene (Plasmid #82559). Individual clones that were Geneticin® (ThermoFisher) resistant were screened for Cas9 and GFP expression by western blotting. The dual guide vectors were co-transfected with a plasmid containing the PiggyBac transposase, and Hygromycin B (ThermoFisher) resistant cells were pooled.

Generation of CTCF binding site deletion

The deletion of the CTCF binding site was generated by transfecting two CRISPR sgRNAs flanking the target sequence along with Cas9. sgRNAs were designed using the online CRISPOR online program (crispor.tefor.net) and cloned into the pX459 plasmid harboring the Cas9 gene. Around 5 million WT serum-grown ESCs were transfected with 1 µg of plasmids using Amaxa 4d Nucleofector (Lonza) and plated at a low density. Ninety-six individual clones were picked and screened by PCR for ~600 bp deletion. Mutated alleles were confirmed by Sanger sequencing of cloned PCR amplicons. sgRNA sequences and genotyping primers can be found in Supplementary Table S1.

CUT&RUN

We performed CUT&RUN according to the original protocol (53) with the following modifications: $5x10^5$ cells were used for each sample, and the primary antibody was incubated overnight at 4°C on a rotator. After incubation with pAG-MNase and performing the MNAse reaction, the samples were placed on a magnetic rack and the supernatant containing the DNA samples was recovered. Following addition of 0.1% SDS and 0.17 mg/ml Proteinase K, samples were incubated at 50°C for 1 h. Purified DNA was obtained by phenol/chloroform extraction and precipitated with 100% ethanol by centrifugation. The DNA pellet was washed in 80% ethanol, spun down and air-dried before being resuspended in 25 µl of 1 mM Tris–HCl pH8.0. We used a primary CTCF antibody (Cell D31H2) or IgG control (Sigma Aldrich I5006) and no secondary antibody was used for these experiments. The pAG-MNase plasmid was obtained from Addgene (#123461), and the protein was purified by the Curiecoretech Recombinant Protein Platform.

Sequencing library preparation was made using the NEBNext® UltraTM II DNA Library Prep Kit for Illumina (NEB) following the procedure described in 'Library Prep for CUT&RUN with NEBNext® UltraTM II DNA Library Prep Kit for Illumina® (E7645) V.1' available in protocols.io. Quality control for the finalized libraries was performed using a TapeStation 420 system (Agilent). Libraries were sequenced by Novogene Co on a NovaSeq using paired-end 150 bp parameters, requesting 4 GB of data per sample, or approximately 13 million reads. The full list of datasets generated in this study are listed in Supplementary Table S2.

Hi-ChIP

Hi-ChIP experiments were performed using the Arima Hi-ChIP kit (Arima Genomics) according to the manufacturer's instructions. 15 µg of chromatin were used per sample and experiments were performed in duplicates. Briefly, cells were cross-linked with 2% formaldehyde for 10 min at room temperature, lysed and chromatin was digested with two different restriction enzymes included in the kit. Overhangs were filled-in in the presence of biotinylated nucleotides, followed by ligation. Ligated DNA was sonicated using the Covaris M220 to an average fragment size of 500 bp with the following parameters (peak incident power: 50; duty factor: 10%; cycles per burst: 200; treatment time: 250 s). DNA was then immunoprecipitated overnight using 2.5 µg of H3K27Ac (Active Motif 91193) or CTCF antibody (Active Motif 91285). After a double-size selection to retain DNA fragments between 200 and 600 bp using Ampure XP beads (Beckman Coulter) the biotin-ligated DNA was precipitated with streptavidin-coupled magnetic beads (included in the kit).

The Hi-ChIP libraries were prepared on beads using the Accel-NGS 2D library Kit (Swift Bioscience) following instructions from the Arima Hi-ChIP kit. Final libraries were analyzed using 4200 TapeStation system (Agilent) and sequenced by Novogene Co on a NovaSeq using paired-end 150 bp parameters.

ChIP-qPCR

CTCF ChIP was performed as previously described (54). Briefly, 4 µl of CTCF rabbit antibody (AbFlex 91285) or 4 µl of IgG control rabbit antibody at 1 mg/ml (SigmaAldrich I5006) were combined to 50 µl of protein A magnetic beads

(Invitrogen 10001D) and added to sonicated chromatin (from 200 to 700 bp, checked on agarose gel) from 7–9 million cells, O/N in the cold room. Beads were washed twice with TF-WBI (20 mM Tris–HCl/pH 7.4, 150 mM NaCl, 0.1% SDS, 1% Triton X − 100, 2 mM EDTA), twice with TFWBIII (250 mM LiCl, 1% Triton X-100, 0.7% DOC and 10 mM Tris–HCl, 1 mM EDTA), and twice with with TET (0.2% Tween-20, 10 mM Tris–HCl/pH 8.0, 1 mM EDTA). Chromatin was eluted and de-crosslinked in 70 μl of elution buffer (0.5% SDS, 300 mM NaCl, 5 mM EDTA, 10 mM Tris–HCl pH 8.0) containing 40 μg of proteinase K in an overnight incubation at 65°C. Eluted and purified DNA (Qiagen 28204) was directly used for qPCR. Data was first normalized to input, then to positive control locus (chr1:63181149–63181244). Primer sequences can be found in Supplementary Table S1.

Circular chromosome conformation capture (4C)

4C material was generated as previously described (55). Naïve ESCs containing the constitutively expressed dCas9-SunTag/TET1construct plus either dual guides targeting a 5meC-sensitive CTCF binding site or a scrambled gRNA control were differentiated to EpiLCs for 4 days. 10–15 million EpiLCs were cross-linked, after which MboI (New England Biolabs) was used as the primary restriction enzyme and NlaIII (New England Biolabs) as the secondary restriction enzyme. The four viewpoints representing the four CTCF binding sites were each amplified with specific primers containing Illumina P5 and P7 sequences, respectively. The P7 primer contained a barcode to distinguish between TET1-edited and scrambled control lines. Primer sequences can be found in Supplementary Table S1.

All eight 4C libraries were pooled at equimolar concentrations prior to sequencing. 150 bp single-end reads were generated on a NextSeq 550 (Illumina) at the I2BC Next Generation Sequencing Core Facility.

RNA extraction, cDNA synthesis and RT-qPCR

RNA extraction from cell pellets was performed using the KingFisher Duo Prime Magnetic Particle Processor and the MagMAX mirVana Total RNA kit, according to the manufacturer's instructions.

First strand cDNA synthesis was performed using the SuperScript III Reverse Transcriptase kit (Invitrogen). 500 ng of total RNA were used for each reaction, along with 1 μl of 50 ng/μl of random primers, 1μl of 10mM dNTP mix and sterile H_2O up to 13 μl. The rest of the procedure was performed following the manufacturer's instruction.

For each RT-qPCR reaction, 1 μl of cDNA was mixed with 5 μl of LightCycler 480 SYBR Green I Master and 0.5 μl of 10μM of each forward and reverse primers as well as sterile H_2O up to 10 μl. The RT-qPCR was run on a LightCycler 480 II (Roche Applied Science) using 384-well plates. The samples first followed an initial incubation at 95°C for 10 min, and then 45 cycles of denaturation at 95°C for 10 s, annealing at 61°C for 20 s and extension at 72°C for 20 s. Samples were amplified in triplicates with appropriate non-template controls. Relative gene expression was calculated using the $2^{-\Delta Ct}$ method and normalized to the geometric mean of the expression levels of the two housekeeping genes *Rrm2* and *Rplp0*. Graphical representation and statistical analysis was performed with GraphPad Prism software. Primer sequences can be found in Supplementary Table S1.

Protein extraction and western blot

For protein extraction, we used a BC250 lysis solution (25 mM Tris pH 7.9, 0.2 mM EDTA, 20% glycerol, 0.25 M KCl) supplemented with complete, EDTA-free protease inhibitors (Roche). Then, the samples were sonicated with a Bioruptor sonication device (high, 30 s on, 30 s off, for three cycles) and the protein concentrations were quantified using Pierce BCA Protein Assay Kit (ThermoFisher) on an Infinite M200 (Tecan) machine. Western blot imaging was performed using the ChemiDoc MP (Biorad). The following antibodies and dilutions were used: Lamin-B1 (abcam ab16048) 1:2000 and H3K27me3 (Cell Signaling C36B11) 1:5000.

Pyrosequencing

Genomic DNA was isolated from cells using the NucleoSpin Tissue kit (Macherey-Nagel). 500 ng–1 μg of genomic DNA was bisulfite converted using the EZ DNA Methylation-Gold kit (Zymo). Bisulfite-converted DNA was PCR amplified, and analyzed using the PyroMark Q24 machine and associated software (Qiagen). Graphical representation and statistical analysis was performed with GraphPad Prism software. Primer sequences can be found in Supplementary Table S1.

LUMA

Genomic DNA (500 ng) was digested with MspI + EcoRI or HpaII + EcoRI (New England BioLabs) in parallel duplicate reactions. HpaII is a methylation-sensitive restriction enzyme, and MspI is its methylation-insensitive isoschizomer. EcoRI was included for internal normalization. The extent of the enzymatic digestions was quantified by pyrosequencing (PyroMark Q24), and global CpG methylation levels were then calculated from the HpaII/MspI normalized peak height ratio.

WGBS analysis

Adapter and low-quality sequences were removed using Trimmomatic (v0.39) (56) and parameters 'ILLUMINACLIP:adapters.fa:2:30:10 SLIDINGWINDOW:4:20 MINLEN:24'. Read quality was assessed using FastQC before alignment to the mm10 genome using Bismark (v0.23.1) (57) and default parameters. Reads with mates that did not survive read trimming or that could not be aligned in paired-end mode were concatenated and realigned. PCR duplicate reads were removed using deduplicate_bismark and CpG methylation information was extracted using bismark_methylation_extractor. In vivo WGBS was downloaded from Wang *et al.* (58) and processed using the same parameters. Parallel coordinate plots were generated in VisRseq (v0.9.42) (59).

CUT&RUN analysis

PE 150 reads were trimmed to 36 using Trimmomatic and parameter 'CROP:36'. PCR duplicate reads were removed using Clumpify (v38.18) (60) and parameters 'dedupe = t k = 19 passes = 6 subs=$substitution_filter', where substitution_filter is calculated by multiplying the rate of Illumina sequencing error (1%) with read length. Subsequently, adapter-derived and low-quality nucleotides were removed as described above for WGBS. Read quality was assessed using FastQC before alignment to the mm10 genome using bowtie2 (v2.4.5) (61) and parameters '–local –very-sensitive –no-mixed –dovetail –no-discordant

–phred33 -I 10 -X 700’. Bigwig files were generated using Deeptools (v3.5.1) (62) bamCoverage and parameter ‘–normalizeUsing CPM –blackListFileName blacklisted_regions.fa –ignoreForNormalization chrX chrM chrY –binSize 1’, removing blacklisted regions defined by the Kundaje lab. Aligned reads were used to call peaks using SEACR (v1.3) (63) and parameters ‘0.01 non stringent’. Peak calls were subsequently filtered for the presence of a CTCF motif using FIMO (v5.5.0) (64), the canonical CTCF motif MA0139.1 from the 2022 JASPAR database (65) by applying a *P*-value cutoff of 1e-3.25, as previously described (66). Peaks from all samples were merged using Bedtools (v2.30.0) (67) and default parameters, resulting in 74319 peaks, which were subsequently used to generate correlograms and PCA plots. Enrichment of CTCF binding was calculated over peaks using VisRseq and RPKM values were used to calculate Spearman correlations using Morpheus (https://software.broadinstitute.org/morpheus). To compare WT and TKO EpiLC CTCF profiles, ESC-specific peaks were removed by filtering on enrichment (RPKM > 1 in any EpiLC dataset) as well as no excess enrichment (RPKM < 5) in control IgG samples, resulting in 66 905 peaks. Genomic distribution of peaks was assessed using ChIPseeker (68). Venn diagrams were generated using pybedtools and matplotlib. PCA plots were generated using the from sklearn PCA package (69) and matplotlib. Average 5meC levels over CTCF peaks were calculated using Bedops (v2.4.40) (70). Differential CTCF enrichment over peaks was calculated using Limma (v3.54.1) (71) and default parameters. Scatterplots and violin plots were generated using VisRseq and matplotlib.

