**Supplemental information for: Bacterial Infection Remodels the DNA Methylation Landscape of Human Dendritic Cells**

Alain Pacis1,2, Ludovic Tailleux3, Alexander M Morin4, John Lambourne5, Julia L Maclsaac4, Vania Yotova1, Anne Dumaine1, Anne Danckaert6, Francesca Luca7, Jean-Christophe Grenier1, Kasper D Hansen8, Brigitte Gicquel3, Miao Yu9, Athma Pai10, Chuan He9, Jenny Tung11,Tomi Pastinen5, Michael S. Kobor4, Roger Pique-Regi7, Yoav Gilad12\*, Luis B Barreiro1,13,\*

\*Correspondence to: [luis.barreiro@umontreal.ca](mailto:luis.barreiro@umontreal.ca) or to [gilad@uchicago.edu](mailto:gilad@uchicago.edu)

**Contents**

Supplemental Methods 3

Sample collection3

*Mycobacterium tuberculosis* preparation and infection of DCs3

CFSE proliferation assay 3

DNA and RNA Extractions 4

MethylC-seq library preparation and sequencing 4

Tet-assisted bisulfite sequencing (TAB-seq)4

5hmC staining 5

Western blotting 5

Quantitative real-time PCR 5

Bisulfite pyrosequencing 6

RNA-seq library preparation and sequencing 6

ChIP-seq library preparation and sequencing 6

ATAC-seq library preparation and sequencing 7

MethylC-seq data processing 7

TAB-seq data processing 7

ChIP-seq data processing 8

RNA-seq data processing and identification of DE genes upon MTB infection 8

Genomic annotation and mRNA TSS collection 8

Integrated analysis of gene expression and 5mC 9

Principal component analysis based on methylomes 9

Identification of MTB-DMRs 9

Assigning MTB-DMRs to genes 10

5hmC analysis 10

Chromatin state annotation and dynamics 10

Enhancer classification of hypomethylated regions based on chromatin state 11

ChIP-seq profiles around MTB-DMRs 11

ATAC-seq data processing and footprinting analysis 11

Peak detection 12

Gene Set Enrichment Analysis 12

Data visualization in the Immune Epigenome Browser 12

Supplemental Figures 14

Supplemental References 31

**Supplemental Methods**

In this section we repeat some of the text from the Method section of the main paper in order to have a comprehensive and uninterrupted description of the study design and statistical analysis.

**Sample collection.** Blood samples were obtained from the Indiana Blood Center. A signed written consent was obtained from all of the participants and the project was approved by the ethics committee at the CHU Sainte-Justine (protocol #4023). All individuals recruited in this study were healthy males of European descent between the ages of 21 and 55 years old. We decided to only focus on males to limit variation in DNA methylation levels due to sex-specific differences.

***Mycobacterium tuberculosis* preparation and infection of DCs.** Peripheral blood mononuclear cells (PBMCs) were isolated from buffy coats by Ficoll-Paque centrifugation. Blood monocytes were then purified from PBMCs by positive selection with magnetic CD14 MicroBeads (Miltenyi Biotech). Pure monocytes were cultured for 5 days in RPMI 1640 (Invitrogen) supplemented with 10% heat-inactivated FCS (Dutscher), L-glutamine (Invitrogen), GM-CSF (20 ng/mL; Immunotools), and IL-4 (20 ng/mL; Immunotools). Cell cultures were fed every 2 days with complete medium supplemented with the cytokines previously mentioned. Before infection, we systematically checked the differentiation/activation status of the monocyte-derived DCs by flow cytometry, using antibodies against CD1a, CD14, CD83, and HLA-DR. Only samples presenting the expected phenotype for non-activated DCs – CD1a+, CD14-, CD83-, and HLA-DRlow – were used in downstream experiments. The resulting monocyte-derived DCs were then infected with MTB for 18 h at a multiplicity of infection of 1-to-1, as previously described ([Barreiro et al. 2012](#_ENREF_5)).

For biosecurity reasons the ChIP-seq and ATAC-seq experiments were performed using heat-killed bacteria instead of live MTB. In order to evaluate the extent to which using heat-killed bacteria could result in a different transcriptional response to that induced by live MTB, we used the Illumina HumanHT-12 v4 Expression BeadChip array to compare the genome-wide transcriptional responses observed in DCs in response to live MTB to those observed when DCs from the same donors were exposed to different amounts of heat-killed MTB bacteria. Low-level microarray processing including normalization of the data and variance stabilizing transformation were performed as previously described ([Barreiro et al. 2012](#_ENREF_5)). We found that using the equivalent of 5 heat-killed bacteria to 1 DC leads to virtually the same transcriptional response at 18 hours to that observed with live MTB (r = 0.91; Supplemental Fig. 13).

**CFSE proliferation assay.** DCs and THP-1 cells were covalently labeled with Carboxyfluorescein Diacetate Succinimidyl Ester (CFSE) (Life Technologies) as described in detail elsewhere ([Quah and Parish 2010](#_ENREF_17)). Briefly cells were washed with PBS and resuspended with 5 mM CFSE. After a 5 min incubation at room temperature, cells were thoroughly washed with PBS containing 5% FCS before plating in complete culture medium.

**DNA and RNA Extractions.** DNA from infected and non-infected DCs was extracted using the PureGene DNA extraction kit (Gentra Systems). Total RNA was extracted from the same samples using the miRNeasy kit (Qiagen). RNA quantity was evaluated spectrophotometrically, and the quality was assessed with the Agilent 2100 Bioanalyzer (Agilent Technologies). Only samples with no evidence for RNA degradation (RNA integrity number > 8) were kept for further experiments.

**MethylC-seq library preparation and sequencing.** DNA from infected and non-infected DCs (6 ug) was spiked with 30 ng of unmethylated cl857 *Sam7* Lambda DNA (Promega, Madison, WI) and sonicated to an average length of ~100bp using a Covaris ultrasonicator under the following settings for 16 cycles: Duty cycle: 10%; Intensity: 5; Cycles/burst: 100. The sonicated product was then subjected to repair of 3’ and 5’ ends followed by the addition of a non-templated dA-tail before ligation to cytosine-methylated adapters provided by Illumina (Illumina, San Diego, CA), as per manufacturerʼs instructions for genomic DNA library construction. Adapter-ligated DNA of 100-200 bp was isolated by 2% agarose gel electrophoresis, and sodium bisulfite conversion was performed on the resulting sample using the MethylCode™ Bisulfite Conversion Kit (Invitrogen) as per manufacturer’s instructions. Half of the bisulfite-converted, adapter-ligated DNA molecules was enriched by six cycles of PCR with the following reaction composition: 2.5 U of uracil-insensitive *PfuTurboCx* Hotstart DNA polymerase (Agilent), 5 μl 10X *PfuTurbo* reaction buffer, 25 μM dNTPs, 1 μl PE Primer 1.0 (Illumina), 1 μl PE Primer 2.0 (Illumina) (50 μl final volume). The thermocycling parameters were: 95ºC 2 min, 98ºC 30 sec, then 6 cycles of 98ºC 15 sec, 60ºC 30 sec and 72ºC 4 min, ending with one 72ºC 10 min step. The reaction products were purified using the QIAquick PCR spin column (Qiagen). Two separate PCR reactions were performed on subsets of the adapter-ligated, bisulfite-converted DNA, yielding two independent libraries from the same biological sample. The quality of the libraries was checked on a Bioanalyzer followed by quantification of the libraries by qPCR using the KAPA Library Quantification Kit prior to sequencing. Samples were sequenced on an Illumina HiSeq 2000 using 50- and 59-bp single-end reads. The sodium bisulfite non-conversion rate was calculated as the percentage of cytosines sequenced at cytosine reference positions in the Lambda genome.

**Tet-assisted bisulfite sequencing (TAB-seq).** TAB-Seq libraries were performed as previously described ([Yu et al. 2012](#_ENREF_57)). Genomic DNA was spiked with 0. 5% of *M. SssI* methylated lambda DNA and 0.25% of 5hmC spike-in control (where all cytosines were 5hmC) and then sonicated to 200-500bp with a Covaris ultrasonicator. The *M. SssI* methylated lambda DNA and the 5hmC spike-in control were used to evaluate the conversion rate of C/5mC and protection rate of 5hmC, respectively (see TAB-Seq data processing section). The mixed genomic DNA was glucosylated in 50 mM HEPES (pH 8.0), 25 mM MgCl2, 2 μM βGT and 200 μM UDP-Glc at 37 oC for 1.5 h. The glucosylated DNA was then purified with QIAquick PCR purification Kit (Qiagen) and eluted in Milli-Q water. The oxidation reactions were performed in multiple 50-μl solution containing 50 mM HEPES (pH 8.0), 100 μM ammonium iron (II) sulfate, 1 mM α-ketoglutarate, 2 mM ascorbic acid, 2.5 mM DTT, 100 mM NaCl, 1.2 mM ATP, 10 ng/μl sheared genomic DNA and 3 μM recombinant mTet1. After incubating the reaction at 37 oC for 1.5 h, 1 μl proteinase K (20 mg/ml) was added, which was followed by another 1 h incubation at 50 oC. The oxidized genomic DNA was cleaned up with Micro Bio-Spin 30 Columns (Bio-Rad) first, then applied to QIAquick PCR purification kit (Qiagen). The purified DNA is eluted in EB buffer. After the treatment, we performed bisulfite conversion and library preparation following a protocol identical to that for the MethylC-seq libraries (described above). Samples were sequenced on an Illumina HiSeq 2000 using 100-bp paired-end reads.

