EpiMethylTag simultaneously detects ATAC-seq or ChIP-seq signals with DNA methylation

Priscillia Lhoumaud[1], Gunjan Sethia[1], Franco Izzo[2,3], Sana Badri[1], MacIntosh Cornwell[1], Kyu-Tae Kim[2,3], Dan Avi Landau[2,3,4], Jane Skok[1#].

Affiliations: [1]New York University Langone Health, New York, NY, USA; [2]New York Genome Center, New York, NY, USA; [3]Meyer Cancer Center, Weill Cornell Medicine, New York, NY, USA; [4]Institute of Computational Biomedicine, Weill Cornell Medicine, New York, NY, USA.

[#] Corresponding author.

Correspondence: Jane.Skok@nyulangone.org

Key words: DNA methylation, chromatin accessibility, CTCF.

Abstract

Activation of regulatory elements is thought to be inversely correlated with DNA methylation levels. However, it is difficult to determine whether DNA methylation is compatible with chromatin accessibility or transcription factor (TF) binding if assays are performed separately. We developed a low input, low sequencing depth method, EpiMethylTag that combines ATAC-seq or ChIP-seq (M-ATAC or M-ChIP) with bisulfite conversion, to simultaneously examine accessibility/TF binding and methylation on the same DNA.

Main

The role of DNA methylation (DNAme) in gene regulation has been widely described[1-4]. In general, methylation is thought to reduce accessibility and prohibit TF binding at enhancers and promoters[5,6]. Nevertheless, TFs are also known to bind methylated DNA[2], but due to limitations in the techniques available for this kind of analysis, few genome wide studies have been performed. As a result, we still know very little about the DNA sequence and chromatin context of TF binding at methylated sites and its significance to gene regulation.

Several techniques have been developed to measure DNAme, some more comprehensive than others. Whole genome bisulfite sequencing (WGBS) covers all genomic regions, however to achieve sufficient sequencing coverage is costly. The alternative, reduced representation bisulfite sequencing (RRBS), which requires less sequencing depth, preferentially captures CpG-dense sequences known as CpG islands that can potentially act as regulatory elements[7]. Nevertheless, both techniques require additional assays on different batches of cells to elucidate the interplay between DNAme, DNA accessibility and TF binding and this does not satisfactorily address the issue of compatibility. Current techniques that simultaneously analyze methylation together with TF binding or accessibility (NOME-seq[8], HT-SELEX[9], ChIP-bisulfite[10], BisChIP-seq[11], ChIP-BisSeq[12]), all have drawbacks such as analysis of DNA rather than chromatin, the requirement of large numbers of cells or high sequencing costs.

To circumvent the high input or sequencing expense associated with WGBS and existing ChIP combined with bisulfite conversion protocols[10-12], we developed 'EpiMethylTag'. This technique combines ATAC-seq or ChIPmentation[13,14] with bisulfite conversion (M-ATAC or M-ChIP, respectively) to specifically determine the methylation status of accessible or TF-bound regions in a chromatin context. EpiMethylTag is based on an approach that was originally developed for tagmentation-based WGBS[15,16]. It involves use of the Tn5 transposase, loaded with adapters

harboring cytosine methylation (**Supplementary table 1**). For M-ATAC or M-ChIP, tagmentation occurs respectively on nuclear lysates as per the conventional ATAC-seq protocol[13], or during chromatin immunoprecipitation as per the ChIPmentation protocol[14]. Following DNA purification, the sample is bisulfite converted and PCR amplified for downstream sequencing (**Fig. 1a**). As shown in **Fig. 1a**, EpiMethylTag can determine whether DNAme and accessibility/TF binding are mutually exclusive (scenario 1) or can coexist in certain locations (scenario 2). The protocol requires fewer cells, less sequencing depth, is quicker than existing methods and can be analyzed using a pipeline we developed that is publicly available online on Github (*"https://github.com/skoklab/EpiMethylTag"*).

M-ATAC and CTCF M-ChIP were performed in duplicate on murine embryonic stem cells (mESC). As controls, we collected aliquots before bisulfite conversion, ATAC-seq and CTCF ChIPmentation with Nextera transposase[13,14]. As shown in **Fig. 1b and Supplementary Fig. 1a**, genome coverage was highly reproducible between M-ATAC replicates and highly correlated with regular ATAC-seq and M-ATAC signal before bisulfite treatment. Thus, bisulfite treatment, or the use of a different transposase does not result in signal bias. High reproducibility was also seen for CTCF M-ChIP, and we observed consistency between our results and data generated by CTCF ChIP-BisSeq, a similar technique that requires a much larger number of cells[12] (**Fig. 1b and Supplementary Fig. 1a)**). Of note, bisulfite conversion does not affect the number of peaks detected, or the Jaccard index of peak overlap (**Supplementary Fig. 1b**), although it leads to shorter reads (**Supplementary Fig. 2**). Of note, average methylation was higher at the edges of the peaks than at the midpoint (**Supplementary Fig. 3**). Comparable DNA methylation levels were found in M-ATAC and CTCF M-ChIP replicates, Pearson correlation = 0.76 and 0.84, respectively (**Supplementary Fig. 4a and 4b**) and these were pooled for subsequent DNAme analysis.

We focused our analysis only at cytosines within peak regions covered by at least five reads, as methylation outside of M-ATAC and M-ChIP peaks has low coverage and is less reliable. We observe positive correlations between DNA methylation from WGBS and M-ATAC (**Fig. 1c,** top panel, Pearson correlation=0.69), and between methylation levels in M-ChIP and WGBS (**Fig. 1c**, bottom panel, Pearson correlation = 0.74). Similar results were observed with the previously published CTCF ChIP-BisSeq method[12] (GSE39739) (Pearson correlation = 0.83, **Supplementary Fig. 4c**). In **Fig. 1b** we highlight the *Klf4* gene, which harbors a peak of chromatin accessibility in the promoter and CTCF binding in the intragenic region associated with low methylation from both EpiMethylTag and WGBS assays (left panel). In contrast, the *Pisd-ps1* intragenic locus contains accessible chromatin that coexists with high levels of DNA methylation as detected by both M-ATAC and WGBS (**Fig. 1b**, middle panel). Interestingly, a proportion of M-ATAC peaks exhibited an intermediate-to-high average methylation level in deeply sequenced WGBS[17], but low methylation in M-ATAC (**Fig. 1c**, top panel, top left corner) as illustrated at the *Slc5a8* locus (**Fig. 1b**, right panel). These data suggest that, as expected open regions are less methylated than closed regions within a population of cells, but that accessibility and methylation can coexist in a small subset of genomic regions, which are depleted for promoter regions and associated with low transcription (**Supplementary Fig. 4d and 4e**). Importantly, M-ATAC is able to identify methylation levels within ATAC peaks, information that cannot be retrieved from integrating data from separate WGBS and ATAC-seq experiments.