HiChIP analysis

HiChIP libraries were sequenced at shallow depth (3–4 million paired-end reads) and the ARIMA MAPS pipeline (v2.0) (72) was used to calculate target sequencing depth. Following deep sequencing, the HiC-Pro pipeline (v3.1.0) (73) was used to digest the mm10 genome (^GATC G^ANTC), align reads and generate contact map matrices. Single-end alignment bam files were used for peak calling using macs2 (v2.2.7.1) (74) and generating bigwigs using bamCoverage (as described above). Adjacent H3K27ac peaks (within 5 kb) were merged using bedtools merge. Chromosome 11 consistently showed more reads in TKO cells and was ignored in subsequent analyses. Correlation between samples was assessed using HiCexplorer hicCorrelate over 5kb matrices. Replicates were merged and PCA plots were generated using fanc (v0.9.25) (75) over chromosome 19 using parameters ‘-Z -s 100000’. A/B compartments were calculated using HiCexplorer hicPCA over 25kb matrices. CTCF enrichment over previously defined autosomal TADs (Bonev 2017) was caulculated using VisRseq. Differential 3D contacts were called using two different pipelines, both of which input individual replicates and consider variation between samples. Firstly, hichipper (v0.7.7) (76) and parameters ‘-mu -pp 1000’ was used to normalise HiChIP signals and call significant loops (FDR < 0.05) and differential loops (between WT and TKO cells) were calculated using diffloop (77) and quickAssoc normalization. bigInteract tracks of differential looping were colour coded based on statistical significance (FDR < 0.05). This identified peak-to-peak 3D interactions. Secondly, to expand the analysis to peak-to-other 3D interactions, we employed fithichip (78) on individual and merged replicates to call significant interactions (5kb bin size, max size 2 Mb, min size 10 kb, FDR 0.05) and the fithichip DiffAnalysisHiChIP.r script was used for differential loop analysis (FDR 0.05, FC > 1.5). Virtual 4C plots were generated in HiCexplorer (v3.7.2) (79–81) using a combination of chichViewpoint (–averageContactBin 4 –range 500000 500000), chicSignificantInteractions (–pValue 0.2 –xFoldBackground 2), chicAggregateStatistic (default parameters), chicDifferentialTest (–alpha 0.25 –statisticTest chi2) and chicPlotViewpoint (–pValueSignificanceLevels 0.1). Since the viewpoint of virtual 4C plots is collapsed, careful manual alignment with genome browser tracks and HiC matrices was performed. hicCompareMatrices and hicPlotMatrix were used to generate heatmaps. UCSC genome browser track hubs (82) were also generated for visualization.

RNAseq analysis

PCR duplicate reads, as well as adapter and low quality sequences, were removed as described above. Trimmed reads were aligned to the mm10 genome using STAR (v2.7.9a) (83) and default parameters. Gene expression levels were quantified over Refseq genes using VisR and uniquely aligned reads (MAPQ $= 255$). Differential expression analysis was conducted using Limma and default parameters. Bigwigs were generated as described above with the additional parameters ‘–minMappingQuality 255’. Bar charts were generated using matplotlib (84).

4C analysis

4C libraries were processed using the c4ctus pipeline (55) using default parameters. Viewpoint coordinates used: *Zdbf2* chr1:63252740–63253821, *Nrp2* chr1:62735878–62736791, *Csf1* chr3:107728368–107729020, *Mob3b* chr4:35108947–35109851. Regions to exclude: *Zdbf2* chr1:63250689–63255821, *Nrp2* chr1:62733878–62738863, *Csf1* chr3:107726363–107731020, *Mob3b* chr4:35106947–35112312. The region over which to normalise 4C signals (viewpoint ± 2 TADs) used: *Zdbf2* 60980000–65189138, *Nrp2* 60319873–65189138, *Csf1* 107090000–109510000, *Mob3b* 32580000–40090000. The delta between 4C signals in cells expressing target sgRNAs versus scrambled sgRNA was calculated using bigWigCompare (v3.5.1).

Results

DNA methylation impacts CTCF binding at a minority of sites

In naïve mouse ESCs, the *de novo* and maintenance DNA methylation is impaired, while active DNA demethylation is stimulated, leading to extremely low 5meC levels: $<10\%$ of all CpGs are methylated, mainly localized to transposable elements (85–89) (Figure 1A, B). To achieve this state, we cultured ESCs in serum-free media, supplemented with MEK and GSK3β inhibitors plus vitamin C (2i + vitC) (39). Given the global DNA hypomethylation, perhaps unsurprisingly the transcriptional landscape of WT naïve ESCs is highly similar to that of TKO ESCs cultured in the same conditions (44). We went on to profile the CTCF binding in both WT and TKO ESCs by Cleavage Under Targets and Release Using Nuclease (CUT&RUN) (53). Consistent with the DNA methylation and transcriptomic data, the CTCF

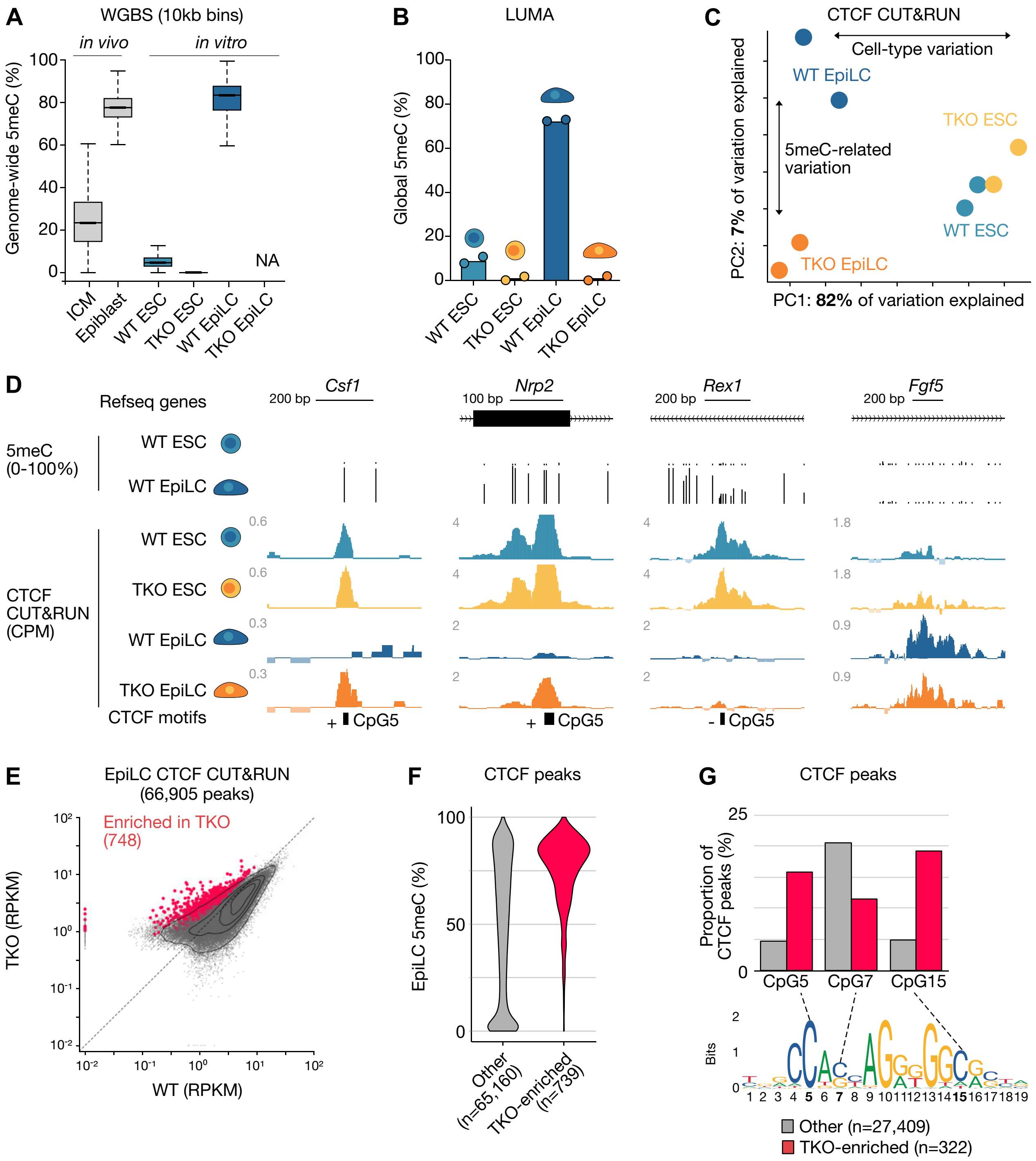


Figure 1. CTCF exhibits potential 5meC-sensitivity at a minority of loci. (**A**) Distribution of average 5meC levels over 10 kb bins ($n = 273\,121$) in E3.5 inner cell mass cells (ICM), the E7.5 epiblast, *Dnmt* WT (WT) and *Dnmt* triple KO (TKO) embryonic stem cells (ESCs) and Epiblast-like cells (EpiLCs). Data from Wang et al. 2014 (*in vivo*), Richard Albert *et al.* (WT ESC & EpiLC) and Domcke *et al.* (TKO ESC). Boxplots represent the median (line inside the box), where 50% of the data are distributed (the box), and whiskers denote the values lying within 1.5 times the interquartile range. (**B**) Luminometric methylation assay (LUMA) depicting global 5meC levels in WT ESCs grown in serum and 2i + vitC, TKO ESCs grown in 2i + vitC and WT and TKO EpiLCs (Day 4). Data are represented as the mean of two replicates, which are included as individual data points. (**C**) Principal component analysis (PCA) plot of CTCF CUT&RUN data in WT and TKO ESCs and EpiLCs. Individual replicates and the percent variation explained by each principal component are shown. (**D**) UCSC genome browser screenshots of putative 5meC-sensitive CTCF binding sites at the *Csf1* and *Nrp2* loci. DNA methylation levels in WT ESCs and EpiLCs and the location of Refseq genes and CTCF binding motifs are included. The strand (±) and position (5, 7 or 15) of the methylated CpG is indicated. Cell-type specific CTCF binding at the naïve pluripotency marker gene *Rex1* (*Zfp42)* and the formative pluripotency marker gene *Fgf5* are shown for comparison. *Csf1*: chr3:107728899–107729432, *Nrp2*: chr1:62738262–62738551, *Rex1*: chr8:43305548–43306216, *Fgf5*: chr5:98255482–98256478. (**E**) 2D scatterplot showing CTCF peak enrichment levels (RPKM) in WT and TKO EpiLCs. Statistically enriched peaks in TKO (linear modeling with Limma, fold-change > 2, t-test adjusted *P* value < 0.05) are highlighted in red. (**F**) Violin plot of the distribution of EpiLC DNA methylation levels within CTCF peaks. Peaks are categorized as in E. (**G**) Bar plot showing the proportion of CTCF peaks that overlap a canonical CTCF binding motif with a CpG at position 5, 7 or 15. The CTCF sequence motif is included (JASPAR MA0139.1).

binding patterns are coherent between WT and mutant conditions (Supplementary Figure S1A, B).

Next we differentiated WT and TKO ESCs to EpiLCs. Within four days, the WT genome is highly methylated (∼80% methylated CpGs), whereas the TKO EpiLC genome remains completely unmethylated (43) (Figure 1A, B). Nevertheless, even in the absence of DNA methylation, we and others have previously demonstrated TKO can not only exit naïve pluripotency, but can do so with similar differentiation kinetics as WT (29,42,44). We reasoned then that EpiLC differentiation would provide a dynamic system in which we could directly compare cell type-specific versus DNA methylation-mediated CTCF regulation. Therefore, as with ESCs, we performed CTCF CUT&RUN in WT and mutant day 4 (D4) of differentiation EpiLCs. The CTCF binding landscape in WT and TKO EpiLCs globally resembled each other more than the TKO EpiLCs resembled naïve ESCs (Figure 1C; Supplementary Figure S1A). This suggests that the cell type plays a more important role in determining CTCF occupancy than DNA methylation, *per se*. Although it is worth noting that there are more differences between WT and TKO EpiLCs than there are between WT and TKO ESCs (Figure 1C; Supplementary Figure S1A, B). These results are in line with a study performed in parallel with ours that used a similar experimental strategy (90).