**5hmC staining.** The protocol was adapted from Santos et al. ([Santos et al. 2003](#_ENREF_18)). DCs were cultured on poly-L-lysine-coated coverslips and fixed for 30 min in 4% paraformaldehyde in PBS and permeabilized with 0.2% Triton X-100 in PBS for 30 min at room temperature (RT). Cells were then washed with 0.05% Tween 20 in PBS and were treated with 1 M HCl plus 0.1% Triton X-100. After 30 min at 37°C, cells were incubated with 100 mM Tris/HCl (pH 8.5) for 30 min and blocked for 2 h in PBS with 1% BSA, 0.05% Tween-20 and 2% goat serum. Cells were incubated with 5-Hydroxymethylcytosine antibody (ActiveMotif), followed by Alexa 488 goat anti-rabbit antibody (Life Technologies) for 1 h at RT. The slides were mounted with Fluoromount G (SouthernBiotech), and cells counterstained with DAPI to localize the nucleus. A laser-scanning microscope (Zeiss LSM 700) in the tile scan mode was used to capture a mosaic of images. Fluorescence was quantified using the Fiji software. Average fluorescence estimates were calculated from 1,769 non-infected cells and 1,532 MTB-infected cells.

**Western blotting.** Cells were lysed in RIPA buffer (Life Technologies) containing complete protease inhibitor cocktail (Roche) and protein concentration was determined by the BCA protein assay. Equal amounts of protein (250 ng protein/lane) were separated using SDS-PAGE and subsequently transferred to Immobilon-P membrane (Millipore). Membranes were blocked in TTBS /5% non-fat dry milk powder and incubated with rabbit anti-TET2 antibody (GeneTex) and HRP-conjugated goat anti-rabbit immunoglobulins (Amersham).

**Quantitative real-time PCR.** First-strand cDNAs were generated from 500 ng of total RNA using qScript™ cDNA SuperMix (Quanta Biosciences) in a final reaction volume of 10 µl. Expression levels were determined using TaqMan® gene expression assays (Life technologies), with probes specifically hybridising *TRAFD1* (Hs00198630\_m1), *IRF4* (Hs01056533\_m1), *REL* (Hs00968440\_m1), *CD83* (Hs00188486\_m1), *NFKB1* (Hs00765730\_m1) and *BCL2* (Hs00608023\_m1). Normalization was performed using an endogenous housekeeping gene encoding *GAPDH* (Hs02758991\_g1). PCR reactions were performed in a final volume of 10 μl, containing 10 ng of cDNA, 1X of probe and 1X of TaqMan® Fast advanced Master Mix (Life technologies). PCR cycle parameters were 50°C for 2 min, 95°C for 20 s, 40 cycles of 95°C for 1 s and 60°C for 20 s. Common threshold fluorescence for all the samples was set into the exponential phase of the amplification and determined the CT, corresponding to the number of amplification cycles needed to reach this threshold. Relative gene expression quantification was performed using the 2−ΔCT method.

**Bisulfite pyrosequencing.** PyroMark Assay Design 2.0 (Qiagen, Inc.) software was used to design the bisulfite pyrosequencing assay covering the targets regions. DNA was subjected to bisulfite conversion using the EZ DNA Methylation Kit (Zymo Research). HotstarTaq DNA polymerase kit (Qiagen, Inc.) was used to amplify the target regions using the biotinylated primer set with the following PCR conditions: 15 minutes at 95°C, 45 cycles of 95°C for 30s, 58°C for 30s, and 72°C for 30s, and a 5 minute 72°C extension step. Streptavidin-coated sepharose beads were bound to the biotinylated-strand of the PCR product and then washed and denatured to yield single-stranded DNA. Sequencing primers were introduced to allow for pyrosequencing (Pyromark™ Q96 MD pyrosequencer, Qiagen, Inc.).

**RNA-seq library preparation and sequencing.** RNA-seq libraries for the six samples for which we collected MethylC-seq were generated via polyA+ selection of mRNA from total RNA using the TruSeq RNA Sample Prep Kit v2 (Illumina). In addition, for the two individuals from whom we collected histone mark ChIP-seq data, we also performed RNA-seq on the whole transcriptome following ribosomal depletion using the Ribo-Zero Gold depletion and the Illumina Total Stranded RNA Library kits (Illumina). We did so in order to be able to capture enhancer RNAs, which are usually non-polyadenylated ([Kim et al. 2010](#_ENREF_11)). RNA-seq libraries were sequenced as 50-bp single-end (polyA+ fraction) and 100-bp paired-end reads (ribo-minus) on an Illumina HiSeq 2500.

**ChIP-seq library preparation and sequencing.** Samples from infected and non-infected DCs from two individuals were crosslinked with 1% w/v formaldehyde for 10 min at RT and immediately quenched for 5 min with 125 mM Glycine at RT. The formaldehyde fixed samples were then sonicated to 100-400 bp using a Bioruptor (Diagenode) and then ChIP-DNA prepared using the IP-Star Compact (Diagenode) Indirect method with an Antibody-Antigen incubation of 10 hr, Bead incubation of 2 hr, and 4x 20 min wash steps. Approximately 1 million cells were used for each ChIP and ~50,000 cells for the input. The following antibodies were used: H3K4me1 (Company: CST, Cat. No.: 5326P, Lot No.: 1), H3K4me3 (CST, 9751BC, 7), H3K9me3 (MABI, 0318, 13001), H3K27me3 (MABI, 0323, 13001), H3K27ac (Abcam, Ab4729, GR119051), and H3K36me3 (MABI, 0333, 12003). ChIP and Input libraries were prepared using the Illumina TruSeq Nano DNA kit, with alterations including: PCR enrichment (14 cycles) prior to size selection and use of the PippinPrep method (SAGE Science) instead of the SPRI method for size selection (200-400 bp). Libraries were sequenced on an Illumina HiSeq 2000. We pooled 8 libraries per lane and sequenced the lane twice to reduce the possibility of lane effects. Each library was sequenced using 50-bp single-end reads.

**ATAC-seq library preparation and sequencing.** ATAC-seq libraries were generated from 100,000 cells, as previously described ([Buenrostro et al. 2013](#_ENREF_6)) and sequenced on an Illumina HiSeq 2500 using 100-bp paired-end reads.

**MethylC-seq data processing.** We used the tool Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim\_galore/) to trim off adapter sequences incorporated in the read and remove bases with a Phred base quality score below 20. PCR duplicates were removed using a Perl script that is part of the Bismark package (deduplicate\_bismark\_alignment\_output.pl). The resulting reads were mapped to the human reference genome (GRCh37/hg19) and lambda phage genome using Bismark ([Krueger and Andrews 2011](#_ENREF_12)) (with the options -p 12 -N 1), which uses Bowtie 2 ([Langmead and Salzberg 2012](#_ENREF_13)) and a bisulfite converted reference genome (C-to-T and a G-to-A) for read mapping. Only reads that had a unique alignment and a maximum of one mismatch were retained. The context of each C was determined, which allowed us to classify each C of the genome as CpG, CHH, or CHG, where H is either an A, T, or C nucleotide. Methylation levels for each CpG site were estimated by counting the number of reported C (‘methylated’ reads) divided by the total number of reported C and T (‘methylated’ plus ‘unmethylated’ reads) at the same position of the reference genome using Bismark’s methylation extractor tool. The same strategy was also applied for non-CpG methylation (CHG context, where H is either an A, T, or C nucleotide). We performed a strand-independent analysis of CpG methylation where counts from the two Cs in a CpG and its reverse complement (position *i* on the plus strand and position *i*+1 on the minus strand) were combined and assigned to the position of the C in the plus strand.

To assess MethylC-seq bisulfite conversion rate, the frequency of unconverted cytosines (C basecalls) at lambda phage CpG reference positions was calculated from reads uniquely mapped to the lambda phage reference genome. Overall, bisulfite conversion rate was >99% in all of the samples (Supplemental Table 1).

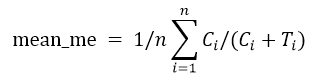
**TAB-seq data processing.** We used Trim Galore in paired-end mode to remove adapter sequences and low quality score bases (Phred score < 20). The resulting reads were mapped in bisulfite mode to the human reference genome (GRCh37/hg19) (and lambda phage + control II sequence) using Bismark with the following parameters: --bowtie2 -p 12 -N 1. PCR duplicates were removed using the deduplicate\_bismark\_alignment\_output.pl script. In total, we obtained ~430 million paired-end reads, of which 87% were unambiguously mapped to the reference genome with a mean sequencing coverage of 10.1X and 9.3X in non-infected and infected DCs, respectively (Table S1). Similar to MethylC-seq data, hydroxymethylation levels for each CpG site were estimated by counting the number of reported C (‘hydroxymethylated’ reads) divided by the total number of reported C and T (‘hydroxymethylated’ plus ‘non-hydroxymethylated’ reads) at the same position of the reference genome using Bismark methylation extractor with parameters --ignore\_r2 2 --no\_overlap. Cytosine non-conversion rate (i.e., failed 5mC conversion by Tet1 and failure of bisulfite conversion) was assessed by calculating the frequency of C base calls at lambda CpG reference positions from reads uniquely mapped to the lambda reference. 5hmC protection rate was calculated likewise using CpG reference positions in control II sequence.

