bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

8 [#] Corresponding author.

9 Correspondence: Jane.Skok@nyulangone.org

10 Key words: DNA methylation, chromatin accessibility, CTCF.

11

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

12 **Abstract**

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

19

20

21

22

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

23 **Main**

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

30

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

42

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

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

49 harboring cytosine methylation (**Supplementary table 1**). For M-ATAC or M-ChIP, tagmentation
50 occurs respectively on nuclear lysates as per the conventional ATAC-seq protocol[13], or during
51 chromatin immunoprecipitation as per the ChIPmentation protocol[14]. Following DNA purification,
52 the sample is bisulfite converted and PCR amplified for downstream sequencing (**Fig. 1a**). As
53 shown in **Fig. 1a**, EpiMethylTag can determine whether DNAme and accessibility/TF binding
54 are mutually exclusive (scenario 1) or can coexist in certain locations (scenario 2). The protocol
55 requires fewer cells, less sequencing depth, is quicker than existing methods and can be
56 analyzed using a pipeline we developed that is publicly available online on Github
57 (*"https://github.com/skoklab/EpiMethylTag"*).
58

59 M-ATAC and CTCF M-ChIP were performed in duplicate on murine embryonic stem cells
60 (mESC). As controls, we collected aliquots before bisulfite conversion, ATAC-seq and CTCF
61 ChIPmentation with Nextera transposase[13,14]. As shown in **Fig. 1b and Supplementary Fig. 1a**,
62 genome coverage was highly reproducible between M-ATAC replicates and highly correlated
63 with regular ATAC-seq and M-ATAC signal before bisulfite treatment. Thus, bisulfite treatment,
64 or the use of a different transposase does not result in signal bias. High reproducibility was also
65 seen for CTCF M-ChIP, and we observed consistency between our results and data generated
66 by CTCF ChIP-BisSeq, a similar technique that requires a much larger number of cells[12] (**Fig.
67 1b and Supplementary Fig. 1a)**). Of note, bisulfite conversion does not affect the number of
68 peaks detected, or the Jaccard index of peak overlap (**Supplementary Fig. 1b**), although it
69 leads to shorter reads (**Supplementary Fig. 2**). Of note, average methylation was higher at the
70 edges of the peaks than at the midpoint (**Supplementary Fig. 3**). Comparable DNA methylation
71 levels were found in M-ATAC and CTCF M-ChIP replicates, Pearson correlation = 0.76 and
72 0.84, respectively (**Supplementary Fig. 4a and 4b**) and these were pooled for subsequent
73 DNAme analysis.

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

74 We focused our analysis only at cytosines within peak regions covered by at least five reads, as
75 methylation outside of M-ATAC and M-ChIP peaks has low coverage and is less reliable. We
76 observe positive correlations between DNA methylation from WGBS and M-ATAC (**Fig. 1c,** top
77 panel, Pearson correlation=0.69), and between methylation levels in M-ChIP and WGBS (**Fig.**
78 **1c**, bottom panel, Pearson correlation = 0.74). Similar results were observed with the previously
79 published CTCF ChIP-BisSeq method[12] (GSE39739) (Pearson correlation = 0.83,
80 **Supplementary Fig. 4c**). In **Fig. 1b** we highlight the *Klf4* gene, which harbors a peak of
81 chromatin accessibility in the promoter and CTCF binding in the intragenic region associated
82 with low methylation from both EpiMethylTag and WGBS assays (left panel). In contrast, the
83 *Pisd-ps1* intragenic locus contains accessible chromatin that coexists with high levels of DNA
84 methylation as detected by both M-ATAC and WGBS (**Fig. 1b**, middle panel). Interestingly, a
85 proportion of M-ATAC peaks exhibited an intermediate-to-high average methylation level in
86 deeply sequenced WGBS[17], but low methylation in M-ATAC (**Fig. 1c**, top panel, top left corner)
87 as illustrated at the *Slc5a8* locus (**Fig. 1b**, right panel). These data suggest that, as expected
88 open regions are less methylated than closed regions within a population of cells, but that
89 accessibility and methylation can coexist in a small subset of genomic regions, which are
90 depleted for promoter regions and associated with low transcription (**Supplementary Fig. 4d**
91 **and 4e**). Importantly, M-ATAC is able to identify methylation levels within ATAC peaks,
92 information that cannot be retrieved from integrating data from separate WGBS and ATAC-seq
93 experiments.