Nevertheless, we were able to determine a substantial number of CTCF peaks that were enriched specifically in TKO EpiLCs: 748 out of 66 905 total peaks (fold-change ≥ 2, adjusted $P < 0.05$) (Figure 1D, E). A number of data points indicate that DNA methylation is antagonizing CTCF binding at these elements. Firstly, TKO-specific peaks were depleted at promoter regions, which are generally DNA methylation-free (Supplementary Figure S1C). As expected, CpG island promoters—which are defined by their high CpG content, and are typically unmethylated—exhibited an exceedingly small number of TKO-specific CTCF peaks (Supplementary Figure S1D). Secondly, using whole genome bisulfite (WGBS) data (45), we could observe that the vast majority of TKO-enriched sites gain DNA methylation in WT (Figure 1F; Supplementary Figure S1E). This stood in contrast with all other peaks, which exhibited much less bias for DNA methylation gain, and harbored a substantial number of sites that remained unmethylated. Thirdly, we examined in detail the DNA methylation state of CpGs within the CTCF binding motif at the TKO-specific sites. The CTCF core binding motif may contain a number of CpG sites that have been demonstrated to impact CTCF-DNA interactions when methylated, with position five (CpG5) in the JASPAR motif (Figure 1G) exhibiting the most substantial effect in biochemical studies (25,28,91). Of the 748 TKO-enriched sites, we were able to determine 322 wherein we could discern a CTCF motif with DNA methylation information at least at one pertinent CpG (Supplementary Figure S1C). This may indicate that DNA methylation in the local chromatin environment, perhaps via interaction with methyl-sensitive DNA binding proteins, may play an important role in shaping the CTCF binding landscape (92). For ensuing analyses, we focused on CTCF binding sites containing CpGs. Consistent with the overall DNA methylation pattern (Figure 1F), two CpG dinucleotides within the CTCF recognition site (CpG5 and CpG15) showed an enrichment of DNA methylation compared to all other peaks (Figure 1G). While the overall percentage of TKO-specific peaks is only ∼1% of the total number of binding sites, there still remains a fairly considerable number of sites where DNA methylation can influence not only CTCF binding, but potentially *cis* regulatory gene control.

3D genome architecture is globally preserved in DNA methylation-deficient pluripotent cells

HiChIP is a variation of Hi-C in which cross-linked chromatin is immunoprecipitated for a chromatin-associated factor or modification prior to sequencing (49). We performed HiChIP of CTCF in WT and TKO ESCs and EpiLCs in order to enrich for CTCF-anchored contacts (see Supplementary Table S2 for the full list of datasets generated). The HiChIP data is versatile in that both CTCF occupancy, as well as genomic contact information data is generated. Consistent with the CTCF CUT&RUN profiles, the HiChIP datasets are grouped more closely by cell type as opposed to genotype (Figure 2A, Supplementary Figure S2A). Moreover, the data indicated that CTCF enrichment over TADs and A/B compartment organization was largely unperturbed in the absence of DNA methylation (Supplementary Figure S2B–D). From merging the data by cell type, we were able to determine 117294 total CTCF peaks in ESCs compared with 121430 in EpiLCs, thus a very modest 3.5% increase. We next analyzed the number of significant contacts from the HiChIP data, focusing on those contacts that link CTCF-bound sites together. Using stringent parameters, we performed differential analyses and uncovered 876 ESC-specific and 1523 EpiLC-specific loops that met our significance thresholds (FDR ≤ 0.05, $\log_2$FC ≥ 2) (Figure 2B). It is worth noting that previous studies have also reported an increase in CTCF-CTCF contacts during ESC differentiation, which may signify cell type-specific gene regulatory programs becoming cemented (7).

Given the absence of an effect on large chromatin structures, we reasoned that differential CTCF binding may rather impact relatively shorter *cis* regulatory contacts (93). Thus, we performed H3K27ac HiChIP in WT and TKO ESCs and EpiLCs in order to establish the ‘enhancer connectome’ in each of these conditions (94). Keeping in line with the CTCF and transcriptome data, the H3K27ac landscape clusters by cell type much more strongly than by genotype (Figure 2C, Supplementary Figure S2E). We were able to identify 3802 ESC-specific H3K27ac contacts, and 1864 in EpiLCs (Figure 2D). This can be readily observed at marker genes for ESCs and EpiLCs, respectively, which exhibited dramatic changes in their H3K27ac-enriched enhancer-promoter contacts independently of the DNA methylation state (Figure 2E, Supplementary Figure S3). The global preservation of chromatin architecture in TKO EpiLCs strongly bolsters our previous findings that DNA methylation is dispensable for exiting naïve pluripotency (44). However, we were curious to pursue whether we could discover a class of genes that are sensitive to CTCF binding in the 5meC mutant, even if the impact on the overall EpiLC state may be more nuanced.

In line with the fact that ESCs exhibit low/absent levels of 5meC, we only observed two differential loops between WT and TKO. More saliently, CTCF-anchored loops enriched in TKO EpiLCs relative to WT—where 5meC levels are very high—may indicate DNA methylation sensitivity. Using the same analysis, we uncovered 43 differential loops in the DNA methylation mutant, using highly stringent thresholding parameters (Figure 3A). Notably, only two loops were enriched in the WT EpiLCs. Analyzing a larger region around each

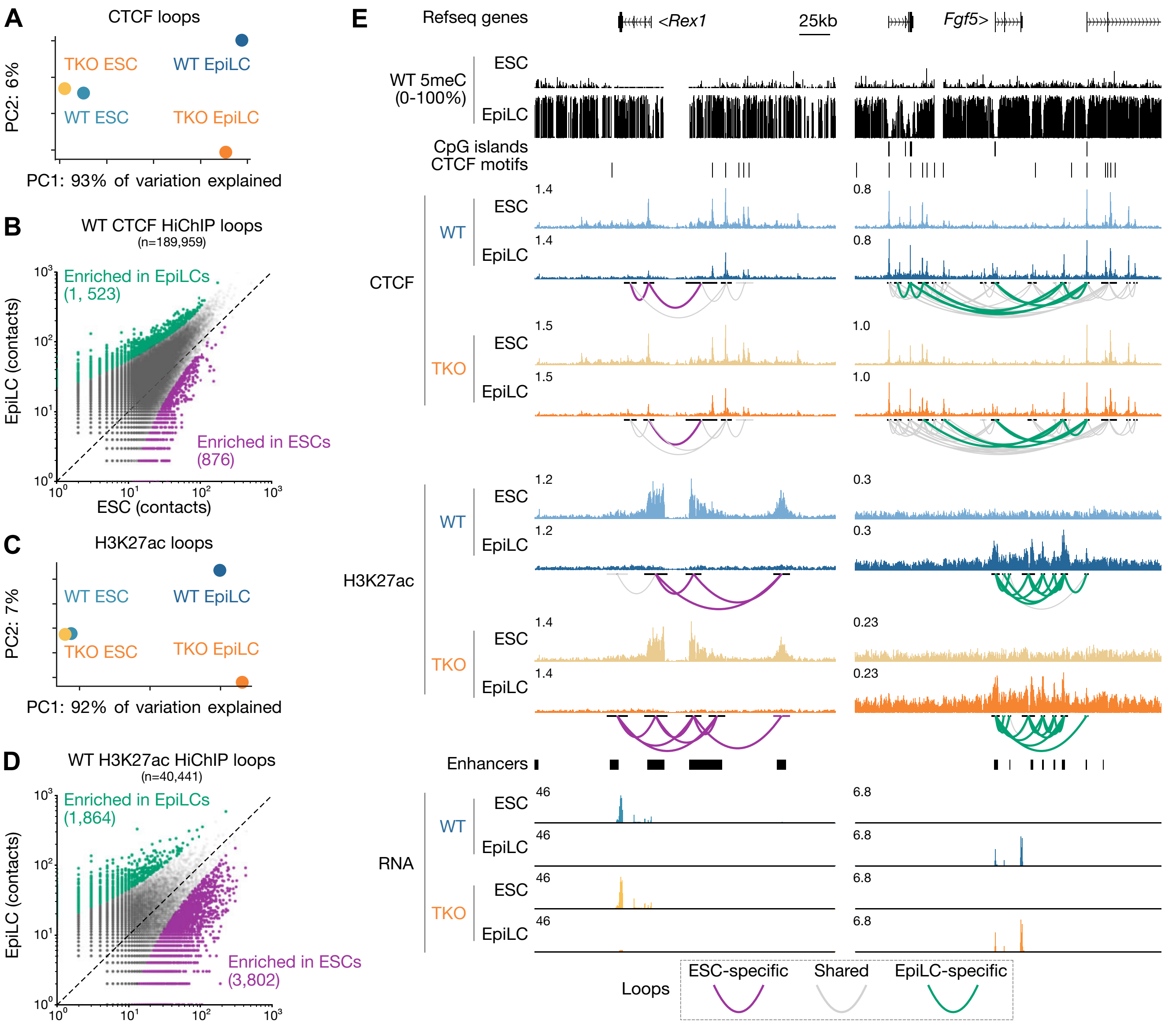


Figure 2. Chromatin architecture is remodeled during EpiLC differentiation in WT and TKO backgrounds. (**A**) PCA plot showing the variation in 3D CTCF contacts between WT and TKO ESCs and EpiLCs. (**B**) 2D scatterplot of CTCF contacts between WT ESCs and EpiLCs. All significant loops are shown in gray. Loops with enriched contacts (>4 reads, FDR < 0.05) in EpiLCs are highlighted in green and ESC-enriched loops are highlighted in purple. (**C**) PCA plot showing the variation in 3D H3K27ac interaction contacts between WT and TKO ESCs and EpiLCs. (**D**) 2D scatterplot of H3K27ac contacts between WT ESCs and EpiLCs as in B. (**E**) UCSC genome browser screenshots of the naïve pluripotency marker gene *Rex1* (*Zfp42*) and the formative pluripotency marker gene *Fgf5* loci. Note the ESC-specific CTCF and H3K27ac binding, CTCF and H3K27ac looping, and gene expression of *Rex1* in WT and TKO backgrounds. EpiLC-specific binding, loop contacts and expression of *Fgf5* is included for comparison. *Rex1*: chr8:43264884–43373313, *Fgf5*: chr5:98139881–98390116. Gray loops represent significant interactions that are consistent across the data being compared, purple are ESC-specific, and green are EpiLC-specific.

anchor (±1750 bp) identifies an additional 96 TKO-specific CTCF loops, and 17 in WT. Thus, by combining both analyses, we defined a final list of 124 differential CTCF contacts, with a noticeable enrichment of contacts in TKO EpiLCs. It is worth noting that most TKO-specific CTCF binding sites could not be associated with TKO-specific loop events, but this appears primarily due to insufficient read coverage at these sites, therefore they remained below our statistical thresholds. Consistently, CTCF peaks from the TKO EpiLC HiChIP data were predominantly DNA methylated in WT (Supplementary Figure S4A, B). It is also worth emphasizing that CpG at position 5 was more enriched than other CpGs in the CTCF binding motif in the TKO EpiLC HiChIP data, suggesting that this is the most deterministic base for 5meC-mediated antagonism (Supplementary Figure S4C).

TKO EpiLC-specific CTCF loops are correlated with gene misregulation at discrete loci

CTCF-mediated chromosome folding can ensure enhancer-promoter contacts allowing for proper gene expression, and at the same time insulate promoters from aberrant enhancer interactions (95) (Figure 3B). Therefore, we set out to determine if the *de novo* DNA methylation program can exert an effect on CTCF-dependent gene control. We first defined all interactions: CTCF-CTCF, CTCF-Other, H3K27ac-H3K27ac