**ChIP-seq data processing.** We started by trimming adapter sequences and low quality score bases using Trim Galore. The resulting reads were mapped to the human reference genome (GRCh37/hg19) and PCR duplicates were removed using Picard tools (http://broadinstitute.github.io/picard/). The alignment software Bowtie 2 was then used with the following options: -p 12 -N 1. Only reads that had a unique alignment and no more than one mismatch were retained. For each of the histone marks in each of the conditions, we obtained an average of 58.5 ± 9.5 SD million reads (Table S1) when combining data from the two biological replicates. Pearson correlation revealed a high concordance between the histone ChIP-seq signals for the two biological replicates sequenced for each of the histone marks (mean r = 0.94 and range = 0.87-0.99; Supplemental Fig. 14), which allowed us to merge them for downstream analyses.

**RNA-seq data processing and identification of differentially expressed genes upon MTB infection.** Adaptor sequences and low quality score bases were first trimmed using Trim Galore. The resulting reads were aligned to the human genome reference sequence (GRCh37/hg19) using the TopHat2 software package ([Kim et al. 2013](#_ENREF_10)) with a TopHat transcript index from RefSeq. The number of read fragments overlapping with annotated exons of genes was tabulated using HTSeq ([Anders et al. 2014](#_ENREF_3)) using the following parameters: -q -m intersection-nonempty -s no. Using normalized gene counts for 6 infected and 6 non-infected samples, we identified genes whose expression levels were significantly altered following MTB infection of DCs using the R package DESeq2 ([Anders et al. 2013](#_ENREF_2)). Using a paired design, we considered a gene as differentially expressed if statistically supported at a Benjamini and Hochberg ([Hochberg 1995](#_ENREF_9)) false discovery rate (FDR) < 1 × 10-4 with a |log2 fold change| > 1. Lowly expressed or non-expressed genes with a mean normalized read count of 0 across all samples were discarded.

**Genomic annotation and mRNA TSS collection.** Gene locations used in Figure 1 were defined based on the GRCh37/hg19 assembly. Annotation of known Ensembl transcripts was obtained from UCSC (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/ensGene.txt.gz). Since genes can have multiple transcripts, we selected the 5’-most transcription start site (TSS) on the positive strand as the single TSS associated with each gene. The reverse (3’ most TSS) was done for genes on the negative strand. We limited downstream analysis to protein-coding genes, resulting in 20,745 TSSs in total. Similarly, annotations for retro-elements (i.e., LINEs and SINEs), CpG islands, exons and introns were downloaded from the UCSC.

**Integrated analysis of gene expression and 5mC.** FPKM (fragments per kilobase of exon per million fragments) values of expression were calculated using Cufflinks ([Trapnell et al. 2010](#_ENREF_22)). Genes were then classified into quartiles based on their gene expression levels at baseline (untreated) state: 1st quartile is lowest and 4th is highest. Gene bodies and 20-kb regions upstream and downstream were each divided into 50 intervals. We gathered methylation data from windows within each of these intervals and plotted the mean methylation level (mean\_me) for all windows overlapping each position. For each bin containing n sites (*i*):



where C = read supporting methylated cytosine, T = read supporting unmethylated cytosine, *i* = position of cytosine and n = total number of cytosine positions.

**Principal component analysis based on methylomes.** Principal component analysis (PCA) of DC methylomes and those of other cell types for which MethylC-seq data was publicly available was performed on a set of 2,724,731 CpG sites that were sequenced at coverage ≥5 across all cell types or tissues using MethylKit tools ([Akalin et al. 2012](#_ENREF_1)). The following cell types or tissues were used: neuroectoderm, neuroepithelial, glia, fetal (fheart, fthymus, fmuscle, fadrenal, fbrain), adipocyte, colon mucosa (cmucosa), substantia nigra (snigra), B-cell, T-cells (cd4, cd8, cd34), dendritic cells (dc81, dc82, dc83, dc87, dc89, dc91, hippocampus, hspc, liver, neutrophil, peripheral blood mononuclear cell (pbmc), and sperm.

**Identification of MTB-DMRs.** The summarized methylation estimates of strand-merged CpG sites from the 6 infected and 6 non-infected samples were used to identify MTB-induced differences in methylation, using the R package BSmooth/BSseq ([Hansen et al. 2012](#_ENREF_8)) with the following parameters: ns = 25 and h = 200. BSmooth implements a smoothing method that uses a local likelihood approach to estimate the smoothed probability of methylation at each site, taking into account the spatial correlation between nearby sites and placing greater weight on sites with higher coverage. To minimize noise in methylation estimates due to low-coverage data, we restricted the differential methylation analysis to CpG sites with coverage of ≥4 sequence reads in at least half of the DC samples in each condition, which still allowed us to interrogate changes in methylation levels at ~20 million CpG sites. Moreover, to eliminate effects caused by polymorphisms, C nucleotides that overlapped with known SNPs (dbSNP132; http://www.ncbi.nlm.nih.gov/SNP/) were removed. We identified MTB-induced differentially methylated regions (MTB-DMRs) as regions containing at least 3 consecutive CpG sites that were significantly differentially methylated using a paired t-test (|t-statistic| > 4.032 at *P* = 0.01) and that exhibited at least a 10% difference in mean methylation levels between treated and untreated samples.

**Assigning MTB-DMRs to genes.** To assign each MTB-DMR to a gene, we use the following rationale: if an MTB-DMR was located within a gene body the MTB-DMR was assigned to that gene; otherwise, we assigned each MTB-DMR to the closest TSS from the center position of the MTB-DMR. If the closest TSS was further away than 250 kb the gene assigned to that MTB-DMR was not included in any of the downstream analysis. These criteria meant that almost all of the MTB-DMRs (93%) were associated with genes. Notably, extending our distance cut-off even further, so that all MTB-DMRs were assigned to genes, produced qualitatively identical GO analysis results: 97.6% of all enriched GO terms (FDR of 20%) identified using the 250 kb cutoff are also identified when all MTB-DMRs were assigned to genes using the same statistical cutoff.

**5hmC analysis.** Metagene profiles of 5hmC were plotted as described above for the 5mC data. To plot 5hmC profiles around MTB-DMRs, the weighted mean methylation was calculated for each contiguous 100-bp bin from 3 kb upstream to 3 kb downstream of the central position of the MTB-DMR. Only CpG sites with sequencing coverage ≥4 were included in the analyses.

**Chromatin state annotation and dynamics.** We used ChromHMM ([Ernst and Kellis 2012](#_ENREF_7)) with default parameters to segment the genome into different chromatin states based on six histone modifications and ChIP input. A model was learned separately for both conditions (i.e., infected and non-infected samples), producing segmentations based on the most likely state assignment of the model. We selected a 12-state model in order to allow sufficient resolution to resolve biologically meaningful chromatin patterns. We further combined segments that had comparable histone patterns, resulting in 7 biologically meaningful chromatin states (Supplemental Fig. 5). To evaluate the enrichment of each chromatin state at MTB-DMRs, we first assigned each MTB-DMR to a particular chromatin state based on the chromHMM segment overlapping with its midpoint. We then calculated the frequency of MTB-DMRs that were assigned to a particular chromatin state, and normalized this value against the expected frequency based on the amount of genome covered by that state. We note that we have also performed similar analyses using a unified model that learns and defines chromatin states in both infected and non-infected DCs at the same time (in contrast to doing it separately in each condition) and all our results and conclusions remain virtually the same (Supplemental Fig. 15).

To test the hypothesis that regions that changed DNA methylation are also more likely to change chromatin state (compared to other regions of the genome), we randomly sampled an equal number of genomic regions (n = 1,714) matched for the same chromatin states observed within hypomethylated regions in non-infected DCs. We then counted the proportion of these random regions that changed chromatin state after infection. The expected distribution of chromHMM state transitions was generated using 1000 simulations and was compared to the proportion of chromatin changes observed among hypomethylated regions. A similar resampling strategy was used to test for an enrichment of hypomethylated regions marked as heterochromatin/repressed before infection (n = 790) and that gained *de novo* enhancer marks upon MTB infection.

**Enhancer classification of hypomethylated regions based on chromatin state.** In order to define different categories of enhancers, we centered our analysis on H3K4me1 signals. If H3K4me1 was present in the baseline (untreated) state, such region was defined as a predefined enhancer. Therefore, predefined enhancers were simply defined as regions that overlapped with a chromHMM segment of either state 3, 4, or 5 (active or inactive enhancers) prior to MTB infection. If H3K4me1 was not found to be enriched against input in the baseline state but H3K4me1 and/or H3K27ac were induced by MTB infection, the region was defined as a *de novo* enhancer. Therefore, *de novo* enhancers were defined as regions that overlapped with a chromHMM segment of state 7 (heterochromatin/repressed) that transitioned to either state 3, 4, or 5 (active or inactive enhancers) after MTB infection.

**ChIP-seq profiles around MTB-DMRs.** Global visualization for chromatin modifications, genome accessibility and RNA patterns around MTB-DMRs was accomplished with ngs.plot package ([Shen et al. 2014](#_ENREF_19)) using default parameters. For each MTB-DMR, data was analyzed from 3 kb upstream to 3 kb downstream of the central position of the MTB-DMR unless otherwise indicated. To compensate for differences in total sequencing read depth among samples, all ChIP-seq read counts were first normalized against the total read count for the same sample. Next, the normalized number of reads was subtracted from the normalized number of reads in the input within a 100-bp scanning window, and the subtracted value was used for further analysis and plotting. For visualization purposes, pseudo counts were added if the resulting values were negative.