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

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

124 and 3) compared to cytosines with low and high levels of methylation (groups 1 and 4) (**Fig.**
125 **3c**). These data provide insight into CTCF binding and indicate an anticorrelation between high
126 accessibility and high methylation, consistent with highly-significant CTCF motif enrichment at
127 cytosines with low levels of methylation (groups 2 and 3) (**Supplementary Table 1**). Consistent
128 with the findings from a recent study that analyzed CTCF binding using oligos rather than
129 genomic DNA methylated at positions C2 and C12[19], CTCF M-ChIP detected higher levels of
130 methylation at C12 compared to C2 (**Fig. 3d**, compare CTCF M-ChIP C2 versus C12, p-value =
131 1.02e-12) Importantly, CTCF M-ChIP is more suitable than WGBS for detecting the differences
132 (**Fig. 3d**, compare CTCF M-ChIP versus WGBS, p-value = 0.023).. In addition, we found that bi-
133 methylation at both CpGs is slightly enriched compared to what is expected by random chance
134 (0.97% versus 0.05%) (**Supplementary Fig. 8a,** χ^2 = 1531, p-value < 0.001). Nonetheless,
135 sequence variation at the C2 and C12 positions appears to have no effect on methylation
136 (**Supplementary Fig. 8b**).

137 In conclusion, we developed a method, “EpiMethylTag”, that allows the simultaneous
138 analysis of DNA methylation with ChIP-seq or ATAC-seq. “EpiMethylTag” can be used to
139 analyze the methylation status and coincident accessibility or binding of other chromatin bound
140 transcription factors. Using this technique, we confirmed that as a general rule, DNA
141 methylation rarely coexists with DNA accessibility or CTCF binding. However, in contrast to
142 WGBS, M-ATAC and CTCF M-ChIP revealed a complex interplay between accessible
143 chromatin, DNA methylation and CTCF binding. Thus, EpiMethylTag can be used to provide
144 information about the DNA sequence and chromatin context of TF binding at methylated sites
145 and its significance to gene regulation and biological processes. This technique can also be
146 adapted for single cell analysis.
147
148

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

149 **Methods**

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

151 XXX.

152

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

153 **Figure Legends**

154 **Fig. 1| EpiMethylTag is a reproducible method to test whether DNAme can coexist or not**
155 **with TF binding (CTCF) or chromatin accessibility genome-wide. a,** Schematic overview of
156 the EpiMethyTag method showing two possible outcomes. **b,** Representative IGV screenshots
157 of EpiMethylTag, at the *Klf4* locus (left panel), the *Pisd-ps1* locus (middle panel), and the *Slc5a8*
158 locus (right panel). ATAC and M-ATAC in green, CTCF in purple and DNA methylation from
159 merged M-ATAC, merged CTCF M-ChIP and WGBS (methylation from 0% in blue to 100% in
160 red). A zoom-in of methylation at the highlighted region is shown at the bottom of each example.
161 The *Klf4* locus illustrates a region that has low methylation as detected by M-ATAC, CTCF M-
162 ChIP and WGBS. *Pisd-ps1* locus illustrates a region that has high methylation as detected by
163 M-ATAC, CTCF M-ChIP and WGBS. *Slc5a8* locus illustrates a region that has low methylation
164 as detected by M-ATAC and high methylation as detected by WGBS. **c,** Density plots of
165 methylation from EpiMethyltag compared with WGBS. Only cytosines inside peaks and with at
166 least 5 reads were considered. Top: average methylation of cytosines per M-ATAC peak in M-
167 ATAC versus WGBS (Pearson Correlation = 0.69, p-value < 2.2e-16; bottom left corner: 27977
168 peaks, top left corner: 8408 peaks, top right corner: 1019 peaks, bottom right corner: 113
169 peaks). Bottom: average methylation per CTCF M-ChIP peak of cytosines in CTCF M-ChIP
170 versus WGBS (Pearson Correlation = 0.74, p-value < 2.2e-16; bottom left corner: 6549 peaks,
171 top left corner: 198 peaks, top right corner: 304 peaks, bottom right corner: 310 peaks).
172