For further analysis, we separated cytosines in M-ATAC peaks according to percentage of methylation (low 0-20%, intermediate 20-80% and high >80%) and read coverage (high > 50 reads and low 5-50 reads) as follows: #1: Low methylation/High coverage; #2: Low Methylation/Low coverage; #3: Intermediate methylation/Low coverage; #4: High methylation/ Low coverage (**Fig. 2a**). As expected, coverage and methylation from M-ATAC are

anticorrelated and we did not detect any cytosines with intermediated or high methylation with high ATAC coverage (>50 reads), a pattern not observed if considering methylation from WGBS (**Supplementary Fig. 5a**). Cytosines in low methylation groups 1 and 2 were enriched at promoters, while cytosines in intermediate and high methylation groups 3 and 4, were respectively enriched in intragenic and intergenic regions (**Fig. 2b**). The average methylation was more negatively correlated with transcriptional output for cytosines at promoters (**Fig. 2c**) than for intragenic cytosines (**Supplementary Fig. 5b**). Intriguingly, H3K4me1 showed a pronounced increase at cytosines with high levels of methylation (group 4) at promoter regions (**Fig. 2d and Supplementary Fig. 5c**). This data suggests that accessible poised promoters have higher levels of methylation. In contrast, H3K27ac and H3K4me3 were enriched at cytosines with low levels of methylation (groups 1 and 2), for both promoters and non-promoters. Using HOMER we detected significant differences in transcription factor motifs in the four groups of methylated cytosines (**Supplementary Table 1**). Interestingly, some *de novo* motifs harboring a CpG were assigned to key pluripotency transcription factors, OCT4, NANOG and KLF4 and the methylation frequency differed between these and known motifs (Supplementary **Fig. 6**). The impact of methylation at these *de novo* motifs on protein binding remains to be elucidated.

As a case study, CTCF M-ChIP was used to analyze the impact of DNAme on CTCF binding in M-ATAC peaks harboring a CTCF motif (**Fig. 3a**). The MA0139.1 CTCF motif from the Jaspar database incorporates 2 CpGs: C2 and/or C12 (**Fig. 3b**). Of note, *de novo* CTCF motifs in CTCF ChIP-seq and Methyl-ChIP peaks were comparable to the MA0139.1 motif (**Supplementary Fig. 7**). Although CTCF occupancy has been inversely correlated with DNA methylation[18], methylation at positions C2 and C12 and the links to CTCF binding have not been examined. Our analysis revealed that M-ATAC peaks containing a CTCF motif have an enriched CTCF intensity at cytosines with low and intermediate levels of methylation (groups 2

and 3) compared to cytosines with low and high levels of methylation (groups 1 and 4) (**Fig. 3c**). These data provide insight into CTCF binding and indicate an anticorrelation between high accessibility and high methylation, consistent with highly-significant CTCF motif enrichment at cytosines with low levels of methylation (groups 2 and 3) (**Supplementary Table 1**). Consistent with the findings from a recent study that analyzed CTCF binding using oligos rather than genomic DNA methylated at positions C2 and C12[19], CTCF M-ChIP detected higher levels of methylation at C12 compared to C2 (**Fig. 3d**, compare CTCF M-ChIP C2 versus C12, p-value = 1.02e-12) Importantly, CTCF M-ChIP is more suitable than WGBS for detecting the differences (**Fig. 3d**, compare CTCF M-ChIP versus WGBS, p-value = 0.023).. In addition, we found that bi-methylation at both CpGs is slightly enriched compared to what is expected by random chance (0.97% versus 0.05%) (**Supplementary Fig. 8a,** χ^2 = 1531, p-value < 0.001). Nonetheless, sequence variation at the C2 and C12 positions appears to have no effect on methylation (**Supplementary Fig. 8b**).

In conclusion, we developed a method, “EpiMethylTag”, that allows the simultaneous analysis of DNA methylation with ChIP-seq or ATAC-seq. “EpiMethylTag” can be used to analyze the methylation status and coincident accessibility or binding of other chromatin bound transcription factors. Using this technique, we confirmed that as a general rule, DNA methylation rarely coexists with DNA accessibility or CTCF binding. However, in contrast to WGBS, M-ATAC and CTCF M-ChIP revealed a complex interplay between accessible chromatin, DNA methylation and CTCF binding. Thus, EpiMethylTag can be used to provide information about the DNA sequence and chromatin context of TF binding at methylated sites and its significance to gene regulation and biological processes. This technique can also be adapted for single cell analysis.

Methods

Methods, including associated accession codes and scripts and references are available at: XXX.