**ATAC-seq data processing and footprinting analysis.** ATAC-seq reads were mapped to the human reference genome (GRCh37/hg19) using BWA-MEM ([Li and Durbin 2009](#_ENREF_14)), using default parameters. Only reads that had a unique alignment and mapping quality of ≥10 were retained. Similarly, ngs.plot was used to plot ATAC-seq profiles around MTB-DMRs. To detect TF binding footprints in the ATAC-seq data we used the program Centipede ([Pique-Regi et al. 2011](#_ENREF_16)) in two steps. In the first step, we determined which transcription factors were active (i.e., had motif instances with footprints) before and after infection using a reduced set of motif instances (5K-15K) for each TF as defined in Moyerbrailean *et al.* ([Moyerbrailean et al. 2014](#_ENREF_15)). In the second step, we scanned the entire genome for motif instances matching the original PWM, and we ran Centipede in parallel for the two conditions in order to make the posterior probabilities comparable. For both steps, to run Centipede the aligned paired-end reads were separated into four bins depending on the fragment length ([40, 99], [100, 139], [140, 179] and [180, 250] in bp). As Tn5 transposase contacts and duplicates 9 bp of DNA ([Buenrostro et al. 2013](#_ENREF_6)) we take as the cleavage site the middle nucleotide. To do so, we shifted 4 bp from the 5'-end positions towards the center of the fragment. Then for each motif we built a matrix that counted Tn5 cleavage events, where each row represented a motif instance (i.e., a candidate binding site), and each column represented a spatial location with respect to the TF binding site in bp (i.e., relative cleavage site). This matrix was constructed separately for each fragment length bin and each strand orientation (with respect to the motif match, or to the reference strand if the motif was palindromic). We used a window size of 300 bp on either side of the motif match. We then concatenated all 8 matrices and fed them as input data to Centipede, together with the PWM score.

To determine which TFs were active in the first step, we calculate a Z-score corresponding to the PWM effect in the prior probability in Centipede’s logistic model and we determined as active those that had a Bonferroni-corrected *P* < 0.05. The Z-score corresponds to the  parameter in:



where  represent the prior probability of binding in Centipede’s model in motif location . In the second step, we first trained Centipede assuming that the footprint was bound in the two conditions. Then, we fixed the model parameters and generated a likelihood ratio and posterior probability  for each condition  separately and for each site . To detect if the footprint was more active in one of the two conditions, we fit a logistic model that included an intercept for each condition ( and ), the PWM effect , and PWM times the treatment effect :



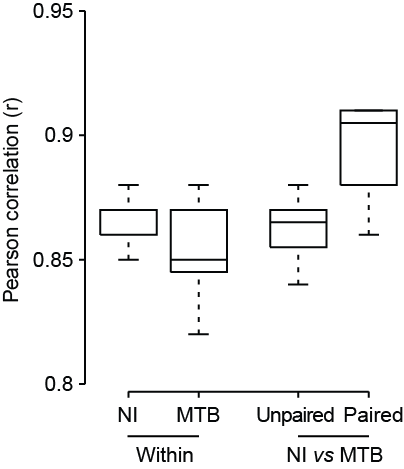
where  is an indicator variable that takes the value 1 if t = “treatment” and 0 if t = “control’’. We then calculated a Z-score for the interaction effect , corresponding to the evidence for condition specific binding.

**Peak detection.** Peaks on ChIP-seq (using input IP as a control) and ATAC-seq data were called using the MACS2 software suite ([Zhang et al. 2008](#_ENREF_24)) with default parameters.

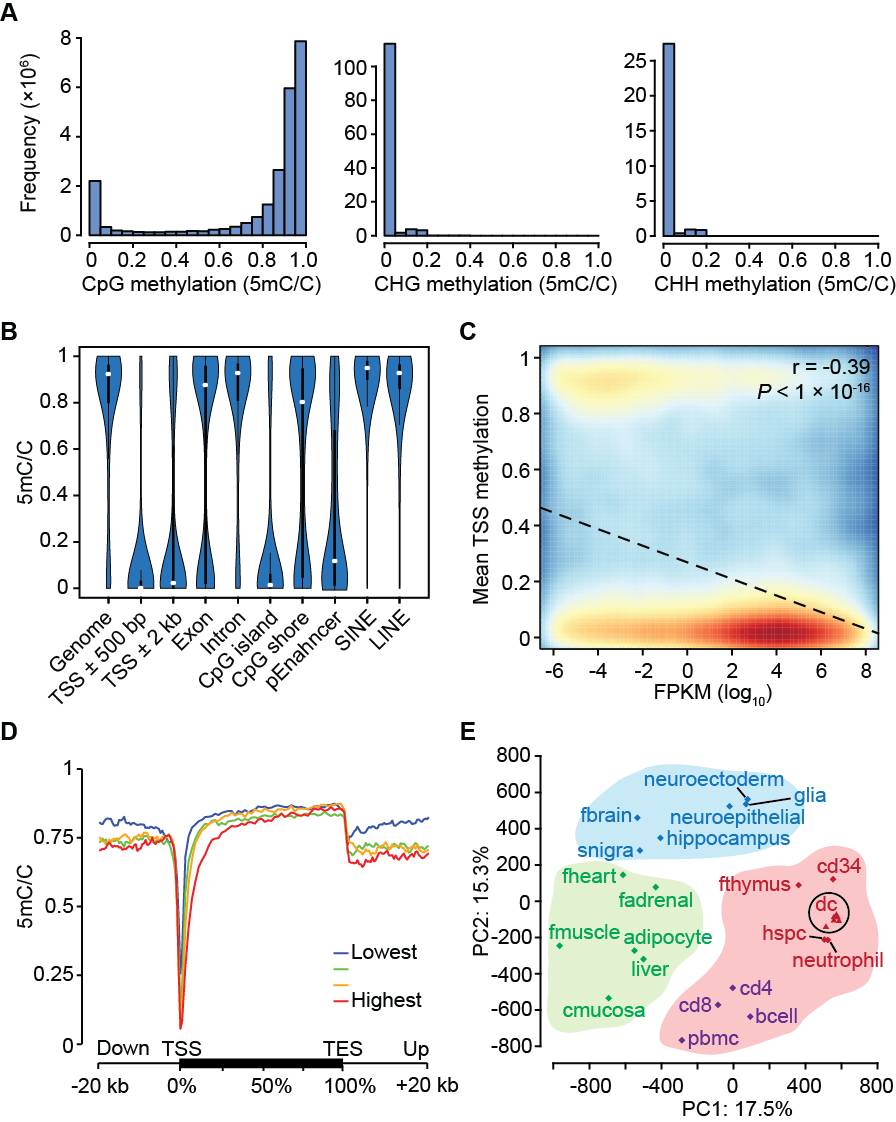
**Gene Set Enrichment Analysis.** We used Genetrail ([Backes et al. 2007](#_ENREF_4)) to test for enrichment of functionally annotated gene sets among genes associated with MTB-DMRs (using 250 kb as distance cutoff), using the set of all Ensembl genes as a background. Analysis was done with default parameters and results were corrected for multiple testing by the method of Benjamini and Hochberg ([Hochberg 1995](#_ENREF_9)) to control the False Discovery Rate (FDR).

**Data visualization in the Immune Epigenome Browser.** The browser, implemented using the WashU Epigenome Browser web interface ([Zhou and Wang 2012](#_ENREF_25)), can be accessed at <http://luis-barreirolab.org/EpigenomeBrowser>. Along with RefSeq gene annotations, it includes 25 data tracks showing(*i*) the genomic location of MTB-DMRs; (*ii*) smoothed site-specific 5mC values; (*iii*) 5hmC values; (*iv*) all histone mark ChIP-seq read signals (H3K4me3, H3K4me1, H3K27ac, H3K27me3, H3K36me3, and H3K9me3); (*v*) Tn5-transposase (i.e., chromatin accessibility) read signals; (*vi*) mRNA read signals; and (*vii*) predicted binding sites for the 55 transcription factors that significantly change genome-wide binding levels in response to MTB infection. All data sets are shown for both non-infected (NI) and MTB-infected (MTB) conditions with respect to the GRCh37/hg19 reference sequence. Note that for ease of visualization, several tracks are not shown under default parameters. These can be added by going to: *Tracks* 🡪 *Custom* *tracks* 🡪 *List of all*.