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

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

178 (39321 cytosines). 4. High methylation (>80%) + Low coverage (5 to 50 reads) (1652
179 cytosines). *** P=0 between groups #1 + 2 and group #3, ***P=3.25e-109 between groups #3
180 and 4 (Wilcoxon text). **b,** Genomic annotations for the 4 groups from **Fig. 2a**. Promoter: TSS -
181 1kb and +100bp; intragenic: introns, exons, 5'UTR, 3'UTR and TTS, intergenic: distal from
182 promoter >1kb and non-coding RNAs. **c,** Expression level of genes associated with the four
183 groups of methylated cytosines from in **Fig. 2a**, for the cytosines at promoters. ****P*=4.2e-33
184 between groups #1 and 2, ****P*=2.8e-75 between groups #2 and 3, **P*=0.034 between groups #3
185 and 4 (Wilcoxon test). **d,** Average profile of M-ATAC, H3K4me1, H3K4me3 and H3K27ac signal
186 associated with the four groups of methylated cytosines from **Fig 2a** at promoters versus non-
187 promoters. Of note, the small number of promoters in group 4 gives an unsmooth pattern for
188 marks such as H3K4me1 and H3K27ac.
189
190 **Fig. 3| CTCF M-ChIP enables analysis of DNA methylation of distinct cytosines in the**
191 **CTCF motif. a,** Schematic illustration representing an ATAC-seq peak with a CTCF motif and
192 CTCF occupancy dependent on C2 and C12 methylation. **b,** CTCF motif from JASPAR
193 database (MA0139.1). The 2 key CpG positions (C2 and C12) are indicated. **c,** Heatmaps (top)
194 and average profiles (bottom) of M-ATAC (left) and CTCF M-ChIP (right) intensity at cytosines
195 in a CTCF motif within M-ATAC peaks for the four groups of cytosines (group #1: 288 cytosines,
196 group #2: 17133 cytosines, group #3 cytosines: 758, group #4: 25 cytosines). **d,** Violin plots of
197 methylation percentage from CTCF M-ChIP and WGBS, at C2 and C12 positions into CTCF
198 motif (MA0139.1). ****P*=1.02e-12 for C2 CTCF M-ChIP versus C12 CTCF M-ChIP (Wilcoxon
199 test), ***P*=0.008 for C2 WGBS versus C12 WGBS (Wilcoxon test), ****P*=9e-12 for C2 CTCF M-
200 ChIP versus C2 WGBS (Wilcoxon test, paired), ****P*=0.00075 for C12 CTCF M-ChIP versus
201 C12 WGBS (Wilcoxon test, paired), **P*=0.023 for CTCF M-ChIP versus WGBS (linear
202 regression model).