Figure Legends

Fig. 1| EpiMethylTag is a reproducible method to test whether DNAme can coexist or not with TF binding (CTCF) or chromatin accessibility genome-wide. a, Schematic overview of the EpiMethyTag method showing two possible outcomes. **b,** Representative IGV screenshots of EpiMethylTag, at the *Klf4* locus (left panel), the *Pisd-ps1* locus (middle panel), and the *Slc5a8* locus (right panel). ATAC and M-ATAC in green, CTCF in purple and DNA methylation from merged M-ATAC, merged CTCF M-ChIP and WGBS (methylation from 0% in blue to 100% in red). A zoom-in of methylation at the highlighted region is shown at the bottom of each example. The *Klf4* locus illustrates a region that has low methylation as detected by M-ATAC, CTCF M-ChIP and WGBS. *Pisd-ps1* locus illustrates a region that has high methylation as detected by M-ATAC, CTCF M-ChIP and WGBS. *Slc5a8* locus illustrates a region that has low methylation as detected by M-ATAC and high methylation as detected by WGBS. **c,** Density plots of methylation from EpiMethyltag compared with WGBS. Only cytosines inside peaks and with at least 5 reads were considered. Top: average methylation of cytosines per M-ATAC peak in M-ATAC versus WGBS (Pearson Correlation = 0.69, p-value < 2.2e-16; bottom left corner: 27977 peaks, top left corner: 8408 peaks, top right corner: 1019 peaks, bottom right corner: 113 peaks). Bottom: average methylation per CTCF M-ChIP peak of cytosines in CTCF M-ChIP versus WGBS (Pearson Correlation = 0.74, p-value < 2.2e-16; bottom left corner: 6549 peaks, top left corner: 198 peaks, top right corner: 304 peaks, bottom right corner: 310 peaks).

Fig. 2| M-ATAC and CTCF M-ChIP reveal complex interplay between accessible chromatin, DNA methylation and CTCF binding. a, Cytosines in M-ATAC peaks were divided into four groups according to methylation and coverage status: 1. Low Methylation (<20%) + High coverage (>50 reads) (22932 cytosines). 2. Low Methylation + Low coverage (5 to 50 reads) (1348931 cytosines). 3. Intermediate methylation (20-80) + Low coverage (5 to 50 reads)

(39321 cytosines). 4. High methylation (>80%) + Low coverage (5 to 50 reads) (1652 cytosines). *** P=0 between groups #1 + 2 and group #3, ***P=3.25e-109 between groups #3 and 4 (Wilcoxon text). **b,** Genomic annotations for the 4 groups from **Fig. 2a**. Promoter: TSS -1kb and +100bp; intragenic: introns, exons, 5'UTR, 3'UTR and TTS, intergenic: distal from promoter >1kb and non-coding RNAs. **c,** Expression level of genes associated with the four groups of methylated cytosines from in **Fig. 2a**, for the cytosines at promoters. ***P=4.2e-33 between groups #1 and 2, ***P=2.8e-75 between groups #2 and 3, *P=0.034 between groups #3 and 4 (Wilcoxon test). **d,** Average profile of M-ATAC, H3K4me1, H3K4me3 and H3K27ac signal associated with the four groups of methylated cytosines from **Fig 2a** at promoters versus non-promoters. Of note, the small number of promoters in group 4 gives an unsmooth pattern for marks such as H3K4me1 and H3K27ac.

Fig. 3| CTCF M-ChIP enables analysis of DNA methylation of distinct cytosines in the CTCF motif. a, Schematic illustration representing an ATAC-seq peak with a CTCF motif and CTCF occupancy dependent on C2 and C12 methylation. **b,** CTCF motif from JASPAR database (MA0139.1). The 2 key CpG positions (C2 and C12) are indicated. **c,** Heatmaps (top) and average profiles (bottom) of M-ATAC (left) and CTCF M-ChIP (right) intensity at cytosines in a CTCF motif within M-ATAC peaks for the four groups of cytosines (group #1: 288 cytosines, group #2: 17133 cytosines, group #3 cytosines: 758, group #4: 25 cytosines). **d,** Violin plots of methylation percentage from CTCF M-ChIP and WGBS, at C2 and C12 positions into CTCF motif (MA0139.1). ***P=1.02e-12 for C2 CTCF M-ChIP versus C12 CTCF M-ChIP (Wilcoxon test), **P=0.008 for C2 WGBS versus C12 WGBS (Wilcoxon test), ***P=9e-12 for C2 CTCF M-ChIP versus C2 WGBS (Wilcoxon test, paired), ***P=0.00075 for C12 CTCF M-ChIP versus C12 WGBS (Wilcoxon test, paired), *P=0.023 for CTCF M-ChIP versus WGBS (linear regression model).

Supplementary Fig. 1| a, Pearson correlation of read counts comparing M-ATAC with unconverted samples (NC) and regular ATAC-seq (top), and CTCF M-ChIP with unconverted samples, a sample from the Schubeler lab generated using ChIP-BisSeq[12] (GSE39739) and regular CTCF ChIP-seq (bottom). **b**, Table showing number of peaks called for each sample, using MACS2. **c,** Jaccard indexes (Jaccard Index = (Intersection / (sample 1 + sample 2 – Intersection)) of peak intersections between ATAC, M-ATAC, M-ATAC-NC samples (left panel) and CTCF ChIP-seq, CTCF M-ChIP and CTCF M-ChIP-NC samples (right panel).

Supplementary Fig. 2| Read lengths for all ATAC, M-ATAC, M-ATAC unconverted (M-ATAC-NC), CTCF ChIP-seq, CTCF M-ChIP and CTCF M-ChIP unconverted (CTCF M-ChIP-NC) samples.

Supplementary Fig. 3| Average cytosine methylation from M-ATAC relative to the position of the cytosines in the peaks for cytosines with coverage of at least 5 reads.

Supplementary Fig. 4| Density plots of average methylation correlations for cytosines with coverage of at least 5 reads. Average cytosine methylation from **a,** M-ATAC replicate 1 versus replicate 2 in M-ATAC peaks (Pearson Correlation = 0.76, p-value < 2.2e-16). **b,** CTCF M-ChIP replicate 1 versus replicate 2 in CTCF M-ChIP peaks (Pearson Correlation = 0.84, p-value < 2.2e-16). **c,** CTCF ChIP-BisSeq (GSE39739) from Dirk Schubeler lab versus WGBS in CTCF ChIP-BisSeq peaks (Pearson Correlation = 0.83, p-value < 2.2e-16). **d,** Genomic annotations for the 4 groups from Fig. 1c. Promoter: TSS - 1kb and +100bp; intragenic: introns, exons, 5'UTR, 3'UTR and TTS, intergenic: distal from promoter >1kb and non-coding RNAs. **e,** Transcriptional output for the 4 groups from **Fig. 1b**, for the cytosines at promoters (left panel, see **Supplementary Fig 3d**). ***P=1.25e-28 between groups #1 and 2, ^{NS}P=0.19 between groups #2 and 3, ^{NS}P=0.58 between groups #3 and 4 (Wilcoxon test), and for the cytosines at

intragenic regions (right panel, introns, exons, 5'UTR, 3'UTR, see **Supplementary Fig. 3d**). ***P= 0.0001 between groups #1 and 2, *P= 0.02 between groups #2 and 3, ^{NS}P= 0.1 between groups #3 and 4 (Wilcoxon test)**.**