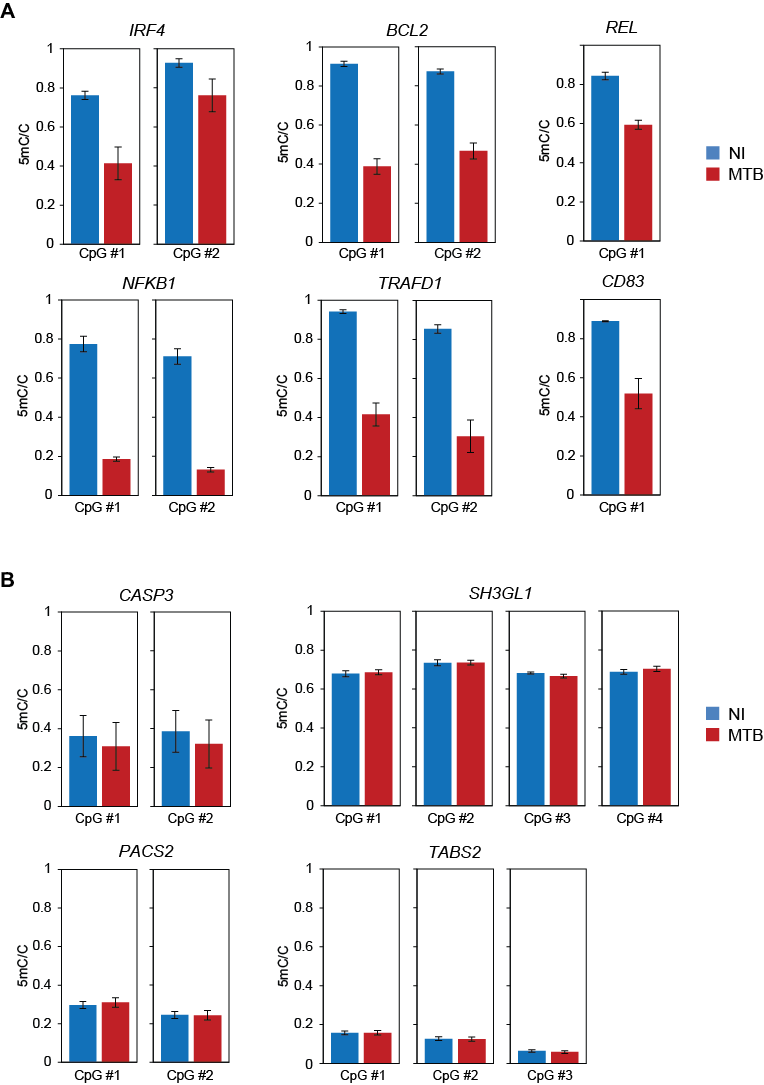
**Supplemental Figures**

****

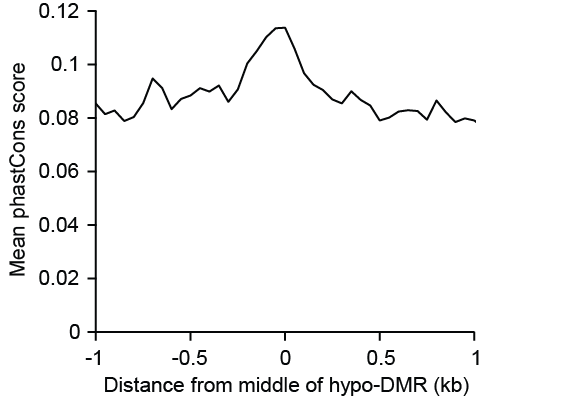
**Supplemental Figure 1.** Correlation between DNA methylation levels among replicates. Boxplots show pairwise comparisons of genome-wide methylation data (CpG sites with ≥5 coverage in all samples) between different individuals *within* either non-infected (NI) and MTB-infected (MTB) groups. Also shown are pairwise comparisons of methylation data between NI and MTB groups from *different individuals* (*unpaired*), and between infected and non-infected samples from the *same individual* (*paired*). Most of the variance in methylation levels is explained by differences between individuals (mean r = 0.87) rather than by differences between conditions (mean r = 0.9). This result is consistent with the fact that, despite our observation that over 1,600 regions are demethylated in response to MTB infection, these regions span a proportionally small subset of all CpG sites in the genome (n = 7,331; 0.035% of all CpG sites that were included in the differential methylation analysis).

****

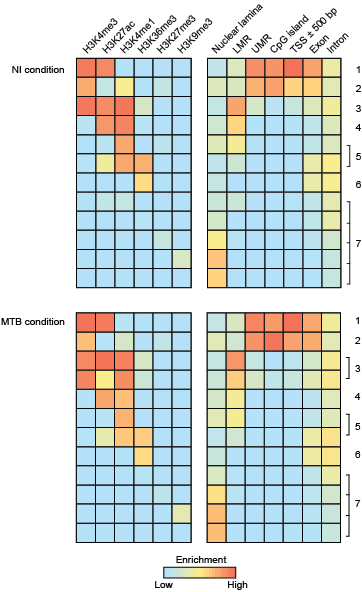
**Supplemental Figure 2.** Global patterns of DNA methylation in human DCs. (*A*) Distribution of genome-wide cytosine methylation in CpG, CHG and CHH (chromosome 1 only) contexts. (*B*) Violin plots showing the distribution of DNA methylation levels at CpG sites located in different genomic regions. The white circles indicate the median DNA methylation levels for each region. Putative enhancers (pEnhancer) represent regions of open chromatin associated with H3K4me1 signal. (*C*) Correlation between mean DNA methylation levels (y-axis) at putative promoter regions (TSS ± 500 bp) and baseline expression levels of the associated genes (x-axis). (*D*) Metagene profiles of DNA methylation levels relative to Ensembl transcripts expressed at different levels in human DCs. We grouped genes into quartiles based on their expression levels in non-infected DCs. (*E*) Principal component analysis based on methylation levels across 21 diverse human cell and tissue samples. Color indicates classification of samples into subgroups of functionally related cells. Blue: brain-associated cell types; Red: hematopoietic cells; Green: other fully differentiated tissues. Detailed sample annotations are listed in the methods section *Principal component analysis based on methylomes*. The 6 dendritic cell (DC) samples profiled in our study are highlighted by a circle. For visual purposes we do not show sperm methylation on the plot because it is a clear outlier on PC2 (PC2 loading = 1369) with respect to all other cell types.

****

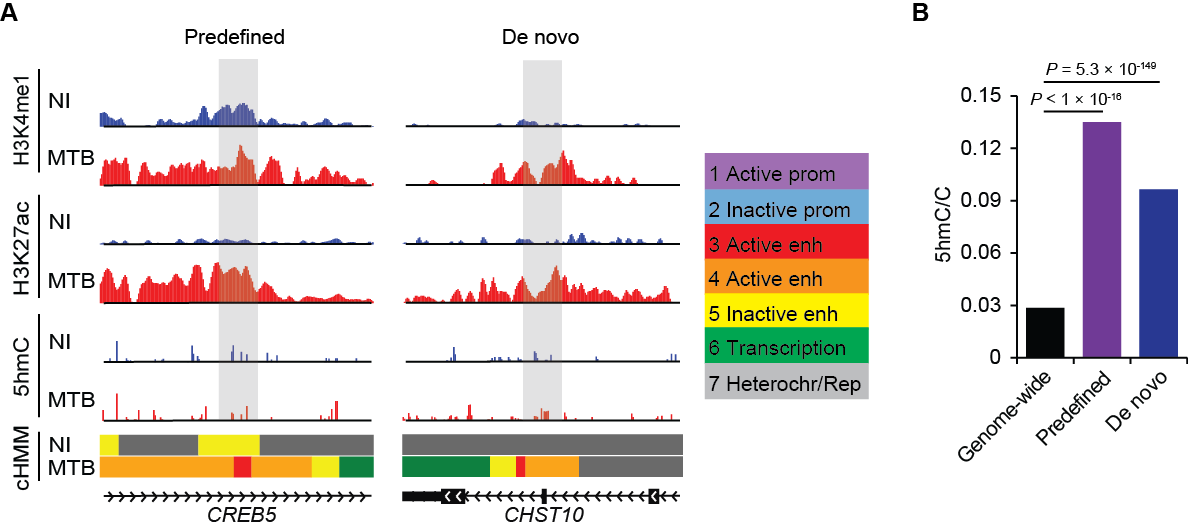
**Supplemental Figure 3.** Validation of differential methylation results by bisulfite pyrosequencing. (*A*) Barplots of the average methylation levels in 10 CpG sites located within 6 hypo-DMRs and (*B*) 11 CpG sites within 4 hyper-DMRs (see Supplemental Table 3 for details). Blue and red boxes represent the DNA methylation levels in non-infected and infected samples, respectively. Data are represented as mean ± s.e.m., n = 5 (except for CpG #2 of *TRAFD1* at MTB condition, where n = 4). The results show high validation rate for hypo-DMRs but not for hyper-DMRs.