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

203 **Supplementary Fig. 1| a,** Pearson correlation of read counts comparing M-ATAC with
204 unconverted samples (NC) and regular ATAC-seq (top), and CTCF M-ChIP with unconverted
205 samples, a sample from the Schubeler lab generated using ChIP-BisSeq[12] (GSE39739) and
206 regular CTCF ChIP-seq (bottom). **b**, Table showing number of peaks called for each sample,
207 using MACS2. **c,** Jaccard indexes (Jaccard Index = (Intersection / (sample 1 + sample 2 –
208 Intersection)) of peak intersections between ATAC, M-ATAC, M-ATAC-NC samples (left panel)
209 and CTCF ChIP-seq, CTCF M-ChIP and CTCF M-ChIP-NC samples (right panel).
210
211 **Supplementary Fig. 2|** Read lengths for all ATAC, M-ATAC, M-ATAC unconverted (M-ATAC-
212 NC), CTCF ChIP-seq, CTCF M-ChIP and CTCF M-ChIP unconverted (CTCF M-ChIP-NC)
213 samples.
214
215 **Supplementary Fig. 3|** Average cytosine methylation from M-ATAC relative to the position of
216 the cytosines in the peaks for cytosines with coverage of at least 5 reads.
217
218 **Supplementary Fig. 4|** Density plots of average methylation correlations for cytosines with
219 coverage of at least 5 reads. Average cytosine methylation from **a,** M-ATAC replicate 1 versus
220 replicate 2 in M-ATAC peaks (Pearson Correlation = 0.76, p-value < 2.2e-16). **b,** CTCF M-ChIP
221 replicate 1 versus replicate 2 in CTCF M-ChIP peaks (Pearson Correlation = 0.84, p-value <
222 2.2e-16). **c,** CTCF ChIP-BisSeq (GSE39739) from Dirk Schubeler lab versus WGBS in CTCF
223 ChIP-BisSeq peaks (Pearson Correlation = 0.83, p-value < 2.2e-16). **d,** Genomic annotations
224 for the 4 groups from Fig. 1c. Promoter: TSS - 1kb and +100bp; intragenic: introns, exons,
225 5'UTR, 3'UTR and TTS, intergenic: distal from promoter >1kb and non-coding RNAs. **e,**
226 Transcriptional output for the 4 groups from **Fig. 1b**, for the cytosines at promoters (left panel,
227 see **Supplementary Fig 3d**). ****P*=1.25e-28 between groups #1 and 2, ^{NS}P=0.19 between
228 groups #2 and 3, ^{NS}P=0.58 between groups #3 and 4 (Wilcoxon test), and for the cytosines at

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

229 intragenic regions (right panel, introns, exons, 5'UTR, 3'UTR, see **Supplementary Fig. 3d**).
230 ****P*= 0.0001 between groups #1 and 2, **P*= 0.02 between groups #2 and 3, ^{NS}*P*= 0.1 between
231 groups #3 and 4 (Wilcoxon test)**.**
232

233 **Supplementary Fig. 5| a,** Cytosines in M-ATAC peaks were divided into three groups
234 according to methylation status from WGBS: 1/ Low Methylation (<20%, 351561 cytosines), 2/
235 Intermediate methylation (20-80, 58655 cytosines), 3/ High methylation (>80%, 17385
236 cytosines). Of note, a cutoff of 5 reads coverage were applied, and as opposed to Fig. 2a, no
237 additional division were made based on coverage. ***P <0.001 (Wilcoxon text). **b,**
238 Transcriptional output for the 4 groups from **Fig. 2a**, for the cytosines at intragenic regions
239 (introns, exons, 5'UTR, 3'UTR, see **Fig. 2b**). **P*= 0.028 between groups #1 and 2, ****P*= 1.38e-
240 38 between groups #2 and 3, ^{NS}*P*= 0.88 between groups #3 and 4 (Wilcoxon test)**. c,** Heatmaps
241 of M-ATAC, H3K4me1, H3K4me3 and H3K27ac signal corresponding to the average profiles
242 shown in Fig. 2D for the 4 groups of cytosines from Fig. 2A at promoters (left panel) versus non-
243 promoters (right panel).

244

245 **Supplementary Fig. 6| a,** Examples of TF motifs from HOMER. The *de novo* motif in group 4 is
246 compared with known motifs for KLF4, NANOG and OCT4 from the HOMER database. **b,** Violin
247 plots showing methylation percentage at KLF4, NANOG and OCT4 at the known and *de novo*
248 motifs from M-ATAC. ****P*= 1.856e-08 for KLF4, **P*= 0.049 for NANOG, ***P*= 0.0017 for OCT4
249 (Wilcoxon test).
250

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

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

254

255 **Supplementary Fig. 8| a,** Tables and histogram representing the number of cytosines at
256 position C2 and C12 in the CTCF motif MA0139.1 in CTCF M-ChIP peaks as well as the
257 frequency of the observed versus expected co-occurrence of methylation at C2 and C12 ($\chi^2 =$
258 1531, p- value < 0.001). **b,** Frequency of methylation in the CTCF motif from CTCF M-ChIP, for
259 the 7 possible combinations of base variations associated with C at positions 2 (1st couple of
260 nucleotides) and 12 (2d couple of nucleotides).