Supplementary Fig. 5| a, Cytosines in M-ATAC peaks were divided into three groups according to methylation status from WGBS: 1/ Low Methylation (<20%, 351561 cytosines), 2/ Intermediate methylation (20-80, 58655 cytosines), 3/ High methylation (>80%, 17385 cytosines). Of note, a cutoff of 5 reads coverage were applied, and as opposed to Fig. 2a, no additional division were made based on coverage. ***P <0.001 (Wilcoxon text). **b,** Transcriptional output for the 4 groups from **Fig. 2a**, for the cytosines at intragenic regions (introns, exons, 5'UTR, 3'UTR, see **Fig. 2b**). *P= 0.028 between groups #1 and 2, ***P= 1.38e-38 between groups #2 and 3, ^{NS}P= 0.88 between groups #3 and 4 (Wilcoxon test)**. c,** Heatmaps of M-ATAC, H3K4me1, H3K4me3 and H3K27ac signal corresponding to the average profiles shown in Fig. 2D for the 4 groups of cytosines from Fig. 2A at promoters (left panel) versus non-promoters (right panel).

Supplementary Fig. 6| a, Examples of TF motifs from HOMER. The *de novo* motif in group 4 is compared with known motifs for KLF4, NANOG and OCT4 from the HOMER database. **b,** Violin plots showing methylation percentage at KLF4, NANOG and OCT4 at the known and *de novo* motifs from M-ATAC. ***P= 1.856e-08 for KLF4, *P= 0.049 for NANOG, **P= 0.0017 for OCT4 (Wilcoxon test).

Supplementary Fig. 7| Comparison of CTCF motifs found in CTCF ChIP-seq and CTCF M-ChIP (this study and GSE39739) from HOMER analysis with the MA0139.1 motif from Jaspar database.

Supplementary Fig. 8| a, Tables and histogram representing the number of cytosines at position C2 and C12 in the CTCF motif MA0139.1 in CTCF M-ChIP peaks as well as the frequency of the observed versus expected co-occurrence of methylation at C2 and C12 (χ^2 = 1531, p- value < 0.001). **b,** Frequency of methylation in the CTCF motif from CTCF M-ChIP, for the 7 possible combinations of base variations associated with C at positions 2 (1st couple of nucleotides) and 12 (2^{d} couple of nucleotides).

Online Methods

Cell culture

Mouse embryonic stem cells were provided by Matthias Stadtfeld. Briefly, KH2 embryonic stem cells (ESCs) [20] were cultured on irradiated feeder cells in KO-DMEM (Invitrogen) supplemented with L-glutamine, penicillin/streptomycin, nonessential amino acids, β-mercaptoethanol, 1,000 U/mL LIF, and 15% FBS (ESC medium). To remove feeder cells from ESCs, cells were trypsin digested and pre-plated in ESC medium for 30 mins. Supernatant containing ESCs was used for further experiments.

Assembly of the transposase

Tn5 transposase was assembled with methylated adaptors as per the T-WGBS protocol[16] and described in detail in the Supplementary Note. Briefly, 10 µl of each adapter with incorporated methylated cytosines (Tn5mC-Apt1 and Tn5mC1.1-A1block; 100 µM each; **Supplementary Table 2**) was added to 80 µl of water and annealed in thermomixer with the following program: 95 °C for 3 min, 70 °C for 3 min, 45 cycles of 30 s with a ramp at −1 °C per cycle to reach 26 °C. 50 µL of annealed adapters was incubated with 50 µl of hot glycerol and 10 µl of this mixture was incubated with 10 µl of Ez-Tn5 transposase (from the EZ-Tn5 insertion kit) at room temperature for 30 min to assemble the transposome.

ATAC-seq and M-ATAC

ATAC-seq and M-ATAC were performed with 50 thousand mESCs as per the original ATAC-seq protocol[13] and as described in detail in the Supplementary Note. Cells were washed in cold

PBS and resuspended in 50 µl of cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3mM $MgCl_2$, 0.1 % IGEPAL CA-630). The tagmentation reaction was performed in 25 µl of TD buffer (Illumina Cat #FC-121-1030), 2.5 µl Transposase (either the Nextera transposase (ATAC-seq) or the transposase containing the methylated adaptors (M-ATAC, see section "assembly of the transposase" for details), and 22.5 µl of Nuclease Free H_2O at 37ºC for 30 min. Purified DNA (on column with the Qiagen Mini Elute kit) was bisulfite converted (see section "Bisulfite conversion for details).

CTCF ChIPseq and M-ChIP

CTCF ChIP-seq and M-ChIP were performed on mESC as per the original ChIPmentation protocol[14] as described in detail in the Supplementary Note. Briefly, 5 µl of CTCF antibody (Millipore) was combined to protein A magnetic beads and added to sonicated chromatin from 10 million mESC, for 3 to 6 hours rotating in cold room. During washes, tagmentation was performed for 1 min at 37°C with either 1 µl of the Nextera transposase (ChIP-seq) or the transposase containing the methylated adaptors (M-ChIP, see section "assembly of the transposase" for details). Chromatin was decrosslinked by adding proteinase K for 2 hours at 55 °C and overnight incubation at 65 °C. Eluted and purified DNA was bisulfite converted (see section "Bisulfite conversion for details).