****

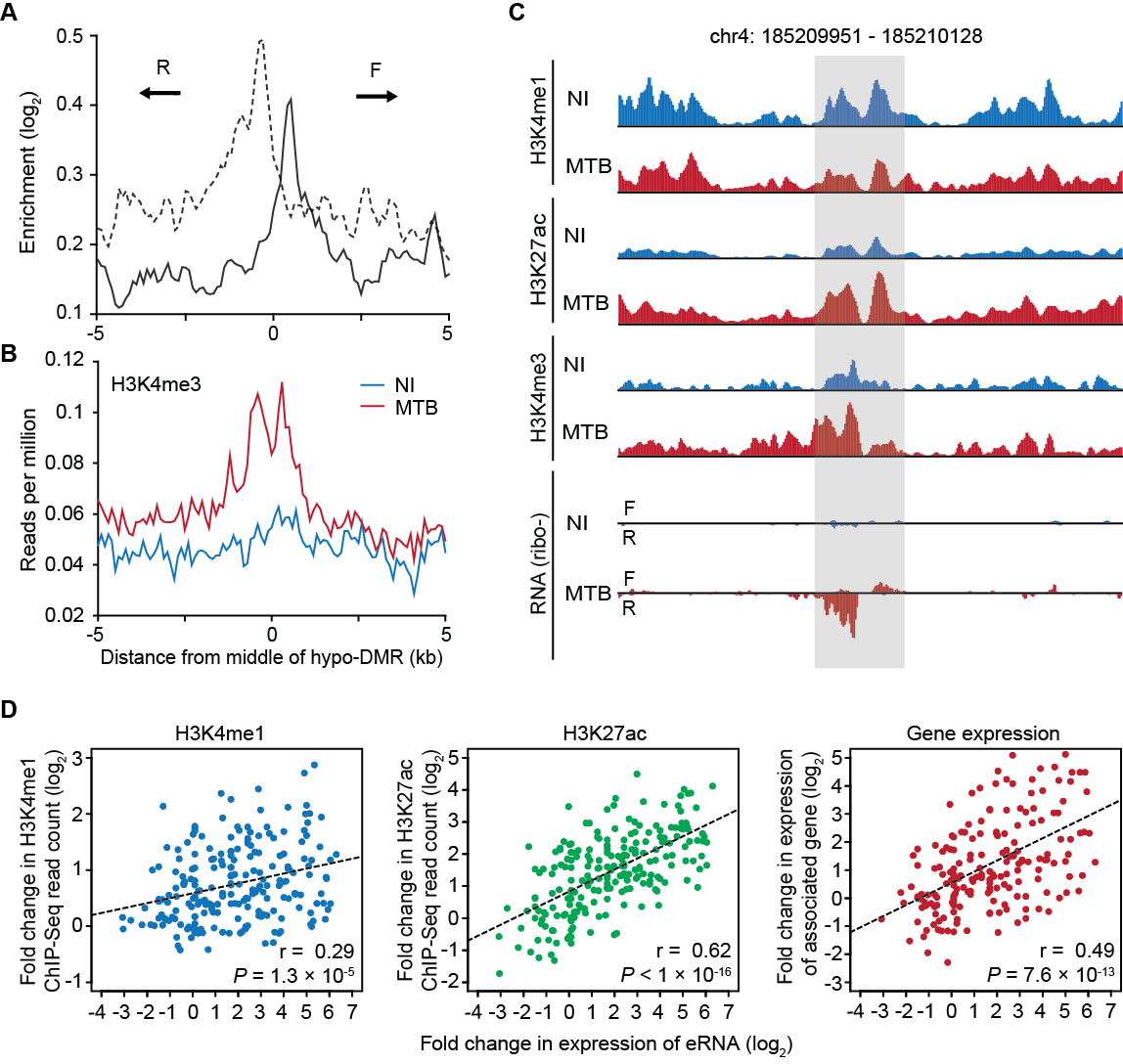
**Supplemental Figure 4.** Average phastCons conservation score ([Siepel et al. 2005](#_ENREF_20)) within 50-bp sliding windows, around the center of hypo-DMRs not associated with promoter regions (i.e, >3kb away from any known TSS).

****

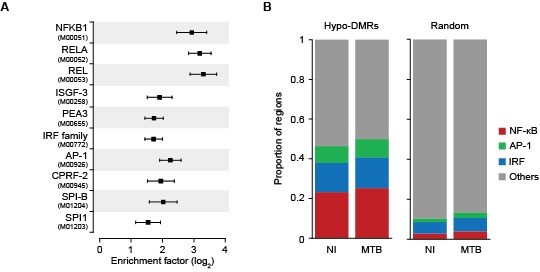
**Supplemental Figure 5.** Chromatin state annotation of infected and non-infected DC genomes. Six histone marks were used to learn a 12-state chromatin model separately for infected and non-infected DCs. Enrichment for emission probabilities for each mark (left panel) and different gene elements (right panel; scaled by column) within each stateare shown, ranging from low/none (blue) to high (red). Chromatin states that had comparable histone patterns and genomic features were combined, resulting in 7 biologically meaningful chromatin states for downstream analyses. Despite the high levels of H3K4me3 we classified state 3 as an active enhancer (in contrast to an active promoter) because of the high levels of H3K4me1 and H3K27ac. Moreover, these regions tend to be located far away from either transcription start sites, CpG islands or unmethylated regions (UMR) and are near lowly methylated regions (LMRs), which represent putative enhancer elements ([Stadler et al. 2011](#_ENREF_21)).

****

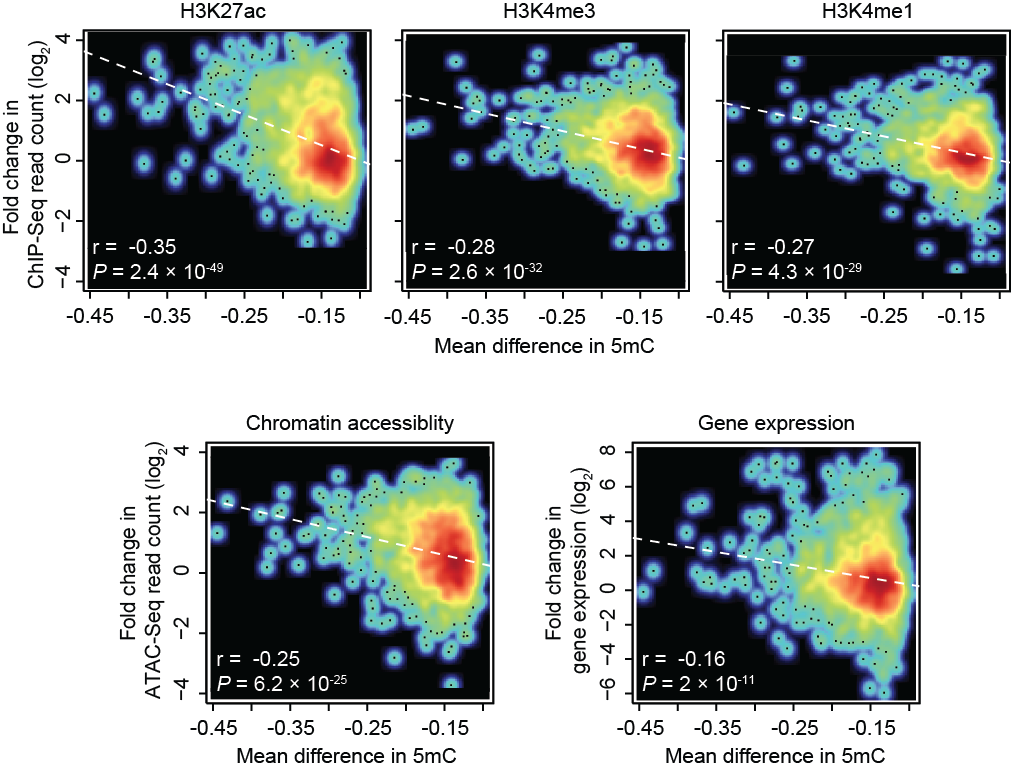
**Supplemental Figure 6.** (*A*) Representative examples of a predefined and *de novo* enhancer at regions exhibiting loss in methylation (gray area). Left panel: a predefined H3K4me1-marked enhancer region that subsequently gains activation marks (H3K27ac) after MTB infection. Right panel: a region that gains *de novo* enhancer marks in response to MTB infection. ChromHMM tracks (cHMM) show the dynamic chromatin state of the locus using combinatorial patterns of 6 histone marks before and after infection. The color code annotation of the chromatin state map is provided on the right. (*B*) Barplots of the hydroxymethylation levels in non-infected DCs showing that both predefined and *de novo* enhancers are enriched in 5hmC prior to MTB-infection.

****

**Supplemental Figure 7.** Relationship between eRNA expression at hypomethylated regions and deposition of histone marks. (*A*) Fold changes (log2) in enhancer-associated RNA expression levels after MTB infection as a function of eRNA location relative to the center of the hypomethylated regions. (F) and (R) denote transcripts originating from the forward and reverse strands, respectively. (*B*) Composite plots of patterns of H3K4me3 before (blue) and after (red) MTB infection ±5 kb around midpoints of hypomethylated regions. (*C*) Gray shading shows an example of a hypomethylated region that displays increased levels of eRNA transcription accompanied by dynamic changes in histone patterns after MTB infection. (*D*) Scatterplots depicting the relationship between the fold change in eRNA expression levels and the fold changes in H3K4me1 (blue), H3K27ac (green) and mRNA expression levels of the associated genes (red). Only regions overlapping with enhancers and showing evidence of eRNA expression in either non-infected or infected DCs (n = 221) were included in the analysis.

****

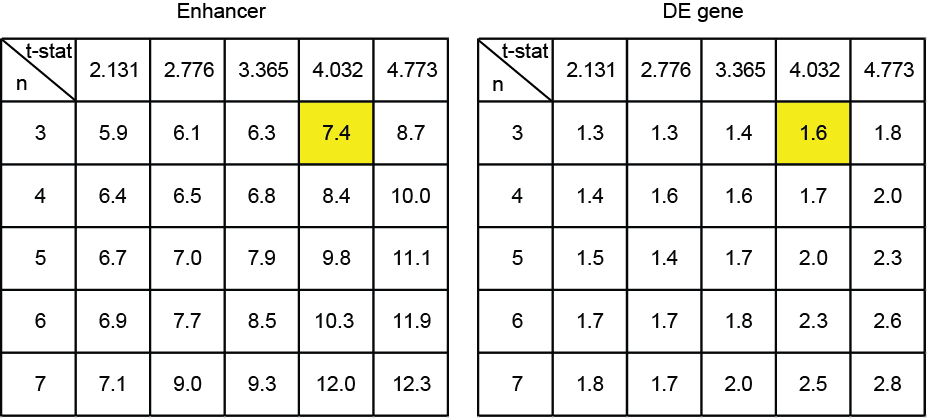
**Supplemental Figure 8.** MTB-DMRs are enriched for signal-dependent TF footprints. (*A*) TF binding motifs for which the number of well-supported footprints (posterior Pr > 0.95) within hypomethylated regions were enriched relative to the genomic background, in MTB infected DCs. The enrichment factors are shown in the x-axis in a log2 scale. The bars around the estimated enrichments reflect the 95% confidence intervals around the estimates. For visualization purposes we only show the top 10 most significantly enriched TF binding motifs. Motif IDs are shown in parentheses below motif names. A complete list of all TF binding motifs for which footprints are enriched within hypomethylated regions can be found in Supplemental Table 5. (*B*) Proportion of hypomethylated regions (y-axis) that have a binding event at motifs associated with TFs in the families of master-regulators NF-κB, AP-1, or IRFs. To test for enrichment of binding of these TFs within these regions, we generated a randomly sampled control set of sequences matched for length and GC content to the observed hypomethylated regions.