261

262

263

264

265

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

266 **Online Methods**

267 **Cell culture**

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

274

275 **Assembly of the transposase**

276 Tn5 transposase was assembled with methylated adaptors as per the T-WGBS protocol[16] and
277 described in detail in the Supplementary Note. Briefly, 10 µl of each adapter with incorporated
278 methylated cytosines (Tn5mC-Apt1 and Tn5mC1.1-A1block; 100 µM each; **Supplementary**
279 **Table 2**) was added to 80 µl of water and annealed in thermomixer with the following program:
280 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.
281 50 µL of annealed adapters was incubated with 50 µl of hot glycerol and 10 µl of this mixture
282 was incubated with 10 µl of Ez-Tn5 transposase (from the EZ-Tn5 insertion kit) at room
283 temperature for 30 min to assemble the transposome.

284

285 **ATAC-seq and M-ATAC**

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

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

288 PBS and resuspended in 50 µl of cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3mM
289 $MgCl_2$, 0.1 % IGEPAL CA-630). The tagmentation reaction was performed in 25 µl of TD buffer
290 (Illumina Cat #FC-121-1030), 2.5 µl Transposase (either the Nextera transposase (ATAC-seq)
291 or the transposase containing the methylated adaptors (M-ATAC, see section "assembly of the
292 transposase" for details), and 22.5 µl of Nuclease Free H_2O at 37ºC for 30 min. Purified DNA
293 (on column with the Qiagen Mini Elute kit) was bisulfite converted (see section "Bisulfite
294 conversion for details).

295

296 **CTCF ChIPseq and M-ChIP**

297 CTCF ChIP-seq and M-ChIP were performed on mESC as per the original ChIPmentation
298 protocol[14] as described in detail in the Supplementary Note. Briefly, 5 µl of CTCF antibody
299 (Millipore) was combined to protein A magnetic beads and added to sonicated chromatin from
300 10 million mESC, for 3 to 6 hours rotating in cold room. During washes, tagmentation was
301 performed for 1 min at 37°C with either 1 µl of the Nextera transposase (ChIP-seq) or the
302 transposase containing the methylated adaptors (M-ChIP, see section "assembly of the
303 transposase" for details). Chromatin was decrosslinked by adding proteinase K for 2 hours at 55
304 °C and overnight incubation at 65 °C. Eluted and purified DNA was bisulfite converted (see
305 section "Bisulfite conversion for details).

306

307 **Bisulfite conversion**

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

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

311 with 2 ng of phage lambda DNA as carrier, 2 µl of dNTP mix (2.5 mM each, 10 mM), 2 µl of 10×
312 Ampligase buffer and 2 µl of replacement oligo (Tn5mC-ReplO1, 10 µM; Table 1) in a
313 thermomixer with the following program: 50 °C for 1 min, 45°C for 10 min, ramp at −0.1 °C per
314 second to reach 37 °C. 1 µl of T4 DNA polymerase and 2.5 µl of Ampligase were added and the
315 gap repair reaction was performed at 37 °C for 30 min. DNA was purified using SPRI AMPure
316 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
317 unconverted control sample, and 45 µl was bisulfite converted using the EZ DNA methylation kit
318 (Zymo). Briefly, the gap repair reaction was performed by adding 5 µl of M-dilution buffer and 15
319 min incubation at 37 °C, and bisulfite treatment was performed by adding 100 µl of liquid CT-
320 conversion reagent in a thermomixer with the following program: 16 cycles of 95 °C for 15
321 sfollowed by 50 °C for 1 hour. Converted DNA was purified on a column and amplified (see
322 section "Amplification of M-ATAC and M-ChIP libraries" for details).

323

324 **Amplification of ATAC-seq and ChIP-seq libraries**

325 Purified DNA (20 µl) was combined with 2.5 µl of each primer and 25 µl of NEB Next PCR
326 master mix as per the original ATAC-seq protocol[13]. For ATAC-seq, DNA was amplified for 5
327 cycles and a monitored quantitative PCR was performed to determine the number of extra
328 cycles needed, and DNA was purified on column with the Qiagen Mini Elute kit (see detail in the
329 Supplementary Note). For ChIP-seq, DNA was amplified as per the ChIPmentation protocol[14] in
330 a thermomixer with the following program: 72 °C for 5 min; 98 °C for 30 s; 14 cycles of 98 °C for
331 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
332 using SPRI AMPure XP beads with a beads-to-sample ratio of 1:1 and eluted in 20 µl of H2O.