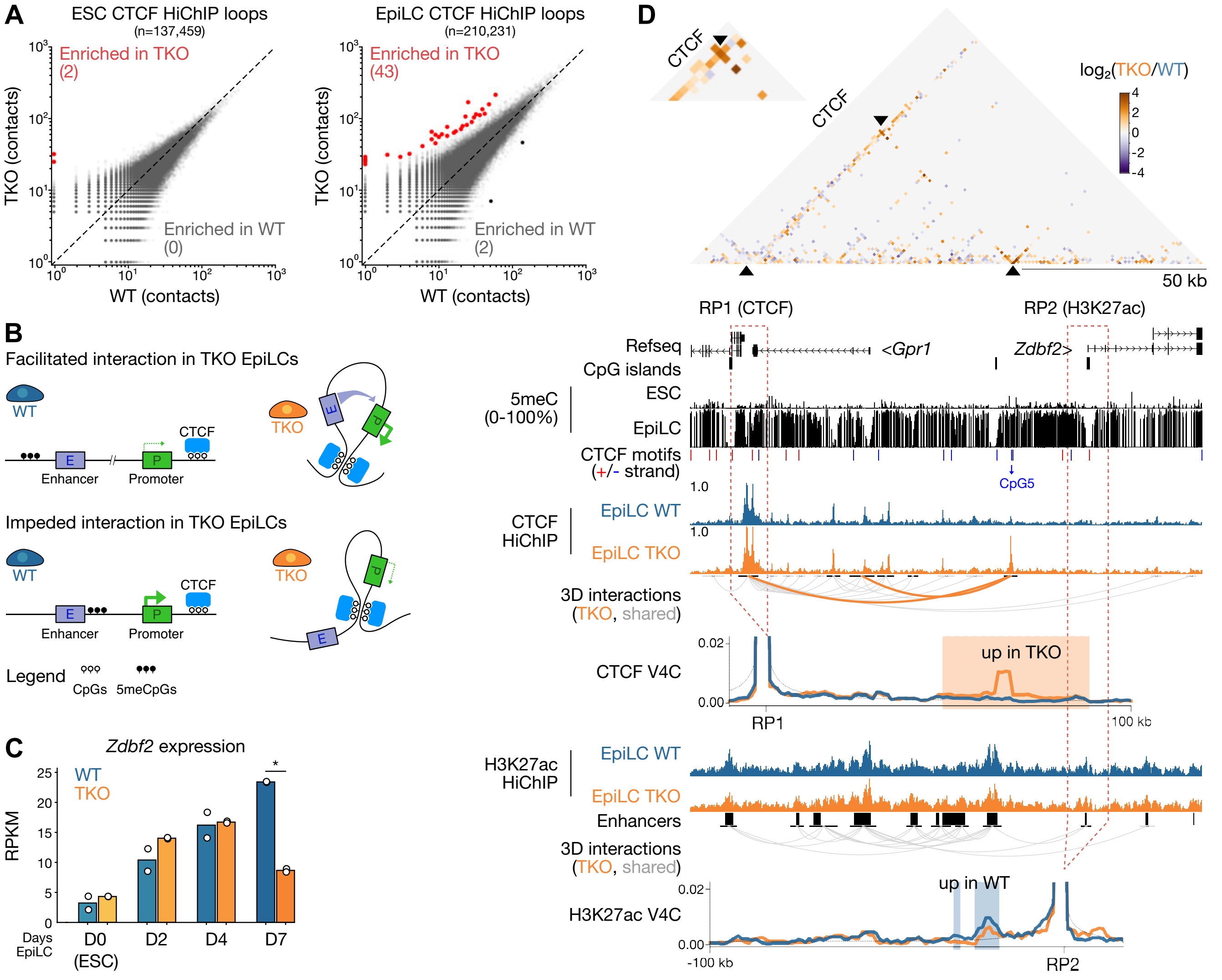


Figure 3. A subset of CTCF loops and promoter-enhancer contacts are disrupted in 5meC-deficient EpiLCs. (**A**) 2D scatterplots of CTCF loop contacts in WT versus TKO ESCs (left) and EpiLCs (right). Loops with significantly enriched contacts in TKO cells (FDR < 0.05) are highlighted in red. (**B**) Schema showing two potential scenarios resulting from global DNA hypomethylation in TKO EpiLCs (right) due to CTCF binding and altered promoter-enhancer contact formation and nearby gene transcription. Top: a hypomethylated CTCF binding site creates a loop that increases enhancer-promoter interactions. Bottom: a hypomethylated CTCF binding site results in a loop that insulates enhancer-promoter interactions. (**C**) Bar plot of *Zdbf2* expression levels (RPKM) in a time course assay of ESC-to-EpiLC differentiation in WT and TKO cells. Bars represent the mean, and replicates are shown as dots. D: days EpiLC differentiation. Statistically significant changes in gene expression between WT and TKO (Linear modeling with Limma, fold-change > 1, t-test adjusted P value < 0.05) are indicated by asterisks. (**D**) Contact matrix (top), genome browser screenshot (middle) and virtual 4C (bottom) plots of the *Zdbf2* locus. Top: differential CTCF contacts between WT and TKO EpiLCs are displayed, where each pixel represents a 1 kb bin. The TKO-enriched contacts between the TAD border and the putative 5meC-sensitive CTCF binding site is magnified in the inset. Reference points 1 (CTCF) and 2 (H3K27ac) for the virtual 4C plots (bottom) are indicated by a dashed box. Middle: browser screenshot showing CpG methylation, CTCF and H3K27ac enrichment levels and loops. Gray loops represent significant interactions that are consistent across the data being compared, orange loops are TKO-specific. Refseq genes, CpG islands, CTCF motifs (positive strand in red, negative strand in blue) are included. The CTCF motif with a CpG at position 5 that underlies a putative 5meC-sensitive CTCF peak is highlighted. Bottom: Virtual 4C plots of reference points 1 (CTCF) and 2 (H3K27ac) showing interaction frequencies between the reference point and adjacent area. The background model is shown as a dotted gray line. Statistically enriched contacts (chi-squared test, alpha < 0.25) are highlighted in blue (WT EpiLC-specific) or orange (TKO EpiLC-specific). Coordinates: chr1:63165972–63304123.

and H3K27ac-Other (Supplementary Figure S5A, B); we found that the majority of interactions were CTCF-CTCF and H3K27ac-H3K27ac, respectively. We next intersected the CTCF and H3K27ac data sets, and found that H3K27ac was enriched at either anchor of roughly half CTCF-CTCF interactions (Supplementary Figure S5C). However, we did not observe a notable difference in H3K27ac enrichment at either anchor of TKO-specific CTCF loops (Supplementary Figure S5D, E). Moreover, we did not interrogate loci where we observed differential H3K27ac-looping with no clear 5meC-CTCF dynamics (Supplementary Figure S6A). Our strategy was to systematically assess the 124 differential CTCF-CTCF loops that were enriched in TKO EpiLCs, and determine if H3K27ac contacts and gene expression were impacted. While we identified many genes overlapping differential CTCF loops that did not change in expression (Supplementary Figure S6B), we identified 76 genes that indicate that DNA methylation could influence gene expression via CTCF antagonism based on their proximity to ectopic CTCF contacts. Finally, at a subset of loci on this list, we observed 5meC-sensitive CTCF-CTCF loops, but we did not see any obvious changes in the underlying H3K27ac (Supplementary Figure S6C).

From the remaining loci, we culled our list to four compelling candidates for further analysis: *Csf1*, *Mob3b, Nrp2* and *Zdbf2*. The *Csf1, Mob3b* and *Nrp2* genes either contain or are adjacent to CTCF binding sites containing a CpG at position 5, and CTCF was enriched at these sites in TKO EpiLCs but not WT (Supplementary Figures S7–S9). Importantly, in all cases the TKO EpiLC-specific binding was associated with the formation of differential loop(s) ($\text{FDR} < 0.05$, fold-change ≥ 2). Finally, these three genes were upregulated in TKO EpiLCs (Supplementary Figure S10A). At least in the case of *Csf1* and *Mob3b*, virtual 4C suggested that the TKO EpiLC-specific loop was also associated with increased interactions between H3K27ac-enriched regions (Supplementary Figures S7 and S8). In other words, the differential CTCF–CTCF looping could be facilitating enhancer-promoter contacts, leading to upregulation (Figure 3B, Supplementary Figures S7B and S8B).

Finally, from our analyses, the most significant differential CTCF loop that was enriched in TKO EpiLCs was found at the imprinted *Zdbf2* locus. As opposed to the previous examples, the presence of the differential loop was correlated with decreased *Zdbf2* expression (Figure 3C). While we uncovered this loop through an agnostic approach, incidentally this is a locus that we have previously characterized. In our culture system, the *Zdbf2* maternal imprint is lost, and both alleles behave like the active paternal allele, thus mitigating the potential for confounding analysis (42). During ESC to EpiLC differentiation, DNA methylation upstream of the *Zdbf2* promoter is required to antagonize polycomb repressive complex 2 (PRC2)-mediated silencing in order to allow proper gene activation (42,96). We also showed that four enhancers upstream of the *Zdbf2* promoter are crucial for its activity (43). The physiological consequences of embryonic *Zdbf2* regulation are life-long: in mouse embryos where DNA methylation is not deposited upstream of the *Zdbf2* promoter, the gene remains constitutively polycomb-repressed, leading to decreased appetite, smaller size, and lower survivability in affected pups with respect to their WT littermates (42,97). As such, *Zdbf2* has emerged as a valuable locus to study the long-lasting effects of epigenetic reprogramming.

The differential loop at the *Zdbf2* locus is anchored in a CTCF binding site that sits between the *Zdbf2* promoter and the aforementioned four enhancers. In WT EpiLCs, CTCF binding was depleted, which was correlated with a gain of 5meC at position 5 in its binding site (Figure 3D). In TKO EpiLCs, where CTCF binding was maintained, our H3K27ac HiChIP data revealed less interactions between the *Zdbf2* promoter with upstream enhancers in the TKO EpiLCs compared with WT (Figure 3D). Thus, we reasoned that in ESCs, the CTCF binding could help insulate *Zdbf2* from precocious activation; this insulation is maintained in the DNA methylation mutant, helping to explain the persistent repression when the CTCF site is not methylated (Figure 3B). Consistent with this interpretation, the *Zdbf2* promoter exhibits substantially increased interactions with the upstream enhancers at later time points in differentiation when the gene is more highly expressed compared with the hypomethylated naïve ESC state (43).

Epigenome editing confirms DNA methylation-CTCF antagonism

While globally the WT and TKO EpiLCs are transcriptionally similar, there are a substantial number of misregulated genes in the DNA methylation mutant (44). Thus, it is possible that the gene misregulation we have described may be indirect of CTCF-mediated action. To formally demonstrate that DNA methylation, *per se*, affects CTCF binding and downstream regulatory defects, we performed locus-specific DNA demethylation using the CRISPR/Cas9 SunTag system. Briefly, catalytically inactive Cas9 (dCas9) fused to five GCN4 epitopes (SunTag) recruits the TET1 catalytic domain fused to GFP and a single chain variable fragment (scFv) that recognizes the SunTag (98). We took advantage of a piggyBac transgenesis-compatible plasmid where all components are expressed as a single transcript driven by a constitutive promoter, and the translated peptide contains the P2A self-cleavable peptide sequence between the dCas9-SunTag and GFP-scFv-TET1 (Supplementary Figure S10B) (45,99). After selecting for GFP positive cells, we used piggyBac-mediated transgenesis to stably integrate single guide RNAs (sgRNAs) that target the epigenome editing machinery to the respective CTCF binding sites. Following selection of sgRNA integration, we differentiated the dCas9-SunTag/TET and control lines to EpiLCs for four days. In all cases, we would expect that persistent DNA demethylation would lead to increased CTCF binding. However, depending on the mode of regulation, we would expect either increased expression (eg, *Csf1*, *Mob3b* and *Nrp2*) or repression (eg, *Zdbf2*) (Figure 4A).

Indeed, we observed robust targeted DNA demethylation for each candidate locus compared to control (Figure 4B, Supplementary Figure S10C). Validating our prediction, reduced 5meC was associated with increased CTCF binding in every case (Figure 4C, Supplementary Figure S11A, B). We next assessed if local gene expression was altered when CTCF binding was increased. In the case of *Csf1*, *Mob3b* and *Nrp2*, we did not observe changes in expression that were consistent with our model (Supplementary Figure S11C).