Bisulfite conversion

Purified DNA was bisulfite converted following the T-WGBS protocol[16] with the EZ DNA methylation kit (Zymo) as described in detail in the Supplementary Note. Briefly, oligonucleotide replacement was performed by incubating 9 µl of tagmented M-ATAC or M-ChIP purified DNA

with 2 ng of phage lambda DNA as carrier, 2 µl of dNTP mix (2.5 mM each, 10 mM), 2 µl of 10× Ampligase buffer and 2 µl of replacement oligo (Tn5mC-ReplO1, 10 µM; Table 1) in a thermomixer with the following program: 50 °C for 1 min, 45°C for 10 min, ramp at −0.1 °C per second to reach 37 °C. 1 µl of T4 DNA polymerase and 2.5 µl of Ampligase were added and the gap repair reaction was performed at 37 °C for 30 min. DNA was purified using SPRI AMPure XP beads with a beads-to-sample ratio of 1.8:1 and eluted in 50 µl of H2O. 5 µl were kept as an unconverted control sample, and 45 µl was bisulfite converted using the EZ DNA methylation kit (Zymo). Briefly, the gap repair reaction was performed by adding 5 µl of M-dilution buffer and 15 min incubation at 37 °C, and bisulfite treatment was performed by adding 100 µl of liquid CT-conversion reagent in a thermomixer with the following program: 16 cycles of 95 °C for 15 sfollowed by 50 °C for 1 hour. Converted DNA was purified on a column and amplified (see section “Amplification of M-ATAC and M-ChIP libraries” for details).

Amplification of ATAC-seq and ChIP-seq libraries

Purified DNA (20 µl) was combined with 2.5 µl of each primer and 25 µl of NEB Next PCR master mix as per the original ATAC-seq protocol[13]. For ATAC-seq, DNA was amplified for 5 cycles and a monitored quantitative PCR was performed to determine the number of extra cycles needed, and DNA was purified on column with the Qiagen Mini Elute kit (see detail in the Supplementary Note). For ChIP-seq, DNA was amplified as per the ChIPmentation protocol[14] in a thermomixer with the following program: 72 °C for 5 min; 98 °C for 30 s; 14 cycles of 98 °C for 10 s, 63 °C for 30 s and 72 °C 30 s; and a final elongation at 72 °C for 1 min. DNA was purified using SPRI AMPure XP beads with a beads-to-sample ratio of 1:1 and eluted in 20 µl of H2O.

Amplification of M-ATAC and M-ChIP libraries

Purified converted DNA was amplified as per the original T-WGBS protocol [16]. Briefly, 10 µl of DNA was combined with 1.25 µl of each primer and 12.5 µl of high-fidelity system KAPA HiFi uracil+ PCR master mix. DNA was amplified for 5 cycles and a monitored quantitative PCR was performed to determine the number of extra cycles needed (see details in the Supplementary Note).

Sequencing of the libraries and data processing

For ATAC-seq, ChIP-seq, M-ATAC and M-ChIP, libraries were quantified using Kapa qPCR kit and sequenced using the HiSeq 2500 for paired-end 50 bp reads (aiming for 50 million paired reads per sample). ChIP-seq for histone modifications in mESC were downloaded from GEO (H3K4me1: GSM1000121, H3K27ac: GSM1000126, H3K4me3: GSM1000124). Data processing was performed as per the pipeline available on Github (link: "https://github.com/skoklab/EpiMethylTag"). Briefly, reads were trimmed using trim-galore/0.4.4, and aligned to the mm10 assembly of mouse genome using bowtie2[21] for ChIP-seq and ATAC-seq, and using Bismark/0.18.1 (bowtie2)[22] for M-ChIP and M-ATAC to account for bisulfite conversion. Reads with quality < 30 and duplicates were removed using Samtools/1.3[23]. Peaks were called using Macs/2.1.0[24] and narrow peaks were considered for further analysis. Bigwigs were generated from bam files with RPKM normalization using Deeptools[25] for visualization on IGV.

Bioinformatic analysis of data

The distribution of fragment lengths were assessed with Deeptools/2.3.3 with option "--maxFragmentLength 1000", and Pearson correlations of reads counts with Deeptools/2.3.3 and default parameters. Heatmaps and average profiles were performed on merged bigwig files using Deeptools/2.3.3. For Fig. 2d and Supplementary Fig. 5b, the plots were centered on cytosines into M-ATAC peaks from the different groups highlighted in Fig. 2a. For Fig. 3c, lists of cytosines were subsampled using BEDTools[26] to consider only the CpGs inside CTCF motifs, and the plots were centered on those CpGs. Genomic annotations were performed using HOMER[27]. CTCF motifs locations in CTCF M-ChIP/ChIP and M-ATAC were determined using the FIMO tool from MEME[28], with the CTCF motif PWM from Jaspar database (MA0139.1). PWM was manually modified to look at methylation frequency at different combinations of C2 and C12 dinucleotides. Transcription factors motifs were identified in M-ATAC peaks using HOMER (Hypergeometric Optimization of Motif EnRichment)[27] to assess for 1/ motif enrichments at the different groups of cytosines for Fig. 2a (considering a window of 10bp around each cytosine of each groups and merging if overlapping) (**Supplementary table 1**) and 2/ frequency of DNA methylation at KLF4, NANOG and OCT4 motifs (**Supplementary Fig. 6**). Scripts are available on Github (link: "https://github.com/skoklab/EpiMethylTag").

References

1 Dor, Y. & Cedar, H. Principles of DNA methylation and their implications for biology and medicine. *Lancet* **392**, 777-786, doi:10.1016/S0140-6736(18)31268-6 (2018).

2 Hu, S. *et al.* DNA methylation presents distinct binding sites for human transcription factors. *Elife* **2**, e00726, doi:10.7554/eLife.00726 (2013).

3 Maurano, M. T. *et al.* Role of DNA Methylation in Modulating Transcription Factor Occupancy. *Cell Rep* **12**, 1184-1195, doi:10.1016/j.celrep.2015.07.024 (2015).

4 Zhu, H., Wang, G. & Qian, J. Transcription factors as readers and effectors of DNA methylation. *Nat Rev Genet* **17**, 551-565, doi:10.1038/nrg.2016.83 (2016).

5 Fouse, S. D. *et al.* Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* **2**, 160-169, doi:10.1016/j.stem.2007.12.011 (2008).