333

334

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

335 **Amplification of M-ATAC and M-ChIP libraries**

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

341

342 **Sequencing of the libraries and data processing**

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

355

356

357

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

358 **Bioinformatic analysis of data**

359 The distribution of fragment lengths were assessed with Deeptools/2.3.3 with option "--
360 maxFragmentLength 1000", and Pearson correlations of reads counts with Deeptools/2.3.3 and
361 default parameters. Heatmaps and average profiles were performed on merged bigwig files
362 using Deeptools/2.3.3. For Fig. 2d and Supplementary Fig. 5b, the plots were centered on
363 cytosines into M-ATAC peaks from the different groups highlighted in Fig. 2a. For Fig. 3c, lists of
364 cytosines were subsampled using BEDTools[26] to consider only the CpGs inside CTCF motifs,
365 and the plots were centered on those CpGs. Genomic annotations were performed using
366 HOMER[27]. CTCF motifs locations in CTCF M-ChIP/ChIP and M-ATAC were determined using
367 the FIMO tool from MEME[28], with the CTCF motif PWM from Jaspar database (MA0139.1).
368 PWM was manually modified to look at methylation frequency at different combinations of C2
369 and C12 dinucleotides. Transcription factors motifs were identified in M-ATAC peaks using
370 HOMER (Hypergeometric Optimization of Motif EnRichment)[27] to assess for 1/ motif
371 enrichments at the different groups of cytosines for Fig. 2a (considering a window of 10bp
372 around each cytosine of each groups and merging if overlapping) (**Supplementary table 1**) and
373 2/ frequency of DNA methylation at KLF4, NANOG and OCT4 motifs (**Supplementary Fig. 6**).
374 Scripts are available on Github (link: "https://github.com/skoklab/EpiMethylTag").

375

376

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

377 ## References

378 1 Dor, Y. & Cedar, H. Principles of DNA methylation and their implications for biology and medicine.
379 *Lancet* **392**, 777-786, doi:10.1016/S0140-6736(18)31268-6 (2018).
380 2 Hu, S. *et al.* DNA methylation presents distinct binding sites for human transcription factors. *Elife*
381 **2**, e00726, doi:10.7554/eLife.00726 (2013).
382 3 Maurano, M. T. *et al.* Role of DNA Methylation in Modulating Transcription Factor Occupancy. *Cell*
383 *Rep* **12**, 1184-1195, doi:10.1016/j.celrep.2015.07.024 (2015).
384 4 Zhu, H., Wang, G. & Qian, J. Transcription factors as readers and effectors of DNA methylation.
385 *Nat Rev Genet* **17**, 551-565, doi:10.1038/nrg.2016.83 (2016).
386 5 Fouse, S. D. *et al.* Promoter CpG methylation contributes to ES cell gene regulation in parallel with
387 Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* **2**, 160-169,
388 doi:10.1016/j.stem.2007.12.011 (2008).
389 6 Natarajan, A., Yardimci, G. G., Sheffield, N. C., Crawford, G. E. & Ohler, U. Predicting cell-type-
390 specific gene expression from regions of open chromatin. *Genome Res* **22**, 1711-1722,
391 doi:10.1101/gr.135129.111 (2012).
392 7 Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution
393 DNA methylation analysis. *Nucleic Acids Res* **33**, 5868-5877, doi:10.1093/nar/gki901 (2005).
394 8 Kelly, T. K. *et al.* Genome-wide mapping of nucleosome positioning and DNA methylation within
395 individual DNA molecules. *Genome Res* **22**, 2497-2506, doi:10.1101/gr.143008.112 (2012).
396 9 Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription
397 factors. *Science* **356**, doi:10.1126/science.aaj2239 (2017).
398 10 Brinkman, A. B. *et al.* Sequential ChIP-bisulfite sequencing enables direct genome-scale
399 investigation of chromatin and DNA methylation cross-talk. *Genome Res* **22**, 1128-1138,
400 doi:10.1101/gr.133728.111 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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).