To assess the impact of TET1-editing on 3D genome organization as well as to gain insight into the the transcriptional regulation of the the three aforementioned genes, we performed high-resolution circular chromosome conformation capture followed by sequencing (4C-seq) (100,101), using the 5meC-sensitive CTCF binding sites as viewpoints. At *Csf1*, we did not observe enrichment of the TKO-specific contact present in the CTCF HiChIP data (Supplementary Figure S12A). This may be due to relatively minor enrichment of CTCF in the TET1-edited line (Supplementary Figure S11A), or an artifact caused by dCas9 binding to chromatin; in any event, this could help explain the absence of an expression change. Conversely, at both *Mob3b* and *Nrp2* we indeed observed increased contacts at the TKO-specific CTCF anchor sites when 5meC was reduced—hence, they are *bona fide* 5meC-sensitive loops (Supplementary Figure S12B, C). Interestingly, we observed reduced expression at *Mob3b*, contrary to our expectations (Supplementary Figure S11C). Further experiments will need to be performed to determine if this loop structure in fact dampens *Mob3b* transcriptional activity or if it simply is transcriptional interference by the dCas9-SunTag complex bound in the body of the *Mob3b* gene (102). Finally, while the TKO-specific CTCF binding and looping dynamics are clearly

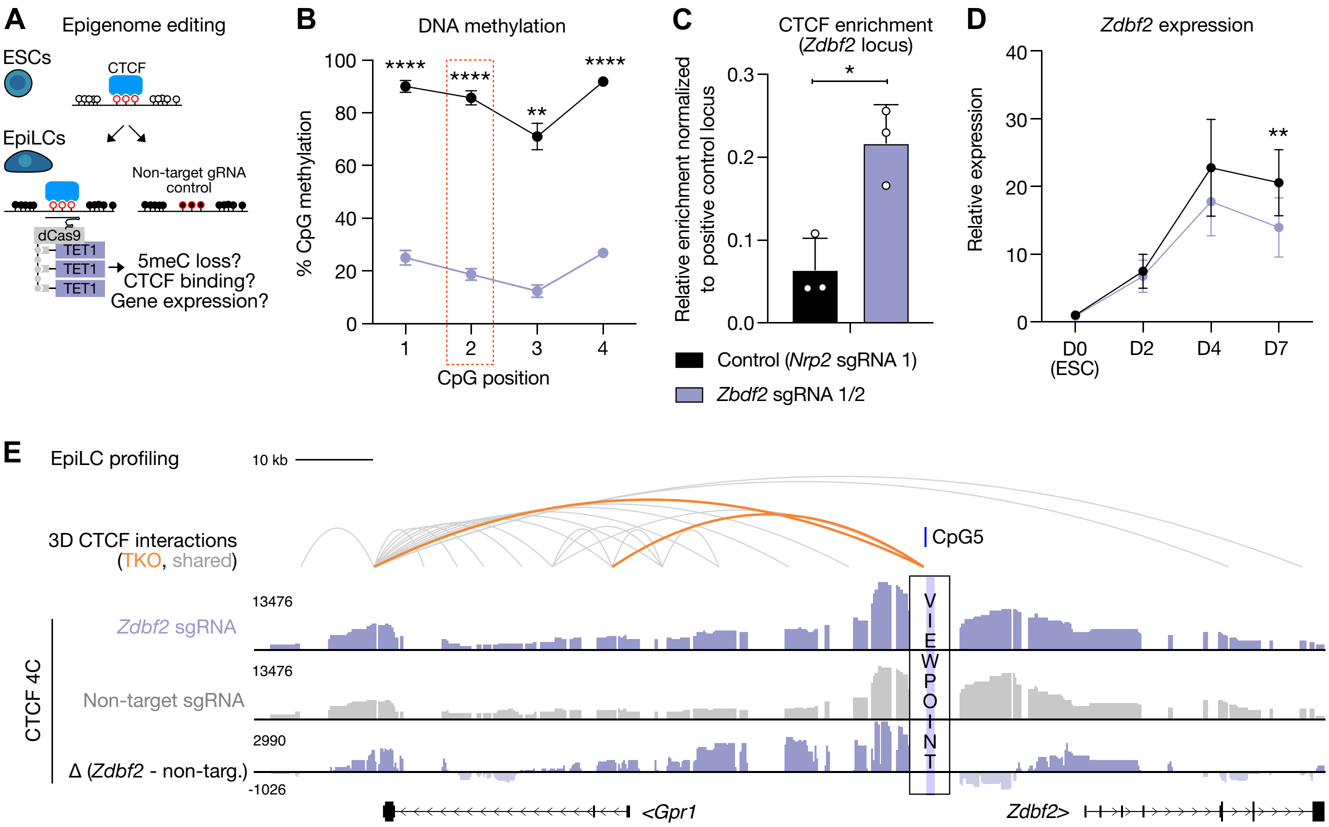


Figure 4. Precision cytosine demethylation at the *Zdbf2* locus results in increased CTCF binding and failure to completely activate gene expression. (**A**) Schema depicting the site-directed 5meC erasure strategy. A catalytically inactive Cas9 (dCas9, gray rectangle) fused to a SunTag (five gray circles) is recruited to target chromatin by one or several gRNAs and in turn recruits the catalytic domain of TET1 (purple) via scFv interactions with the SunTag. Cells are differentiated for 4 days (left) and assessed for 5meC levels, CTCF binding, nearby gene expression and 3D conformation compared to control cells expressing non-target gRNA (right). (**B**) Bisulphite-pyrosequencing results of cells expressing non-target sgRNA (black) and cells expressing sgRNAs (purple) targeted to the *Zdbf2* CTCF binding site. The position of each CpG within the amplicon is indicated, and the CpG corresponding to CpG5 in the canonical CTCF binding motif is highlighted by the red box. Data are shown as mean ± standard error for three replicates. (**C**) ChIP-qPCR results of the same cells as in B. Data are shown as mean ± standard error for three replicates represented by unfilled circles. (**D**) RT-qPCR results of the same cells as in B over a time course of 7 days of EpiLC differentiation. Expression of each replicate was normalized to two housekeeping genes (*Rrm2* & *Rplp0*), and then to WT ESCs. Data are shown as mean ± standard error for three replicates. (**E**) UCSC genome browser screenshot of the *Zdbf2* locus showing 4C-seq results of the same cells as in B. The TKO-specific CTCF HiChIP loop (orange) is included for reference. Grey loops represent significant interactions shared between WT and TKO EpiLCs. The TKO-specific CTCF binding site was used as a viewpoint. Note that genomic contacts generally increase on the left side of the plot in the TET1-edited condition, when CTCF binding at the viewpoint is enriched. *P*-values were calculated by two-tailed paired t-test assuming unequal variance: $^{*}P < 0.05$, $^{**}P < 0.01$, $^{**}P < 0.001$, $^{****}P < 0.0001$.

recreated at the TET1-edited *Nrp2* locus, there is no clear regulatory impact (Supplementary Figure S11C).

However, fitting with our model, the enriched CTCF at the epigenome-edited *Zdbf2* locus led to a significant decrease in *Zdbf2* expression (Figure 4C, D). Consistently, 4C-seq data indicated the presence of the TKO-specific loop structure when TET1 was targeted to the CTCF-binding site (Figure 4E). We went on to characterize the CTCF-mediated control of this locus in more depth.

CTCF and Polycomb both coordinate repression of *Zdbf2* in the hypomethylated state

We were intrigued by our dCas9-SunTag/TET1 editing results at *Zdbf2*, and were motivated to perform further genetic tests to substantiate our model of CTCF insulating the four enhancers from contacting the *Zdbf2* promoter (Figure 5A, B). To do this, we generated a homozygous deletion mutant of the CTCF binding site in WT ESCs (Figure 5A, Supplementary Figure S13A). In ESCs lacking the CTCF binding site, indeed we observed a minor increase in *Zdbf2* expression (Figure 5C). As mentioned, *Zdbf2* is polycomb repressed in ESCs (Figure 5A, B), and we previously demonstrated that addition of a PRC2 inhibitor to ESC culture media leads to mild de-repression (43). We performed the same experiment here, and recapitulated the mild effect observed in the WT background (Figure 5C, Supplementary Figure S13B). Strikingly, we observed a substantial upregulation (∼17 fold) when we added the PRC2 inhibitor to cells lacking the CTCF binding site (Figure 5C). These data strongly suggest that polycomb and CTCF synergistically cooperate to maintain *Zdbf2* repression in the hypomethylated state—H3K27me3 is enriched TKO EpiLCs as well (42,45)—and the *de novo* DNA methylation program is required to release both of these means of control (Figure 5A-C). Finally, consistent with our prediction, in EpiLCs when the DNA methylation levels are high and CTCF is no longer bound, the deletion of the CTCF binding site did not lead to an effect on *Zdbf2* expression (Supplementary Figure S13C).

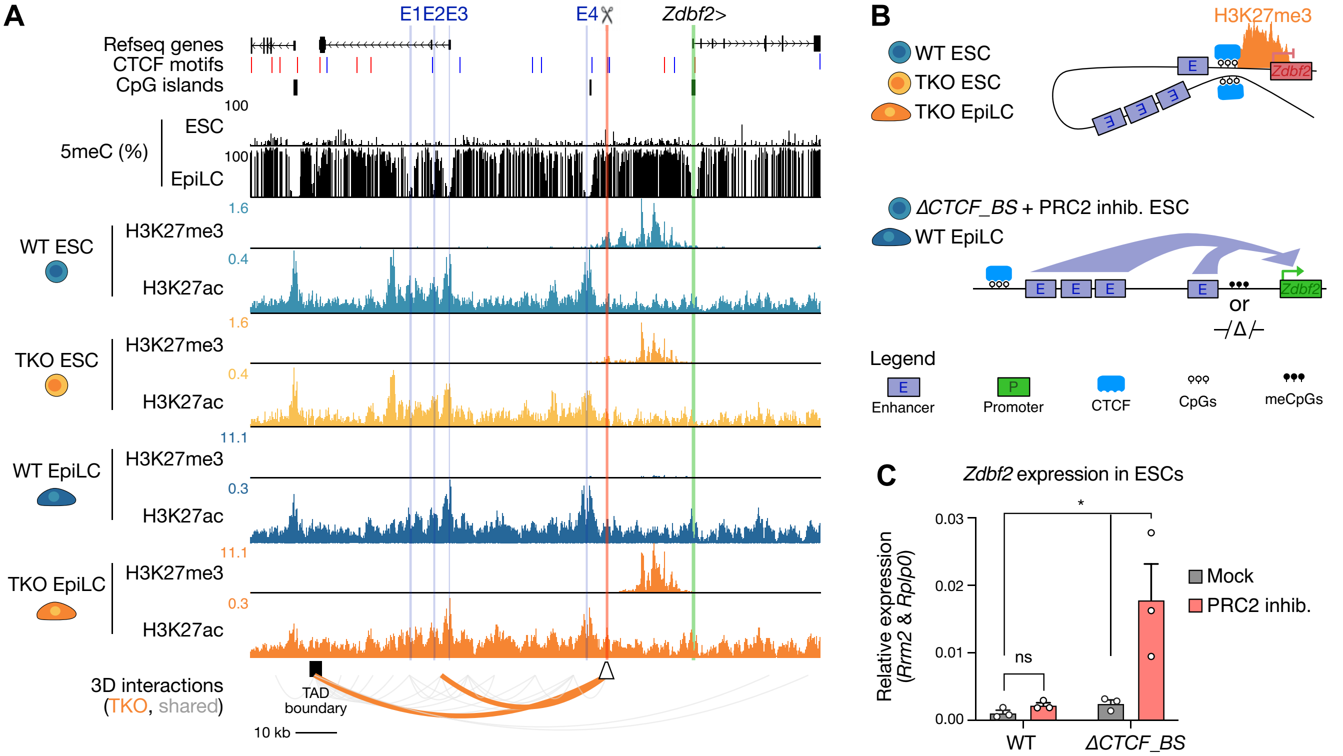


Figure 5. CTCF insulates promoter-enhancer interactions at the *Zdbf2* locus. (**A**) UCSC genome browser screenshot of the *Zdbf2* locus. CpG methylation levels in WT ESCs and EpiLCs are shown, followed by H3K27me3 (45) and H3K27ac (HiChIP) profiles in WT and TKO ESC and EpiLCs. Refseq genes, CpG islands, CTCF binding motifs and the TAD boundary are shown. Previously defined enhancers (E1-4) are shown in blue, the *Zdbf2* CGI promoter in green, and the CTCF binding site deletion is highlighted in red. Statistically enriched CTCF loops in TKO EpiLCs (as shown in Figure 3) are included (orange). Gray loops represent significant interactions that are consistent across WT and TKO EpiLCs. *Zdbf2*: chr1:63165972–63304123. (**B**) Schema showing the hypothesized insulating effect of CTCF binding on *Zdbf2* promoter-enhancer looping. In hypomethylated cells where CTCF is bound and enhancer sequences are insulated, the promoter sequence is marked by PRC2-associated H3K27me3 and repressed (top). When the CTCF binding site is DNA methylated or deleted, the enhancers have increased access to the *Zdbf2* promoter, leading to increased expression (bottom). (**C**) Bar chart showing the relative expression levels of *Zdbf2* by RT-qPCR in WT and *ΔCTCF_Binding Site* (*ΔCTCF_BS*) ESCs treated with UNC1999 (PRC2 inhibitor) or UNC2400 (mock). Data are shown as mean ± standard error for three replicates. *P*-values were calculated by two-tailed paired *t*-test: $*P < 0.05$.

Discussion

The CTCF binding landscape varies substantially between cell types, most likely related to its role in orchestrating the *cis*-regulatory interactions that help define cell identity. As 5meC antagonizes CTCF in certain contexts, we utilized an ESC to EpiLC differentiation system as this particular trajectory is associated with a dramatic increase in global DNA methylation levels. Given the dramatic genome-wide gain of 5meC, one could conclude that only 1% of CTCF binding sites impacted represents a minor effect. On the other hand, 1% comprises hundreds of sites, and impaired CTCF binding could directly influence the expression of hundreds or thousands of genes. This upper limit would constitute a substantial effect on genome regulation. Indeed, at day four of EpiLC differentiation, 1459 genes were misregulated in TKO cells (FC ≥ 2, adjusted $P \leq 0.05$). However our analyses would suggest that a small fraction of the total number of misregulated genes was likely due to direct CTCF-mediated control; it should be emphasized that expression of a greater subset of genes could be impacted via downstream effects, and there were more subtle changes in looping and/or expression that did not meet our thresholds.