6 Natarajan, A., Yardimci, G. G., Sheffield, N. C., Crawford, G. E. & Ohler, U. Predicting cell-type-specific gene expression from regions of open chromatin. *Genome Res* **22**, 1711-1722, doi:10.1101/gr.135129.111 (2012).

7 Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* **33**, 5868-5877, doi:10.1093/nar/gki901 (2005).

8 Kelly, T. K. *et al.* Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res* **22**, 2497-2506, doi:10.1101/gr.143008.112 (2012).

9 Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, doi:10.1126/science.aaj2239 (2017).

10 Brinkman, A. B. *et al.* Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res* **22**, 1128-1138, doi:10.1101/gr.133728.111 (2012).

401 11 Statham, A. L. *et al.* Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq)
402 directly informs methylation status of histone-modified DNA. *Genome Res* **22**, 1120-1127,
403 doi:10.1101/gr.132076.111 (2012).
404 12 Feldmann, A. *et al.* Transcription factor occupancy can mediate active turnover of DNA
405 methylation at regulatory regions. *PLoS Genet* **9**, e1003994, doi:10.1371/journal.pgen.1003994
406 (2013).
407 13 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native
408 chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins
409 and nucleosome position. *Nat Methods* **10**, 1213-1218, doi:10.1038/nmeth.2688 (2013).
410 14 Schmidl, C., Rendeiro, A. F., Sheffield, N. C. & Bock, C. ChIPmentation: fast, robust, low-input ChIP-
411 seq for histones and transcription factors. *Nat Methods* **12**, 963-965, doi:10.1038/nmeth.3542
412 (2015).
413 15 Adey, A. & Shendure, J. Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing.
414 *Genome Res* **22**, 1139-1143, doi:10.1101/gr.136242.111 (2012).
415 16 Wang, Q. *et al.* Tagmentation-based whole-genome bisulfite sequencing. *Nat Protoc* **8**, 2022-
416 2032, doi:10.1038/nprot.2013.118 (2013).
417 17 Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions.
418 *Nature* **480**, 490-495, doi:10.1038/nature10716 (2011).
419 18 Wang, H. *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res*
420 **22**, 1680-1688, doi:10.1101/gr.136101.111 (2012).
421 19 Hashimoto, H. *et al.* Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF
422 to DNA. *Mol Cell* **66**, 711-720 e713, doi:10.1016/j.molcel.2017.05.004 (2017).

423 20 Beard, C., Hochedlinger, K., Plath, K., Wutz, A. & Jaenisch, R. Efficient method to generate single-
424 copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* **44**, 23-28,
425 doi:10.1002/gene.20180 (2006).
426 21 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-
427 359, doi:10.1038/nmeth.1923 (2012).
428 22 Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq
429 applications. *Bioinformatics* **27**, 1571-1572, doi:10.1093/bioinformatics/btr167 (2011).
430 23 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079,
431 doi:10.1093/bioinformatics/btp352 (2009).
432 24 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-
433 2008-9-9-r137 (2008).
434 25 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis.
435 *Nucleic Acids Res* **44**, W160-165, doi:10.1093/nar/gkw257 (2016).
436 26 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features.
437 *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
438 27 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-
439 regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589,
440 doi:10.1016/j.molcel.2010.05.004 (2010).
441 28 Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**,
442 W202-208, doi:10.1093/nar/gkp335 (2009).