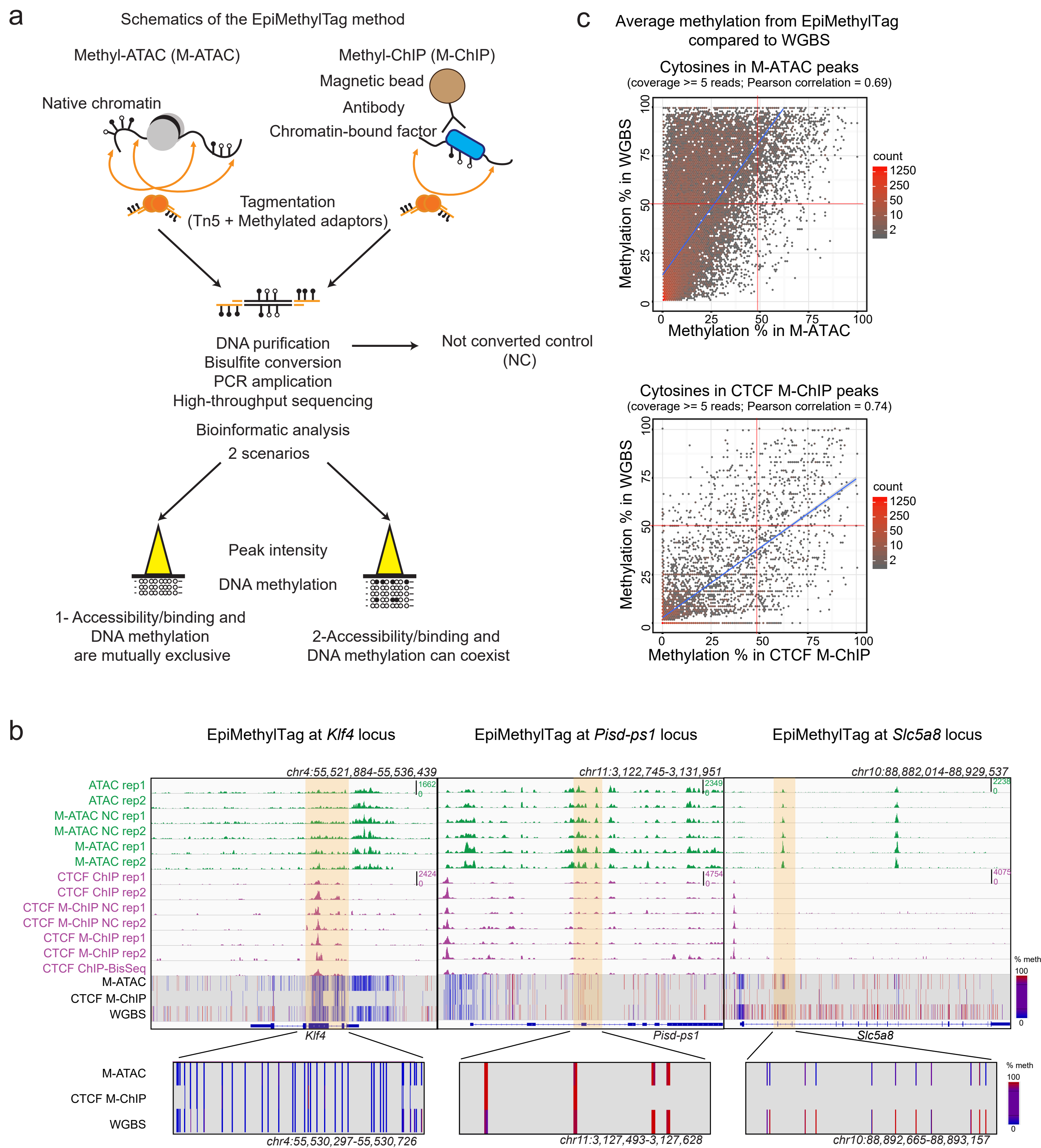
bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