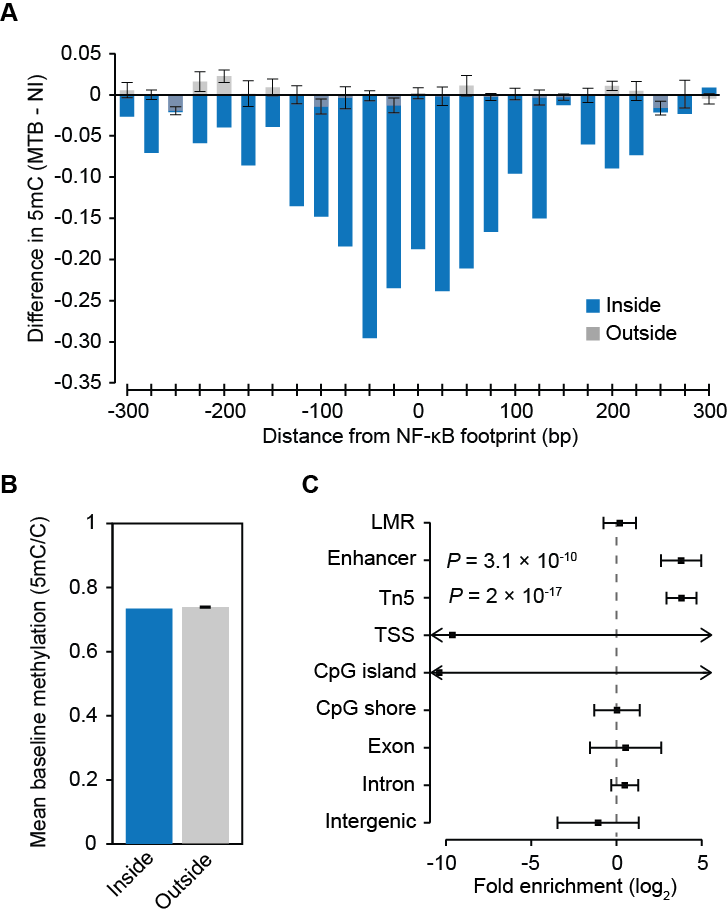
**Supplemental Figure 9.** Smooth scatterplots depicting the relationship between the magnitude of loss in DNA methylation level after infection and changes (in log2 scale) in H3K27ac, H3K4me3 and H3K4me1 histone marks, chromatin accessibility, and mRNA expression levels of nearby genes.

****

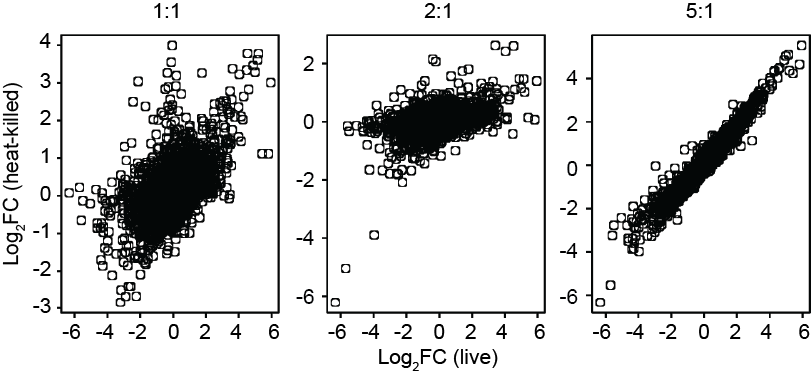
**Supplemental Figure 10.** (*A*) Barplots showing the transcript levels (in FPKM) of *TET1*, *TET2* and *TET3* before (NI) and 18 hours post-infection (MTB) using RNA-Seq. Data are represented as mean ± s.e.m., n = 6. (*B*) Distribution of mean baseline gene expression values (FPKM at log10 scale) from 6 untreated DC samples. Arrows denote the relative expression of *TET1*, *TET2* (2) and *TET3* (3). (*C*) Western blot showing protein levels of TET2 before (NI) and after MTB-infection (MTB), at 18 and 48 hours (h).



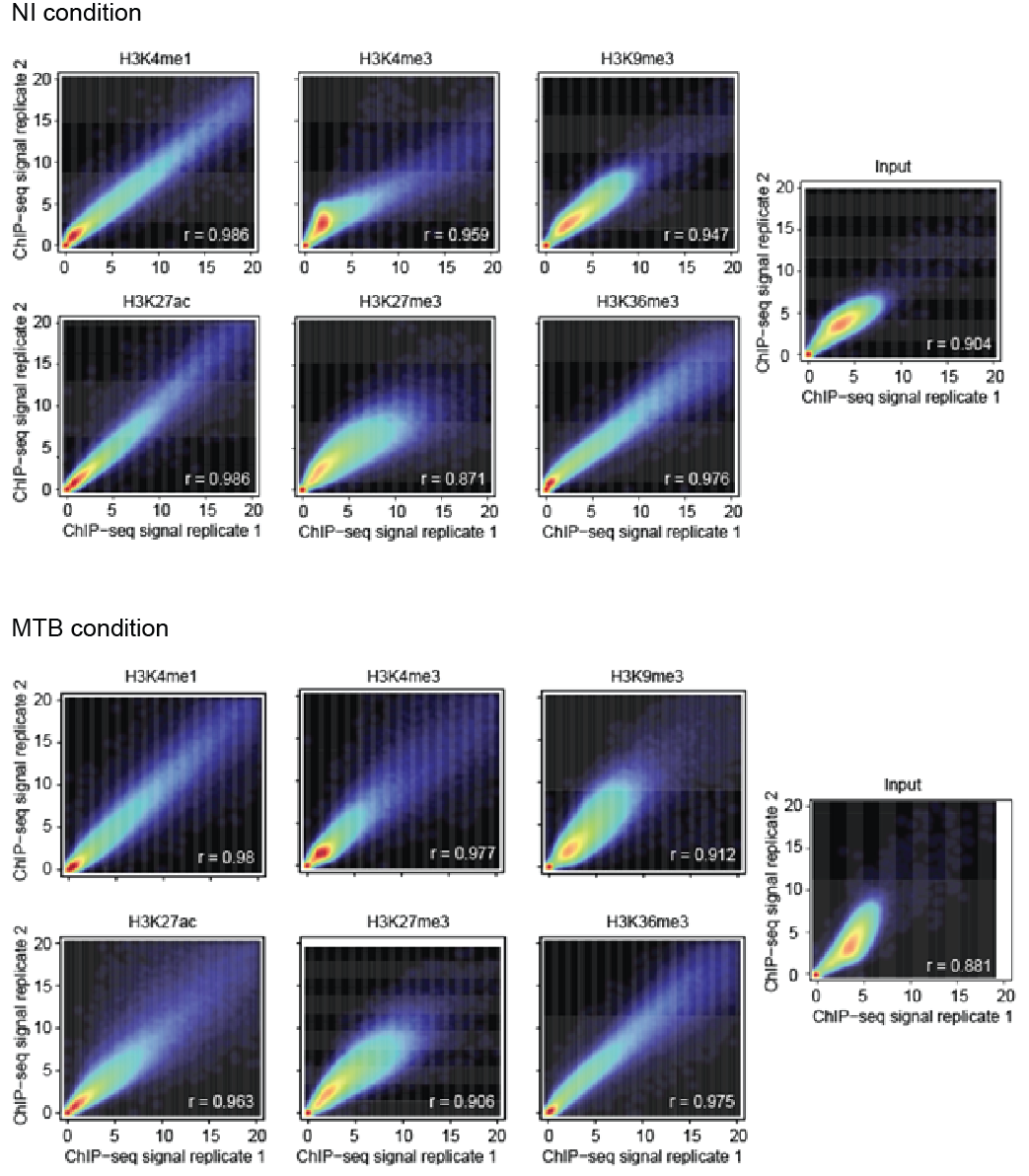
**Supplemental Figure 11.** The enrichment for overlap with enhancer elements and DE genes is robust to the cutoffs used to define hypomethylated regions.The tables show the observed fold enrichments of overlap with enhancer elements (left panel) and differentially expressed genes (right panel) when classifying hypomethylated regions using a different combination of cutoffs for the number of consecutive CpG sites exhibiting a significant difference in methylation (n, row labels) at a given t-statistic value (t-stat, column labels). The minimum difference in methylation levels was fixed at |0.1| as with the cutoff used on the manuscript. The cells in yellow indicate the enrichments observed with the cutoffs used to define hypomethylated regions in the main text.