443

Figure 1

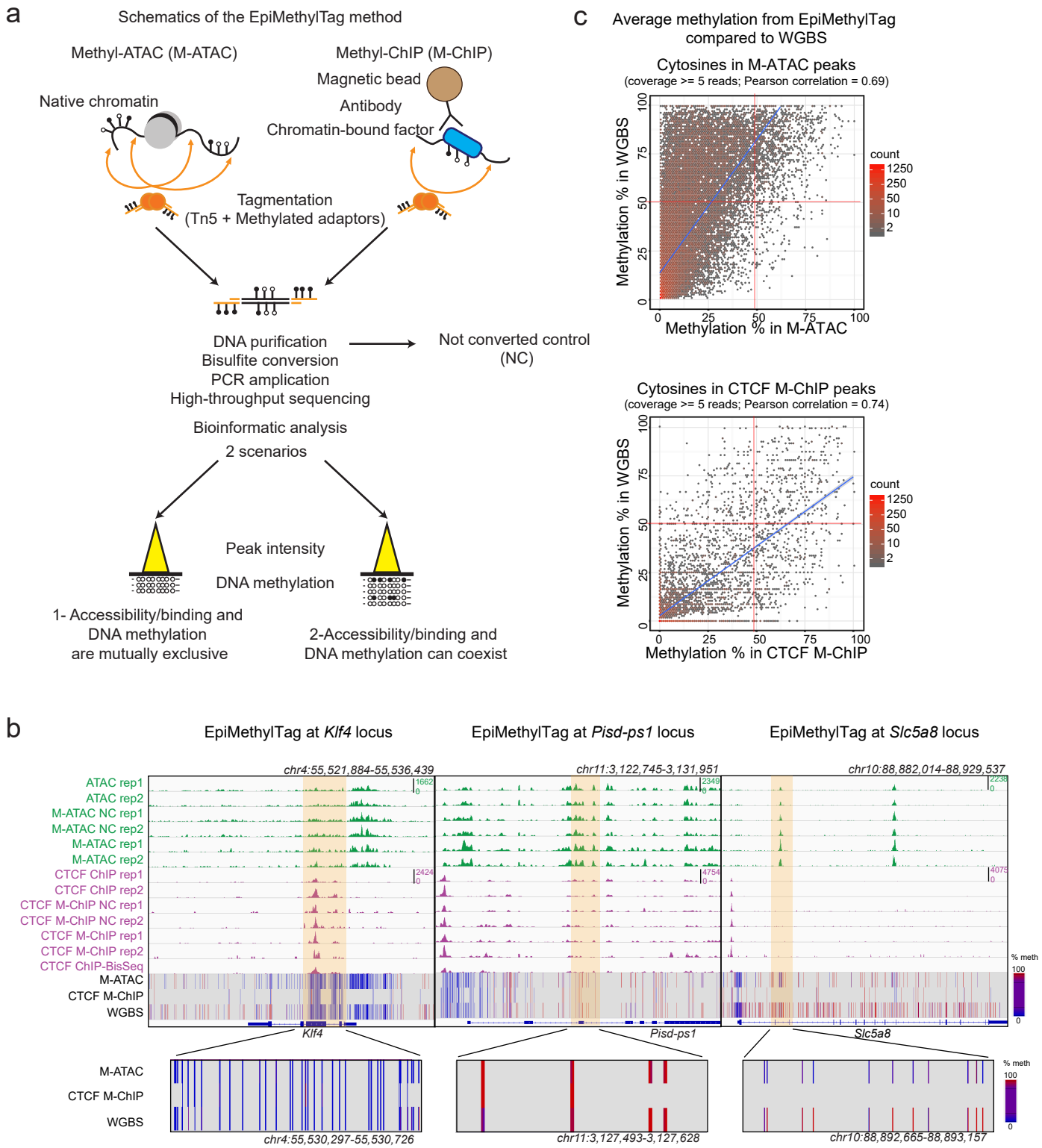

a
Schematics of the EpiMethylTag method
Methyl-ATAC (M-ATAC)
Methyl-ChIP (M-ChIP)
Magnetic bead
Antibody
Chromatin-bound factor
Native chromatin
Tagmentation
(Tn5 + Methylated adaptors)
DNA purification
Bisulfite conversion
PCR amplication
High-throughput sequencing
Not converted control
(NC)
Bioinformatic analysis
2 scenarios
Peak intensity
DNA methylation
1- Accessibility/binding and
DNA methylation
are mutually exclusive
2-Accessibility/binding and
DNA methylation can coexist
c
Average methylation from EpiMethylTag
compared to WGBS
Cytosines in M-ATAC peaks
(coverage >= 5 reads; Pearson correlation = 0.69)
Methylation % in WGBS
Methylation % in M-ATAC
count
1250
250
50
10
2
Cytosines in CTCF M-ChIP peaks
(coverage >= 5 reads; Pearson correlation = 0.74)
Methylation % in WGBS
Methylation % in CTCF M-ChIP
b
EpiMethylTag at Klf4 locus
EpiMethylTag at Pisd-ps1 locus
EpiMethylTag at Slc5a8 locus
chr4:55,521,884-55,536,439
chr11:3,122,745-3,131,951
chr10:88,882,014-88,929,537
ATAC rep1
ATAC rep2
M-ATAC NC rep1
M-ATAC NC rep2
M-ATAC rep1
M-ATAC rep2
CTCF ChIP rep1
CTCF ChIP rep2
CTCF M-ChIP NC rep1
CTCF M-ChIP NC rep2
CTCF M-ChIP rep1
CTCF M-ChIP rep2
CTCF ChIP-BisSeq
M-ATAC
CTCF M-ChIP
WGBS
% meth
Klf4
Pisd-ps1
Slc5a8
chr4:55,530,297-55,530,726
chr11:3,127,493-3,127,628
chr10:88,892,665-88,893,157