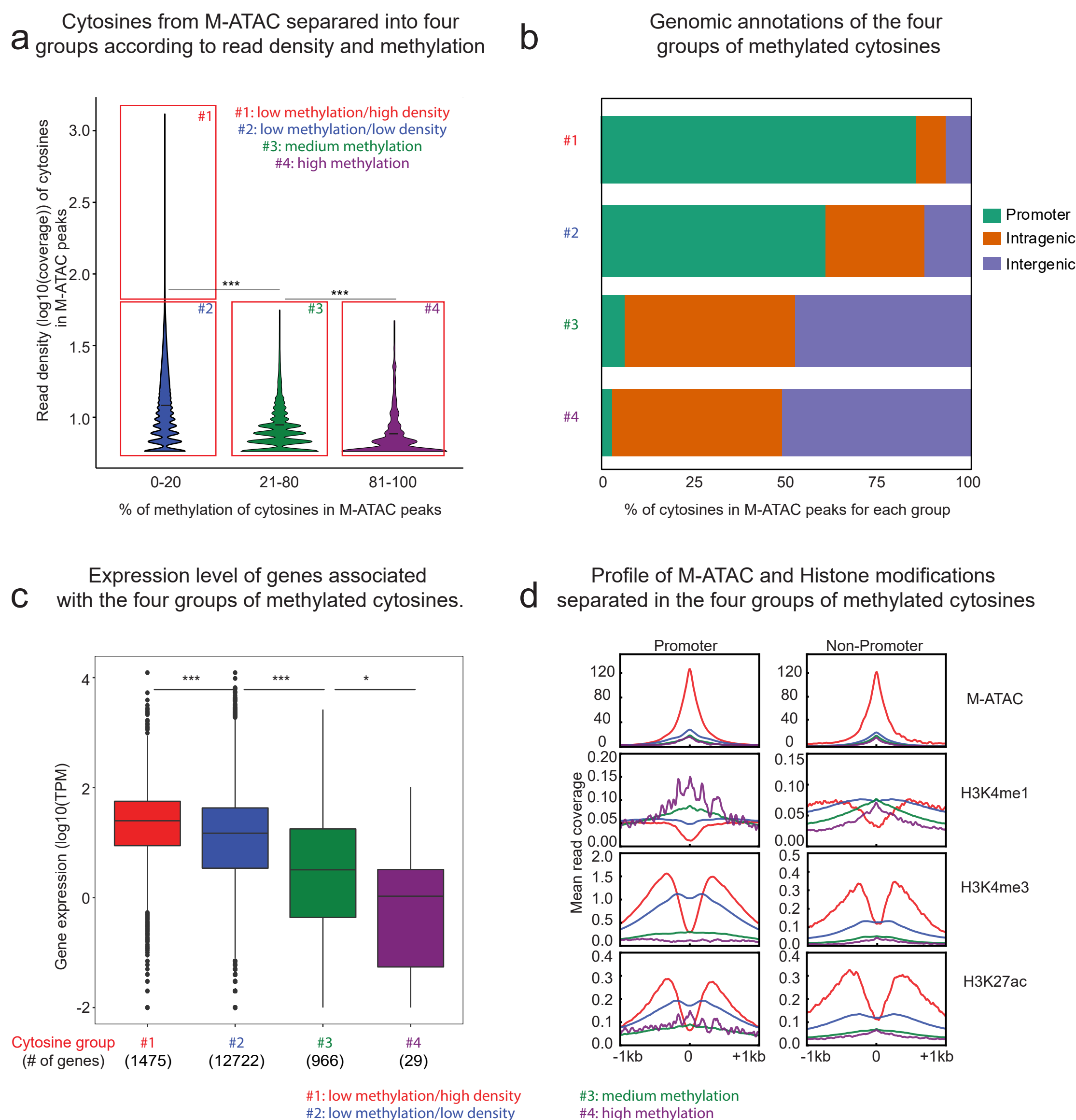
bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1



bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2



bioRxiv preprint doi: https://doi.org/10.1101/549550; this version posted February 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3