Nevertheless, we went on to examine four loci due to the likely effect that 5meC had on CTCF binding and looping. We did not discover a clear biological process linking these four genes. *Csf1* encodes a macrophage-stimulating factor that is widely expressed across mouse tissues, and is important for maintenance of tissue-specific macrophage populations (103). *Mob3b* is part of the highly conserved monopolar spindle-one-binder (MOB) gene family, and while the protein products have no known enzymatic function, they are thought to be scaffold proteins, and are linked with a number of human diseases (104). *Nrp2* codes for a transmembrane protein that contributes to a number of signaling pathways that contribute to the cytoskeleton, angiogenesis, and cancer progression (105). Conversely, we have extensively described *Zdbf2*. Interestingly, although it is expressed during early embryonic stages, *Zdbf2* mutants exhibit no obvious phenotypes either in ESCs nor in the *in vivo* embryo; rather, *Zdbf2* expression appears important in the postnatal hypothalamus (42,97). Thus, it is possible that while the genes we described in this study were protected from ectopic CTCF-mediated gene control in WT EpiLCs (and potentially the *in vivo* epiblast) via 5meC deposition, we uncovered regulatory mechanisms that are biologically relevant in other cell types at later developmental stages. One could imagine that toggling the DNA methylation at the CTCF binding sites—via natural TET-mediated demethylation, for example (92), akin to what we did with our artificial dCas9 system—could tune the expression of the linked genes to ensure proper cellular function.

Related to this point, a substantial number of potentially DNA methylation sensitive binding sites could be masked due to the fact that even in WT EpiLCS, when global 5meC levels were elevated, the CTCF-bound region remained hypomethylated (Figure 1E). CTCF itself may protect from *de novo* DNA methylation at a substantial number of its binding sites (22,106,107). Additionally, these elements are likely protected by TET protein activity (92) and/or combinatorial transcription factor binding (106,108,109). In other words, there is potential that DNA methylation could be substantially disruptive to the 3D regulatory structure of the genome absent of factors that deter DNA methylation machinery. This is hinted at by the fact that in some cancers, where the DNA methylome is broadly misregulated, there are notable examples of 5meC accumulation leading to CTCF loss resulting in oncogene expression (32,33), or repression of a tumor suppressor (110). A previous study examined CTCF binding in a mouse ESC line harboring mutations in *Tet1* and *Tet2* (as opposed to our study here, the cells were cultured in conditions such that DNA methylation levels were globally high) (92). They observed that increased DNA methylation and loss of CTCF binding adjacent to promoters led to reduced gene expression; however, they employed a metastable cellular state and did not formally assess the impact of differential CTCF binding on chromatin conformation. Fodder for future studies will be to remove protection mechanisms, such as TET enzymes or pertinent transcription factors, in order to observe if ectopic gain of DNA methylation at discrete sites will have a more severe impact on CTCF-regulated genes than a complete loss of 5meC.

We utilized the EpiLC differentiation technique because it allowed for us to interrogate a cellular transition in the complete absence of DNA methylation. In most differentiated cell types DNA methylation is absolutely required, and acute 5meC loss leads to widespread epigenetic misregulation and cell death (111). However, EpiLCs are distinguished from other highly DNA methylated cell types in that they are still pluripotent, and express many of the same transcription factors that are linked with the pluripotency network (112). It is possible that this property is directly related with the resilience EpiLCs show in the absence of DNA methylation. In other words, there could be the same mechanisms in place in EpiLCs that allow DNA hypomethylated naïve ESCs to proliferate without fatal genomic instability. This could include ensuring that CTCF-mediated genome architecture largely stays intact independently of the underlying DNA methylation state. It is possible that in other more differentiated cell types, acute depletion of DNA methylation may lead to a more drastic effect than we observed in our TKO EpiLCs.

How then to bypass the cell death phenotype in DNA methylation mutants? Many chromatin conformation studies take advantage of degron technology and assay genome folding in the window between protein depletion and cell death. Such techniques have been successfully utilized to understand the role of CTCF (18), cohesin (113,114), the Mediator complex (115), and RNA polymerase II (116–118). With high resolution techniques, such as HiChIP or Micro-C (119–121), a degron system can be coupled with an assessment of the *cis*-regulatory interactome (118). Such techniques could be adapted for DNA methylation degrons (eg, DNMT1) in differentiated cell types in order to gauge the impact of 5meC on the 3D genome.

Nevertheless, our EpiLC system did reveal a number of DNA methylation-sensitive CTCF binding events. The emergence of epigenome editing has enabled the direct assessment of the effect of a chromatin modification at a locus of interest without generating genetic mutants that exhibit potential confounding effects. Not only are these powerful tools to modify CTCF binding in cell culture systems, as described here and elsewhere (98,122,123), but they can also be implemented *in vivo* (123). Indeed, a SunTag/TET system highly similar to the one we utilized here has been successfully employed in mouse embryos to target the *H19-Igf2* imprint, which disrupted CTCF binding and *Igf2* expression (99,124). The prospect of using epigenome editing in the developing embryo proper to modify chromatin architecture and distill the physiological consequences presents a compelling endeavor for future studies.

Data availability

High throughput sequencing data was uploaded to NCBI GEO under accession number GSE246984. See Supplementary Table S2 for the full list of data analyzed in this study. Custom scripts are available under an GNU General Public License v3.0 on GitHub: https://github.com/julienrichardalbert/3DNAmethylation/releases/tag/v0.1 and Zenodo https://zenodo.org/records/11654445.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank Ana Paula Azambuja for sharing differential looping R scripts and Joël Marchand for maintaining computational resources. We thank Ignacio Mazon Pelaez from Arima Genomics for advice for HiChIP experiments. We thank Roshani Singh and Julie Segueni for their assistance with the project. We thank Teresa Urli for critical reading of the manuscript. We thank members of the Greenberg lab for useful discussions.

Author contributions: A.M.-S., J.R.A. and M.V.C.G. conceived the study and designed the experiments. A.M-S. and M.S. performed the experiments. J.R.A. performed the genomic analyses, with assistance from A.M-S and D.N.. D.N. provided guidance for HiChIP analyses and interpretation. A.M.S., J.R.A. and M.V.C.G. wrote the paper. All authors edited the paper.

Funding

Work in the Greenberg group is supported by the European Research Council [ERC-StG-2019 DyNAmecs]; Laboratoire d'excellence Who Am I? [Labex 11-LABX-0071] Emerging Teams Grant and funds from the Agence National de Recherche [ANR, project ANR-21-CE12-0015-03]; A.M.S. is supported by FRM [SPF202004011789]; ARC [ARCPDF12020070002563] postdoctoral fellowships; J.R.A is supported by a Fondation pour la Recherche Médicale, Post doc France Fellowship [SPF202110014238]; work in the Noordermeer group is supported by funds from the ANR [ANR-21-CE12-0034-01, ANR-22-CE12-0016-03, ANR-22-CE14-0021-02]; PlanCancer [19CS145-00]; we acknowledge the sequencing and bioinformatics expertise of the I2BC High-throughput sequencing facility, supported by France Génomique (funded by the French National Program 'Investissement d'Avenir' ANR-10-INBS-09). Funding for open access charge: ERC-StG-2019 DyNAmecs [851054].

Conflict of interest statement

None declared.

References

1. Greenberg,M.V.C. and Bourc'his,D. (2019) The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.*, **20**, 590–607.
2. Nichols,J., Silva,J., Roode,M. and Smith,A. (2009) Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development*, **136**, 3215–3222.
3. Argelaguet,R., Clark,S.J., Mohammed,H., Stapel,L.C., Krueger,C., Kapourani,C.-A., Imaz-Rosshandler,I., Lohoff,T., Xiang,Y., Hanna,C.W., *et al.* (2019) Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature*, **576**, 487–491.
4. Lieberman-Aiden,E., van Berkum,N.L., Williams,L., Imakaev,M., Ragoczy,T., Telling,A., Amit,I., Lajoie,B.R., Sabo,P.J., Dorschner,M.O., *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
5. Cremer,T. and Cremer,M. (2010) Chromosome territories. *Cold Spring Harb. Perspect. Biol.*, **2**, a003889.
6. 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.
7. Bonev,B., Mendelson Cohen,N., Szabo,Q., Fritsch,L., Papadopoulos,G.L., Lubling,Y., Xu,X., Lv,X., Hugnot,J.P., Tanay,A., *et al.* (2017) Multiscale 3D genome rewiring during mouse neural development. *Cell*, **171**, 557–572.
8. Chen,Y., Zhang,Y., Wang,Y., Zhang,L., Brinkman,E.K., Adam,S.A., Goldman,R., van Steensel,B., Ma,J. and Belmont,A.S. (2018) Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J. Cell Biol.*, **217**, 4025–4048.
9. 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.
10. da Costa-Nunes,J.A. and Noordermeer,D. (2023) TADs: dynamic structures to create stable regulatory functions. *Curr. Opin. Struct. Biol.*, **81**, 102622.
11. Nora,E.P., Dekker,J. and Heard,E. (2013) Segmental folding of chromosomes: a basis for structural and regulatory chromosomal neighborhoods?*Bioessays*, **35**, 818–828.
12. Tolhuis,B., Palstra,R.J., Splinter,E., Grosveld,F. and de Laat,W. (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell*, **10**, 1453–1465.
13. Schoenfelder,S. and Fraser,P. (2019) Long-range enhancer–promoter contacts in gene expression control. *Nat. Rev. Genet.*, **20**, 437–455.
14. Zheng,H. and Xie,W. (2019) The role of 3D genome organization in development and cell differentiation. *Nat. Rev. Mol. Cell Biol.*, **20**, 535–550.
15. Monk,D. (2015) Germline-derived DNA methylation and early embryo epigenetic reprogramming: the selected survival of imprints. *Int. J. Biochem. Cell Biol.*, **67**, 128–138.
16. de Wit,E., Vos,E.S.M., Holwerda,S.J.B., Valdes-Quezada,C., Verstegen,M.J.A.M., Teunissen,H., Splinter,E., Wijchers,P.J., Krijger,P.H.L. and de Laat,W. (2015) CTCF binding polarity determines chromatin looping. *Mol. Cell*, **60**, 676–684.
17. Li,Y., Haarhuis,J.H.I., Sedeño Cacciatore,Á., Oldenkamp,R., van Ruiten,M.S., Willems,L., Teunissen,H., Muir,K.W., de Wit,E., Rowland,B.D., *et al.* (2020) The structural basis for cohesin–CTCF-anchored loops. *Nature*, **578**, 472–476.
18. Nora,E.P., Goloborodko,A., Valton,A.L., Gibcus,J.H., Uebersohn,A., Abdennur,N., Dekker,J., Mirny,L.A. and Bruneau,B.G. (2017) Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell*, **169**, 930–944.
19. Arzate-Mejía,R.G., Recillas-Targa,F. and Corces,V.G. (2018) Developing in 3D: the role of CTCF in cell differentiation. *Development*, **145**, dev137729.
20. Kubo,N., Ishii,H., Xiong,X., Bianco,S., Meitinger,F., Hu,R., Hocker,J.D., Conte,M., Gorkin,D., Yu,M., *et al.* (2021) Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nat. Struct. Mol. Biol.*, **28**, 152–161.
21. Moore,J.M., Rabaia,N.A., Smith,L.E., Fagerlie,S., Gurley,K., Loukinov,D., Disteche,C.M., Collins,S.J., Kemp,C.J., Lobanenkov,V.V., *et al.* (2012) Loss of maternal CTCF is associated with peri-implantation lethality of ctcf null embryos. *PLoS One*, **7**, e34915.
22. Kemp,C.J., Moore,J.M., Moser,R., Bernard,B., Teater,M., Smith,L.E., Rabaia,N.A., Gurley,K.E., Guinney,J., Busch,S.E., *et al.* (2014) CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer. *Cell Rep.*, **7**, 1020–1029.
23. Wang,H., Maurano,M.T., Qu,H., Varley,K.E., Gertz,J., Pauli,F., Lee,K., Canfield,T., Weaver,M., Sandstrom,R., *et al.* (2012) Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.*, **22**, 1680–1688.
24. Sahu,B., Hartonen,T., Pihlajamaa,P., Wei,B., Dave,K., Zhu,F., Kaasinen,E., Lidschreiber,K., Lidschreiber,M., Daub,C.O., *et al.* (2022) Sequence determinants of human gene regulatory elements. *Nat. Genet.*, **54**, 283–294.
25. Kreibich,E., Kleinendorst,R., Barzaghi,G., Kaspar,S. and Krebs,A.R. (2023) Single-molecule footprinting identifies context-dependent regulation of enhancers by DNA methylation. *Mol. Cell*, **83**, 787–802.
26. Kaaij,L.J.T., Mohn,F., van der Weide,R.H., de Wit,E. and Bühler,M. (2019) The ChAHP complex counteracts chromatin looping at CTCF sites that emerged from SINE expansions in mouse. *Cell*, **178**, 1437–1451.
27. Monteagudo-Sánchez,A., Noordermeer,D. and Greenberg,M.V.C. (2024) The impact of DNA methylation on CTCF-mediated 3D genome organization. *Nat. Struct. Mol. Biol.*, **31**, 404–412.
28. Hashimoto,H., Wang,D., Horton,J.R., Zhang,X., Corces,V.G. and Cheng,X. (2017) Structural basis for the versatile and methylation-dependent binding of CTCF to DNA. *Mol. Cell*, **66**, 711–720.
29. Hassan-Zadeh,V., Rugg-Gunn,P. and Bazett-Jones,D.P. (2017) DNA methylation is dispensable for changes in global chromatin architecture but required for chromocentre formation in early stem cell differentiation. *Chromosoma*, **126**, 605–614.
30. Nothjunge,S., Nührenberg,T.G., Grüning,B.A., Doppler,S.A., Preissl,S., Schwaderer,M., Rommel,C., Krane,M., Hein,L. and Gilsbach,R. (2017) DNA methylation signatures follow preformed chromatin compartments in cardiac myocytes. *Nat. Commun.*, **8**, 1667.
31. Jiang,Q., Ang,J.Y.J., Lee,A.Y., Cao,Q., Li,K.Y., Yip,K.Y. and Leung,D.C.Y. (2020) G9a Plays distinct roles in maintaining DNA methylation, retrotransposon silencing, and chromatin looping. *Cell Rep.*, **33**, 108315.
32. Flavahan,W.A., Drier,Y., Liau,B.B., Gillespie,S.M., Venteicher,A.S., Stemmer-Rachamimov,A.O., Suvà,M.L. and Bernstein,B.E. (2016) Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*, **529**, 110–114.