Figure 2

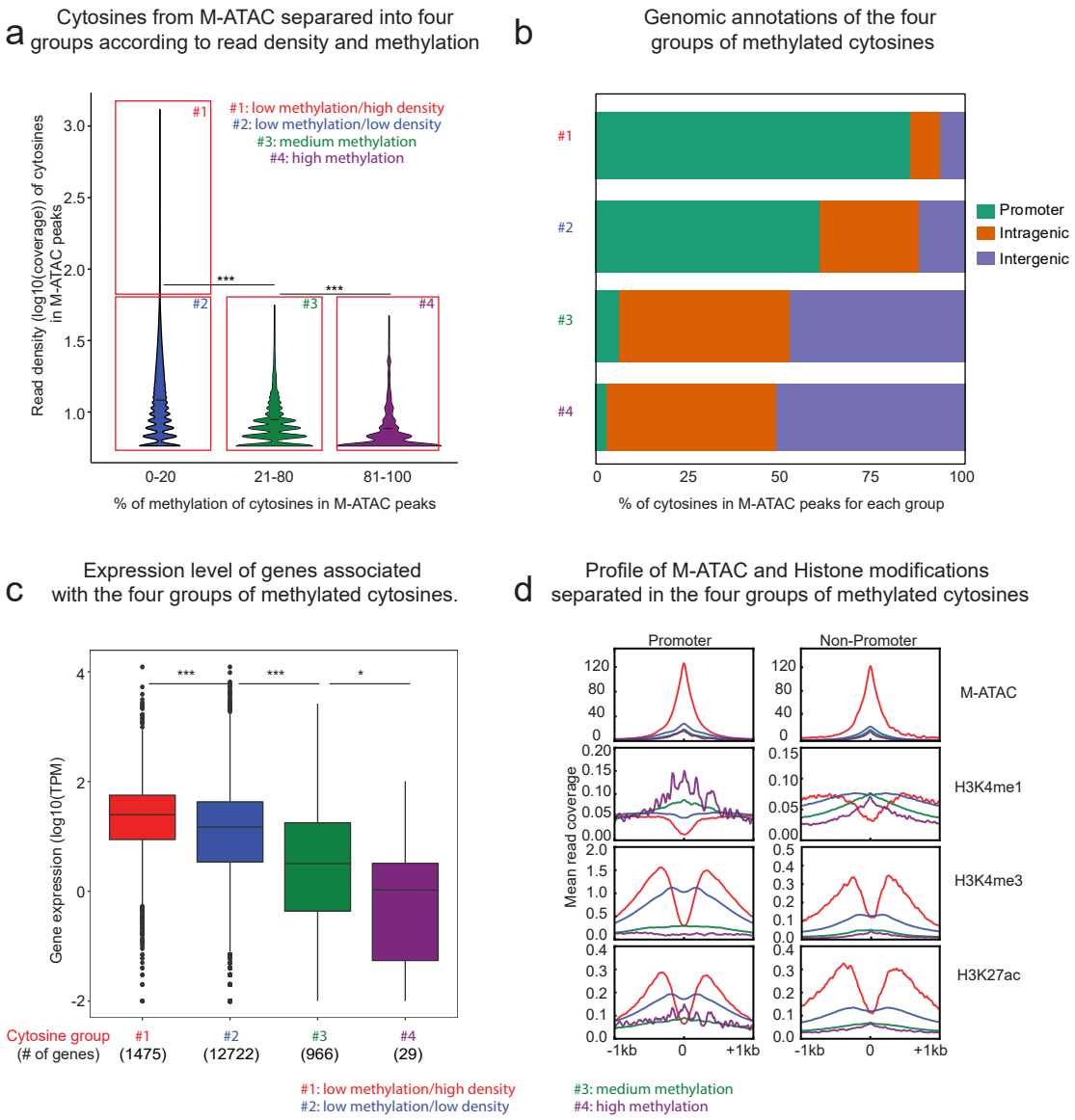
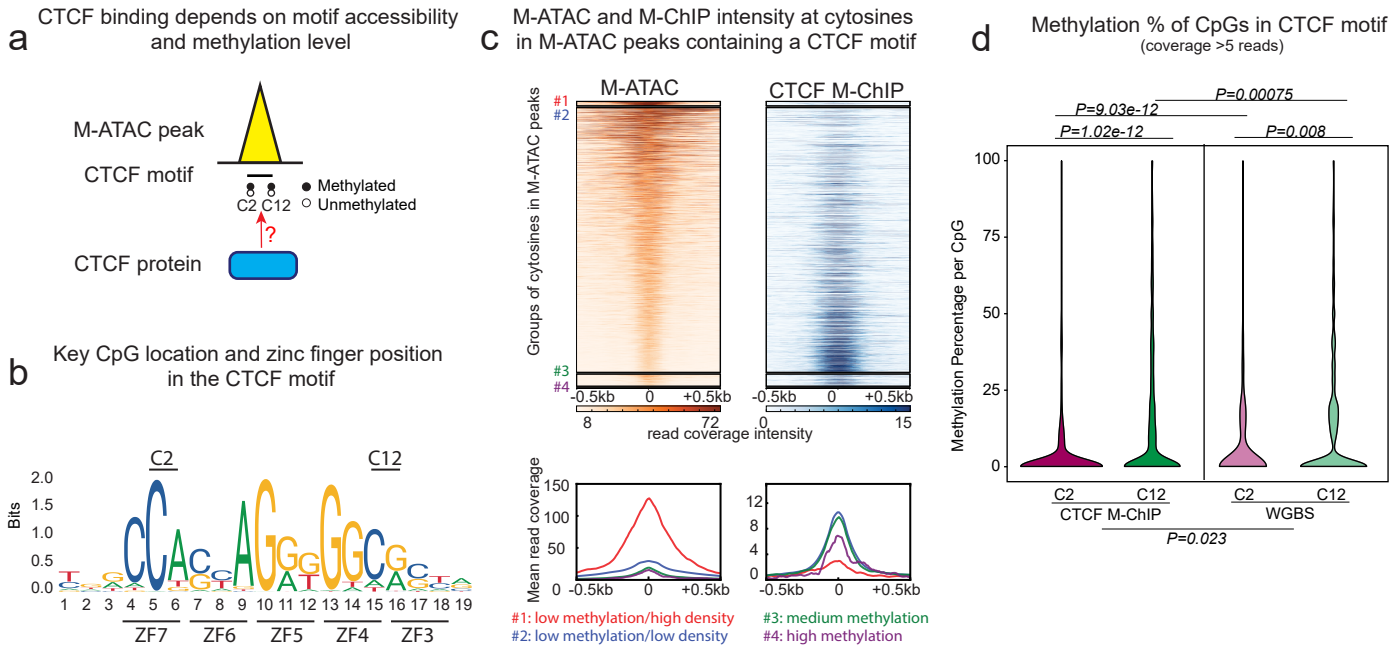

a
Cytosines from M-ATAC separared into four groups according to read density and methylation
#1: low methylation/high density
#2: low methylation/low density
#3: medium methylation
#4: high methylation
Read density (log10(coverage)) of cytosines in M-ATAC peaks
3.0
2.5
2.0
1.5
1.0
***
***
#1
#2
#3
#4
0-20
21-80
81-100
% of methylation of cytosines in M-ATAC peaks
b
Genomic annotations of the four groups of methylated cytosines
#1
#2
#3
#4
Promoter
Intragenic
Intergenic
0
25
50
75
100
% of cytosines in M-ATAC peaks for each group
c
Expression level of genes associated with the four groups of methylated cytosines.
Gene expression (log10(TPM))
4
2
0
-2
***
***
*
Cytosine group
(# of genes)
#1
(1475)
#2
(12722)
#3
(966)
#4
(29)
#1: low methylation/high density
#2: low methylation/low density
#3: medium methylation
#4: high methylation
d
Profile of M-ATAC and Histone modifications separated in the four groups of methylated cytosines
Promoter
Non-Promoter
M-ATAC
H3K4me1
H3K4me3
H3K27ac
Mean read coverage
-1kb
0
+1kb

Figure 3


a
CTCF binding depends on motif accessibility and methylation level
M-ATAC peak
CTCF motif
C2 C12
Methylated
Unmethylated
?
CTCF protein
b
Key CpG location and zinc finger position in the CTCF motif
C2
C12
Bits
2.0
1.5
1.0
0.5
0.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
ZF7
ZF6
ZF5
ZF4
ZF3
c
M-ATAC and M-ChIP intensity at cytosines in M-ATAC peaks containing a CTCF motif
M-ATAC
CTCF M-ChIP
Groups of cytosines in M-ATAC peaks
#1
#2
#3
#4
-0.5kb 0 +0.5kb
8 72
0 15
read coverage intensity
Mean read coverage
150
100
50
0
12
8
4
0
#1: low methylation/high density
#2: low methylation/low density
#3: medium methylation
#4: high methylation
d
Methylation % of CpGs in CTCF motif
(coverage >5 reads)
P=0.00075
P=9.03e-12
P=1.02e-12
P=0.008
Methylation Percentage per CpG
100
75
50
25
0
C2
C12
CTCF M-ChIP
WGBS
P=0.023