****

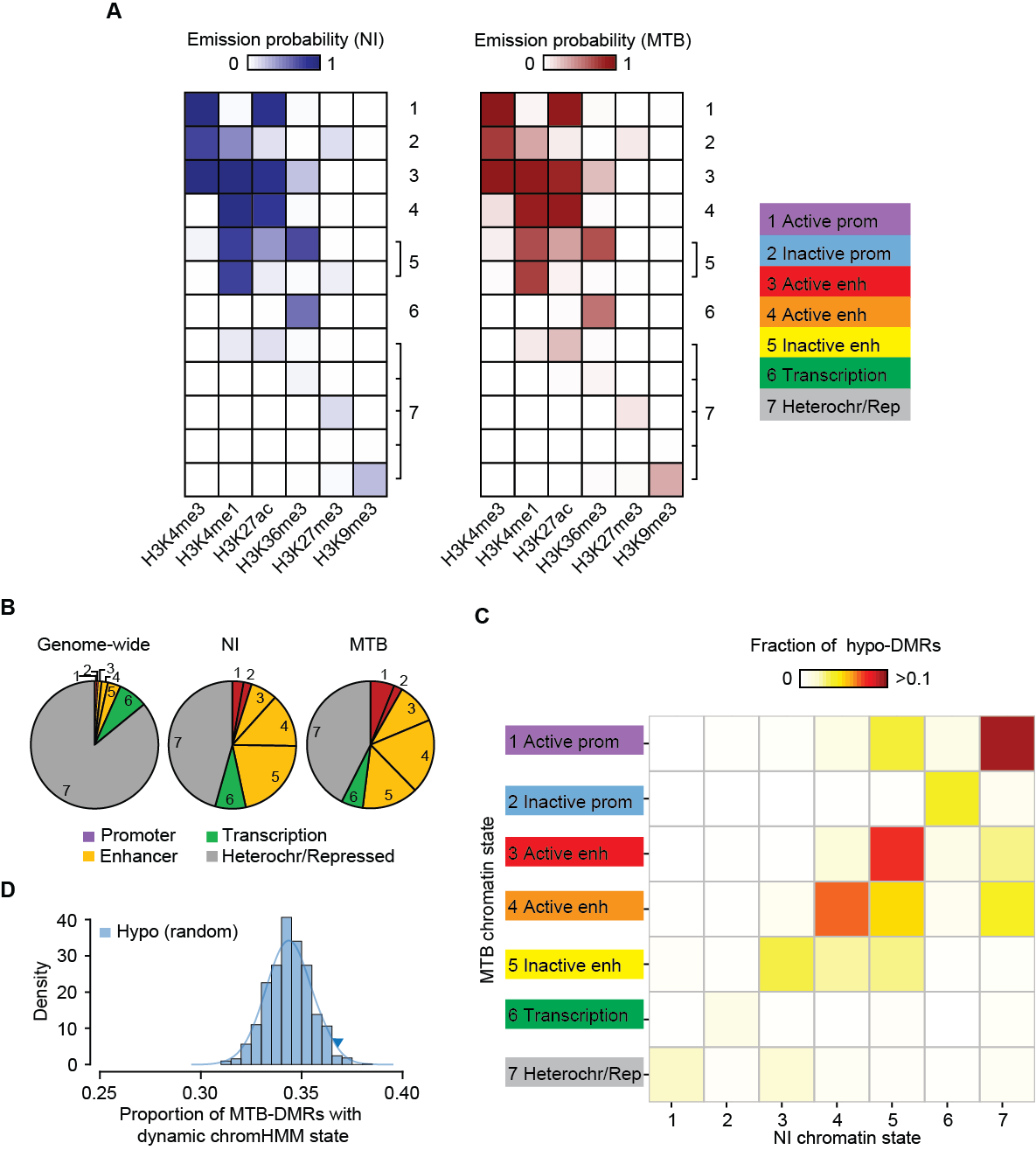
**Supplemental Figure 12.** TF binding alone is not sufficient to induce loss in methylation levels.(*A*) Changes in methylation ±300 bp around sites (25-bp sliding window), inside (blue) and outside (gray) hypomethylated regions, where NF-κB binding events occurred after infection (i.e., posterior probability of <0.05 in non-infected cells and posterior probability of >0.95 in MTB-infected cells) for motifs. Note that the overwhelming majority of MTB-induced NF-κB binding events occur outside hypomethylated regions (n = 51 events inside). To test for changes in methylation within these regions, we generated five control sets of 51 regions by randomly sampling NF-κB binding sites, found outside of hypomethylated regions, with (*B*) matching number of CpG sites and baseline methylation within ±100 bp as those sites found inside. The error bars derived from each of the five sets represent the standard error. (*C*) Results from a logistic regression used to evaluate which genomic features best predict TF binding changes that are associated with changes in DNA methylation. For changes in methylation to occur, TF binding event must occur inside an enhancer element (as defined by ChromHMM) and/or in a region of open chromatin (Tn5 peaks). Thus, in addition to TF binding, other regulatory factors specifically recruited at enhancer elements are required for changes in methylation to occur. LMR stands for lowly methylated region.

****

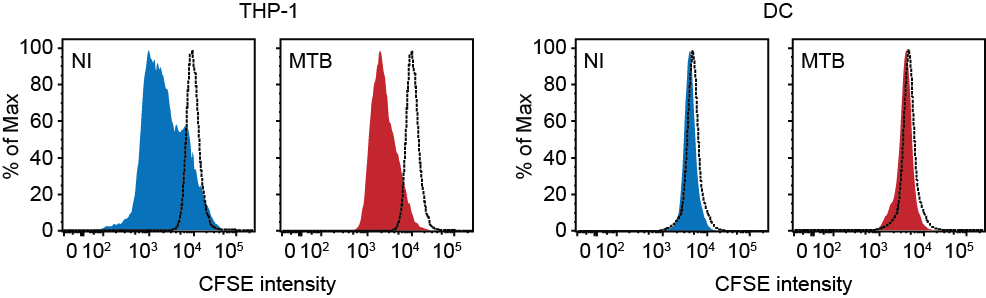
**Supplemental Figure 13.** Transcriptional responses using live MTB and heat-killed bacteria at different ratios.Scatterplots depicting the correlation between the log2 fold changes (log2FC) in gene expression levels using live MTB and heat-killed bacteria at different ratios.

****

**Supplemental Figure 14.** Correlation of ChIP-Seq signals for each histone mark between biological replicates. Smoothed scatterplots showing the correlation of ChIP-Seq signals within 10000-bp sliding windows for each histone mark between 2 biological replicates.

****

**Supplemental Figure 15.** Chromatin state emission probabilities and characterization of dynamic changes in histone marks in MTB-DMRs, based on a unified ChromHMM model. (*A*) ChromHMM emission probabilities for each chromatin state using a unified ChromHMM model that was learned from both infected and non-infected DCs at the same time. (*B*) Pie charts showing the distribution of chromatin state annotations genome-wide and within MTB-DMRs as defined by chromHMM based on ChIP-Seq data for six histone marks. (*C*) Heatmap of the proportion of hypomethylated regions by chromatin transition state. The x-axis represents the chromatin states defined in non-infected DCs and the y-axis the chromatin state of the same region in MTB-infected DCs. (*D*) Histogram showing the observed proportion of hypomethylated regions that change chromatin state after infection (any transition) when sampling 1000 random sets of regions of the genome matched for the chromatin states found in non-infected samples. Each random set contains the same number of hypomethylated regions that we identified in the true data. The blue triangle represents the observed proportion of hypomethylated regions that changed chromatin state in response to MTB infection.



**Supplemental Figure 16.** CFSE-labeled THP-1 (left) and CFSE-labeled DCs (right). Proliferation was assayed in either non-infected cells or cells infected for 48 hours with MTB.

**Supplemental References**

Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. 2012. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol* **13**: R87.

Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD. 2013. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat Protoc* **8**: 1765-1786.

Anders S, Pyl PT, Huber W. 2014. HTSeq - A Python framework to work with high-throughput sequencing data. *Bioinformatics*.

Backes C, Keller A, Kuentzer J, Kneissl B, Comtesse N, Elnakady YA, Muller R, Meese E, Lenhof HP. 2007. GeneTrail--advanced gene set enrichment analysis. *Nucleic Acids Res* **35**: W186-192.

Barreiro LB, Tailleux L, Pai AA, Gicquel B, Marioni JC, Gilad Y. 2012. Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. *Proc Natl Acad Sci U S A* **109**: 1204-1209.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213-1218.

Ernst J, Kellis M. 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**: 215-216.

Hansen KD, Langmead B, Irizarry RA. 2012. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol* **13**: R83.

Hochberg YBaY. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)* **Vol. 57**: 289-300.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**: R36.

Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S et al. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**: 182-187.

Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**: 1571-1572.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.

Moyerbrailean GA, Harvey CT, Kalita CA, Wen X, Luca F, Pique-Regi R. 2014. Are all genetic variants in DNase I sensitivity regions functional? *bioRxiv*.

Pique-Regi R, Degner JF, Pai AA, Gaffney DJ, Gilad Y, Pritchard JK. 2011. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res* **21**: 447-455.

Quah BJ, Parish CR. 2010. The use of carboxyfluorescein diacetate succinimidyl ester (CFSE) to monitor lymphocyte proliferation. *J Vis Exp*.

Santos F, Zakhartchenko V, Stojkovic M, Peters A, Jenuwein T, Wolf E, Reik W, Dean W. 2003. Epigenetic marking correlates with developmental potential in cloned bovine preimplantation embryos. *Curr Biol* **13**: 1116-1121.

Shen L, Shao N, Liu X, Nestler E. 2014. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15**: 284.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**: 1034-1050.

Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D et al. 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**: 490-495.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511-515.

Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B et al. 2012. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**: 1368-1380.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.

Zhou X, Wang T. 2012. Using the Wash U Epigenome Browser to examine genome-wide sequencing data. *Curr Protoc Bioinformatics* **Chapter 10**: Unit10 10.