33. Flavahan,W.A., Drier,Y., Johnstone,S.E., Hemming,M.L., Tarjan,D.R., Hegazi,E., Shareef,S.J., Javed,N.M., Raut,C.P., Eschle,B.K., *et al.* (2019) Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs. *Nature*, **575**, 229–233.
34. Hark,A.T., Schoenherr,C.J., Katz,D.J., Ingram,R.S., Levorse,J.M. and Tilghman,S.M. (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature*, **405**, 486–489.
35. Bell,A.C. and Felsenfeld,G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, **405**, 482–485.
36. Llères,D., Moindrot,B., Pathak,R., Piras,V., Matelot,M., Pignard,B., Marchand,A., Poncelet,M., Perrin,A., Tellier,V., *et al.* (2019) CTCF modulates allele-specific sub-TAD organization and imprinted gene activity at the mouse Dlk1-Dio3 and Igf2-H19 domains. *Genome Biol.*, **20**, 272.
37. Fang,C., Wang,Z., Han,C., Safgren,S.L., Helmin,K.A., Adelman,E.R., Serafin,V., Basso,G., Eagen,K.P., Gaspar-Maia,A., *et al.* (2020) Cancer-specific CTCF binding facilitates oncogenic transcriptional dysregulation. *Genome Biol.*, **21**, 247.
38. Rahme,G.J., Javed,N.M., Puorro,K.L., Xin,S., Hovestadt,V., Johnstone,S.E. and Bernstein,B.E. (2023) Modeling epigenetic lesions that cause gliomas. *Cell*, **186**, 3674–3685.
39. Blaschke,K., Ebata,K.T., Karimi,M.M., Zepeda-Martínez,J.A., Goyal,P., Mahapatra,S., Tam,A., Laird,D.J., Hirst,M., Rao,A., *et al.* (2013) Vitamin C induces tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature*, **500**, 222–226.
40. Guo,G., Yang,J., Nichols,J., Hall,J.S., Eyres,I., Mansfield,W. and Smith,A. (2009) Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development*, **136**, 1063–1069.
41. Tsumura,A., Hayakawa,T., Kumaki,Y., Takebayashi,S.I., Sakaue,M., Matsuoka,C., Shimotohno,K., Ishikawa,F., Li,E., Ueda,H.R., *et al.* (2006) Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells*, **11**, 805–814.
42. Greenberg,M.V.C., Glaser,J., Borsos,M., Marjou,F.E., Walter,M., Teissandier,A. and Bourc'his,D. (2017) Transient transcription in the early embryo sets an epigenetic state that programs postnatal growth. *Nat. Genet.*, **49**, 110–118.
43. Greenberg,M., Teissandier,A., Walter,M., Noordermeer,D. and Bourc'his,D. (2019) Dynamic enhancer partitioning instructs activation of a growth-related gene during exit from naïve pluripotency. *eLife*, **8**, e44057.
44. Schulz,M., Teissandier,A., de la MataSantaella,E., Armand,M., Iranzo,J., El Marjou,F., Gestraud,P., Walter,M., Kinston,S., Göttgens,B., *et al.* (2024) DNA methylation restricts coordinated germline and neural fates in embryonic stem cell differentiation. *Nat. Struct. Mol. Biol.*, **31**, 102–114.
45. Richard Albert,J., Urli,T., Monteagudo-Sánchez,A., Le Breton,A., Sultanova,A., David,A., Schulz,M. and Greenberg,M.V.C. (2024) DNA methylation shapes the polycomb landscape during the exit from naïve pluripotency. bioRxiv doi: https://doi.org/10.1101/2023.09.14.557729, 14 September 2023, preprint: not peer reviewed.
46. Feldmann,A., Ivanek,R., Murr,R., Gaidatzis,D., Burger,L. and Schübeler,D. (2013) Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet.*, **9**, e1003994.
47. Stadler,M.B., Murr,R., Burger,L., Ivanek,R., Lienert,F., Schöler,A., van Nimwegen,E., Wirbelauer,C., Oakeley,E.J., Gaidatzis,D., *et al.* (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, **480**, 490–495.
48. Guo,G., Pinello,L., Han,X., Lai,S., Shen,L., Lin,T.-W., Zou,K., Yuan,G.-C. and Orkin,S.H. (2016) Serum-based culture conditions provoke gene expression variability in mouse embryonic stem cells as revealed by single-cell analysis. *Cell Rep.*, **14**, 956–965.
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. Dubois,A., Vincenti,L., Chervova,A., Greenberg,M.V.C., Vandormael-Pournin,S., Bourc'his,D., Cohen-Tannoudji,M. and Navarro,P. (2022) H3K9 tri-methylation at Nanog times differentiation commitment and enables the acquisition of primitive endoderm fate. *Development*, **149**, dev201074.
51. Zuin,J., Roth,G., Zhan,Y., Cramard,J., Redolfi,J., Piskadlo,E., Mach,P., Kryzhanovska,M., Tihanyi,G., Kohler,H., *et al.* (2022) Nonlinear control of transcription through enhancer–promoter interactions. *Nature*, **604**, 571–577.
52. Holoch,D., Wassef,M., Lövkvist,C., Zielinski,D., Aflaki,S., Lombard,B., Héry,T., Loew,D., Howard,M. and Margueron,R. (2021) A cis-acting mechanism mediates transcriptional memory at Polycomb target genes in mammals. *Nat. Genet.*, **53**, 1686–1697.
53. Skene,P.J. and Henikoff,S. (2017) An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife*, **6**, e21856.
54. Schmidl,C., Rendeiro,A.F., Sheffield,N.C. and Bock,C. (2015) ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat. Methods*, **12**, 963–965.
55. Miranda,M., Noordermeer,D. and Moindrot,B. (2022) Detection of allele-specific 3D chromatin interactions using high-resolution In-nucleus 4C-seq. *Methods Mol. Biol.*, **2532**, 15–33.
56. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
57. Krueger,F. and Andrews,S.R. (2011) Bismark: a flexible aligner and methylation caller for bisulfite-Seq applications. *Bioinformatics*, **27**, 1571–1572.
58. Wang,L., Zhang,J., Duan,J., Gao,X., Zhu,W., Lu,X., Yang,L., Zhang,J., Li,G., Ci,W., *et al.* (2014) Programming and inheritance of parental DNA methylomes in mammals. *Cell*, **157**, 979–991.
59. Younesy,H., Möller,T., Lorincz,M.C., Karimi,M.M. and Jones,S.J.M. (2015) VisRseq: r-based visual framework for analysis of sequencing data. *BMC Bioinf.*, **16**, S2.
60. Bushnell,B. (2014) In: *BBMap: A Fast, Accurate, Splice-aware Aligner Lawrence Berkeley National Lab*. LBNL, Berkeley, CA (United States).
61. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
62. Ramírez,F., Ryan,D.P., Grüning,B., Bhardwaj,V., Kilpert,F., Richter,A.S., Heyne,S., Dündar,F. and Manke,T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.*, **44**, W160–W5.
63. Meers,M.P., Tenenbaum,D. and Henikoff,S. (2019) Peak calling by Sparse Enrichment analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin*, **12**, 42.
64. Bailey,T.L., Boden,M., Buske,F.A., Frith,M., Grant,C.E., Clementi,L., Ren,J., Li,W.W. and Noble,W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202–W208.
65. Castro-Mondragon,J.A., Riudavets-Puig,R., Rauluseviciute,I., Lemma,R.B., Turchi,L., Blanc-Mathieu,R., Lucas,J., Boddie,P., Khan,A., Manosalva Pérez,N., *et al.* (2022) JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **50**, D165–D173.
66. Chang,L.-H., Ghosh,S., Papale,A., Luppino,J.M., Miranda,M., Piras,V., Degrouard,J., Edouard,J., Poncelet,M., Lecouvreur,N., *et al.* (2023) Multi-feature clustering of CTCF binding creates robustness for loop extrusion blocking and topologically associating domain boundaries. *Nat. Commun.*, **14**, 5615.

67. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
68. Wang,Q., Li,M., Wu,T., Zhan,L., Li,L., Chen,M., Xie,W., Xie,Z., Hu,E., Xu,S., *et al.* (2022) Exploring epigenomic datasets by ChIPseeker. *Curr. Protoc.*, **2**, e585.
69. Pedregosa,F., Varoquaux,G., Gramfort,A., Michel,V., Thirion,B., Grisel,O., Blondel,M., Müller,A., Nothman,J., Louppe,G., *et al.* (2011) Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830.
70. Neph,S., Kuehn,M.S., Reynolds,A.P., Haugen,E., Thurman,R.E., Johnson,A.K., Rynes,E., Maurano,M.T., Vierstra,J., Thomas,S., *et al.* (2012) BEDOPS: high-performance genomic feature operations. *Bioinformatics*, **28**, 1919–1920.
71. Ritchie,M.E., Phipson,B., Wu,D., Hu,Y., Law,C.W., Shi,W. and Smyth,G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.
72. Juric,I., Yu,M., Abnousi,A., Raviram,R., Fang,R., Zhao,Y., Zhang,Y., Qiu,Y., Yang,Y., Li,Y., *et al.* (2019) MAPS: model-based analysis of long-range chromatin interactions from PLAC-seq and HiChIP experiments. *PLoS Comput. Biol.*, **15**, e1006982.
73. Servant,N., Varoquaux,N., Lajoie,B.R., Viara,E., Chen,C.-J., Vert,J.-P., Heard,E., Dekker,J. and Barillot,E. (2015) HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.*, **16**, 259.
74. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W., *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
75. Kruse,K., Hug,C.B. and Vaquerizas,J.M. (2020) FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. *Genome Biol.*, **21**, 303.
76. Lareau,C.A. and Aryee,M.J. (2018) hichipper: a preprocessing pipeline for calling DNA loops from HiChIP data. *Nat. Methods*, **15**, 155–156.
77. Lareau,C.A. and Aryee,M.J. (2018) diffloop: a computational framework for identifying and analyzing differential DNA loops from sequencing data. *Bioinformatics*, **34**, 672–674.
78. Bhattacharyya,S., Chandra,V., Vijayanand,P. and Ay,F. (2019) Identification of significant chromatin contacts from HiChIP data by FitHiChIP. *Nat. Commun.*, **10**, 4221.
79. Wolff,J., Rabbani,L., Gilsbach,R., Richard,G., Manke,T., Backofen,R. and Grüning,B.A. (2020) Galaxy HiCExplorer 3: a web server for reproducible hi-C, capture hi-C and single-cell hi-C data analysis, quality control and visualization. *Nucleic Acids Res.*, **48**, W177–W184.
80. Wolff,J., Bhardwaj,V., Nothjunge,S., Richard,G., Renschler,G., Gilsbach,R., Manke,T., Backofen,R., Ramírez,F. and Grüning,B.A. (2018) Galaxy HiCExplorer: a web server for reproducible hi-C data analysis, quality control and visualization. *Nucleic Acids Res.*, **46**, W11–W16.
81. Ramírez,F., Bhardwaj,V., Arrigoni,L., Lam,K.C., Grüning,B.A., Villaveces,J., Habermann,B., Akhtar,A. and Manke,T. (2018) High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.*, **9**, 189.
82. Raney,B.J., Dreszer,T.R., Barber,G.P., Clawson,H., Fujita,P.A., Wang,T., Nguyen,N., Paten,B., Zweig,A.S., Karolchik,D., *et al.* (2014) Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics*, **30**, 1003–1005.
83. 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.
84. Hunter,J.D. (2007) Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.*, **9**, 90–95.
85. Leitch,H.G., McEwen,K.R., Turp,A., Encheva,V., Carroll,T., Grabole,N., Mansfield,W., Nashun,B., Knezovich,J.G., Smith,A., *et al.* (2013) Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.*, **20**, 311–316.
86. Hackett,J.A. and Surani,M.A. (2014) Regulatory principles of pluripotency: from the ground State up. *Cell Stem Cell*, **15**, 416–430.
87. Marks,H., Kalkan,T., Menafra,R., Denissov,S., Jones,K., Hofemeister,H., Nichols,J., Kranz,A., Stewart,A.F., Smith,A., *et al.* (2012) The transcriptional and epigenomic foundations of ground state pluripotency. *Cell*, **149**, 590–604.
88. von Meyenn,F., Iurlaro,M., Habibi,E., Liu,N.Q., Salehzadeh-Yazdi,A., Santos,F., Petrini,E., Milagre,I., Yu,M., Xie,Z., *et al.* (2016) Impairment of DNA methylation maintenance is the main cause of global demethylation in naive embryonic stem cells. *Mol. Cell*, **62**, 848–861.
89. Walter,M., Teissandier,A., Pérez-Palacios,R., Bourc'his,D., Akalin,A., Kormaksson,M., Li,S., Garrett-Bakelman,F.E., Figueroa,M.E., Melnick,A., *et al.* (2016) An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife*, **5**, e11418.
90. Parry,A., Krueger,C., Lohoff,T., Wingett,S., Schoenfelder,S. and Reik,W. (2023) Dynamic DNA methylation turnover at the exit of pluripotency epigenetically primes gene regulatory elements for hematopoietic lineage specification. bioRxiv doi: https://doi.org/10.1101/2023.01.11.523441, 11 January 2023, preprint: not peer reviewed.
91. Maurano,M.T., Wang,H., John,S., Shafer,A., Canfield,T., Lee,K. and Stamatoyannopoulos,J.A. (2015) Role of DNA methylation in modulating transcription factor occupancy. *Cell Rep.*, **12**, 1184–1195.
92. Wiehle,L., Thorn,G.J., Raddatz,G., Clarkson,C.T., Rippe,K., Lyko,F., Breiling,A. and Teif,V.B. (2019) DNA (de)methylation in embryonic stem cells controls CTCF-dependent chromatin boundaries. *Genome Res.*, **29**, 750–761.
93. Ren,G., Jin,W., Cui,K., Rodrigez,J., Hu,G., Zhang,Z., Larson,D.R. and Zhao,K. (2017) CTCF-mediated enhancer-promoter interaction is a critical regulator of cell-to-cell variation of gene expression. *Mol. Cell*, **67**, 1049–1058.
94. Mumbach,M.R., Satpathy,A.T., Boyle,E.A., Dai,C., Gowen,B.G., Cho,S.W., Nguyen,M.L., Rubin,A.J., Granja,J.M., Kazane,K.R., *et al.* (2017) Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.*, **49**, 1602–1612.
95. Kim,S., Yu,N.-K. and Kaang,B.-K. (2015) CTCF as a multifunctional protein in genome regulation and gene expression. *Exp. Mol. Med.*, **47**, e166.
96. Duffié,R., Ajjan,S., Greenberg,M.V., Zamudio,N., Escamilla del Arenal,M., Iranzo,J., Okamoto,I., Barbaux,S., Fauque,P. and Bourc'his,D. (2014) The Gpr1/Zdbf2 locus provides new paradigms for transient and dynamic genomic imprinting in mammals. *Genes Dev.*, **28**, 463–478.
97. Glaser,J., Iranzo,J., Borensztein,M., Marinucci,M., Gualtieri,A., Jouhanneau,C., Teissandier,A., Gaston-Massuet,C. and Bourc'his,D. (2022) The imprinted Zdbf2 gene finely tunes control of feeding and growth in neonates. *eLife*, **11**, e65641.
98. Morita,S., Noguchi,H., Horii,T., Nakabayashi,K., Kimura,M., Okamura,K., Sakai,A., Nakashima,H., Hata,K., Nakashima,K., *et al.* (2016) Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nat. Biotechnol.*, **34**, 1060–1065.
99. Horii,T., Morita,S., Kimura,M. and Hatada,I. (2022) Efficient generation of epigenetic disease model mice by epigenome editing using the piggyBac transposon system. *Epigenetics Chromatin*, **15**, 40.
100. Noordermeer,D., Leleu,M., Splinter,E., Rougemont,J., De Laat,W. and Duboule,D. (2011) The dynamic architecture of hox gene clusters. *Science*, **334**, 222–225.
101. van de Werken,H.J., Landan,G., Holwerda,S.J., Hoichman,M., Klous,P., Chachik,R., Splinter,E., Valdes-Quezada,C., Oz,Y., Bouwman,B.A., *et al.* (2012) Robust 4C-seq data analysis to

screen for regulatory DNA interactions. *Nat. Methods*, **9**, 969–972.
102. Zukher,I., Dujardin,G., Sousa-Luís,R. and Proudfoot,N.J. (2023) Elongation roadblocks mediated by dCas9 across human genes modulate transcription and nascent RNA processing. *Nat. Struct. Mol. Biol.*, **30**, 1536–1548.
103. Sehgal,A., Irvine,K.M. and Hume,D.A. (2021) Functions of macrophage colony-stimulating factor (CSF1) in development, homeostasis, and tissue repair. *Semin. Immunol.*, **54**, 101509.
104. Gundogdu,R. and Hergovich,A. (2019) MOB (Mps one Binder) proteins in the Hippo pathway and cancer. *Cells*, **8**, 569.
105. Grandclement,C., Pallandre,J.R., Valmary Degano,S., Viel,E., Bouard,A., Balland,J., Rémy-Martin,J.-P., Simon,B., Rouleau,A., Boireau,W., *et al.* (2011) Neuropilin-2 expression promotes TGF-β1-mediated epithelial to mesenchymal transition in colorectal cancer cells. *PLoS One*, **6**, e20444.
106. Lienert,F., Wirbelauer,C., Som,I., Dean,A., Mohn,F. and Schübeler,D. (2011) Identification of genetic elements that autonomously determine DNA methylation states. *Nat. Genet.*, **43**, 1091–1097.
107. Pant,V., Mariano,P., Kanduri,C., Mattsson,A., Lobanenkov,V., Heuchel,R. and Ohlsson,R. (2003) The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the H19 imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. *Genes Dev.*, **17**, 586–590.
108. Krebs,A.R., Dessus-Babus,S., Burger,L. and Schübeler,D. (2014) High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. *eLife*, **3**, e04094.
109. Kremsky,I. and Corces,V.G. (2020) Protection from DNA re-methylation by transcription factors in primordial germ cells and pre-implantation embryos can explain trans-generational epigenetic inheritance. *Genome Biol.*, **21**, 118.
110. Rodriguez,C., Borgel,J., Court,F., Cathala,G., Forné,T. and Piette,J. (2010) CTCF is a DNA methylation-sensitive positive regulator of the INK/ARF locus. *Biochem. Biophys. Res. Commun.*, **392**, 129–134.
111. Jackson-Grusby,L., Beard,C., Possemato,R., Tudor,M., Fambrough,D., Csankovszki,G., Dausman,J., Lee,P., Wilson,C., Lander,E., *et al.* (2001) Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat. Genet.*, **27**, 31–39.
112. Takahashi,S., Kobayashi,S. and Hiratani,I. (2018) Epigenetic differences between naïve and primed pluripotent stem cells. *Cell. Mol. Life Sci.*, **75**, 1191–1203.
113. Rao,S.S.P., Huang,S.-C., Glenn St Hilaire,B., Engreitz,J.M., Perez,E.M., Kieffer-Kwon,K.-R., Sanborn,A.L., Johnstone,S.E., Bascom,G.D., Bochkov,I.D., *et al.* (2017) Cohesin loss eliminates all loop domains. *Cell*, **171**, 305–320.
114. Wutz,G., Várnai,C., Nagasaka,K., Cisneros,D.A., Stocsits,R.R., Tang,W., Schoenfelder,S., Jessberger,G., Muhar,M., Hossain,M.J., *et al.* (2017) Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.*, **36**, 3573–3599.
115. El Khattabi,L., Zhao,H., Kalchschmidt,J., Young,N., Jung,S., Van Blerkom,P., Kieffer-Kwon,P., Kieffer-Kwon,K.-R., Park,S., Wang,X., *et al.* (2019) A pliable mediator acts as a functional rather than an architectural bridge between promoters and enhancers. *Cell*, **178**, 1145–1158.
116. Jiang,Y., Huang,J., Lun,K., Li,B., Zheng,H., Li,Y., Zhou,R., Duan,W., Wang,C., Feng,Y., *et al.* (2020) Genome-wide analyses of chromatin interactions after the loss of Pol I, Pol II, and Pol III. *Genome Biol.*, **21**, 158.
117. Sun,F., Sun,T., Kronenberg,M., Tan,X., Huang,C. and Carey,M.F. (2021) The Pol II preinitiation complex (PIC) influences mediator binding but not promoter–enhancer looping. *Genes Dev.*, **35**, 1175–1189.
118. Zhang,S., Übelmesser,N., Barbieri,M. and Papantonis,A. (2023) Enhancer-promoter contact formation requires RNAPII and antagonizes loop extrusion. *Nat. Genet.*, **55**, 832–840.
119. Krietenstein,N., Abraham,S., Venev,S.V., Abdennur,N., Gibcus,J., Hsieh,T.-H.S., Parsi,K.M., Yang,L., Maehr,R., Mirny,L.A., *et al.* (2020) Ultrastructural details of mammalian chromosome architecture. *Mol. Cell*, **78**, 554–565.
120. Hsieh,T.-H.S., Weiner,A., Lajoie,B., Dekker,J., Friedman,N. and Rando,O.J. (2015) Mapping nucleosome resolution chromosome folding in yeast by micro-C. *Cell*, **162**, 108–119.
121. Hsieh,T.-H.S., Cattoglio,C., Slobodyanyuk,E., Hansen,A.S., Rando,O.J., Tjian,R. and Darzacq,X. (2020) Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol. Cell*, **78**, 539–553.
122. Tarjan,D.R., Flavahan,W.A. and Bernstein,B.E. (2019) Epigenome editing strategies for the functional annotation of CTCF insulators. *Nat. Commun.*, **10**, 4258.
123. Liu,X.S., Wu,H., Ji,X., Stelzer,Y., Wu,X., Czauderna,S., Shu,J., Dadon,D., Young,R.A. and Jaenisch,R. (2016) Editing DNA methylation in the mammalian genome. *Cell*, **167**, 233–247.
124. Horii,T., Morita,S., Hino,S., Kimura,M., Hino,Y., Kogo,H., Nakao,M. and Hatada,I. (2020) Successful generation of epigenetic disease model mice by targeted demethylation of the epigenome. *Genome Biol.*, **21**, 77.

Received: November 20, 2023. **Revised:** July 23, 2024. **Editorial Decision:** July 27, 2024. **Accepted:** August 21, 2024