Université de Montréal

Epigenetic regulation of innate immune responses to infection

par

Alain Pacis

Département de biochimie et médecine moléculaire

Faculté de médecine

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Ph. D.
en Bio-informatique

Mars 2018

© Alain Pacis, 2018

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée :

Epigenetic regulation of innate immune responses to infection

présentée par :

Alain Pacis

a été évaluée par un jury composé des personnes suivantes :

Adrian Serohijos, président-rapporteur

Luis Barreiro, directeur de recherche

Serge McGraw, membre du jury

Guillaume Bourque, examinateur externe

Claude Perreault, représentant de la doyenne de la FES

Résumé

L'importance des modifications épigénétiques sur le contrôle de l'expression génique est clairement établie dans la littérature. Il demeure cependant incertain si les marques épigénétiques modulent l'activité transcriptionnelle ou si ce sont plutôt des conséquences découlant de facteurs régulateurs qui modulent préalablement cette activité. Pour ma thèse, j'ai investigué le rôle de la méthylation de l'ADN dans le contexte de l'activation du système immunitaire inné. Plus précisément, j'ai conduit une analyse intégrant des données de méthylation à l'échelle génomique, de modifications d'histones, d'accessibilité à la chromatine et d'expression génique sur des cellules dendritiques avant et après une infection provoquée par *Mycobacterium tuberculosis* (MTB). Dans le cadre du projet, je montre que la réponse immunitaire à l'infection est associée à la perte de méthylation sur des milliers de sites CpG, indépendamment de la prolifération cellulaire. Les déméthylations actives se trouvent principalement sur des éléments amplificateurs éloignés des sites d'initiation de la transcription et sont fortement associées à l'induction de gènes situés dans leur voisinage. Cependant, une analyse longitudinale indique que la plupart des changements d'expression se produisent avant les changements perceptibles de méthylation. Une analyse de footprint de l'ADN a révélé que le recrutement de facteurs de transcriptions impliqués dans la réponse immunitaire, tel que NF-κB/Rel, précède les pertes de méthylation observées. Il est intéressant de noter que les niveaux de méthylation dans les régions déméthylées ne sont pas rétablis durant l'infection, même pour des gènes dont l'expression retourne à l'état basal. Ces résultats suggèrent que la déméthylation de l'ADN n'est probablement pas cruciale à la mise en place du programme de régulation central enclenché par les cellules du système immunitaire en réponse aux pathogènes. Celle-ci pourrait cependant jouer un rôle dans la mémoire épigénétique et pourrait permettre une réponse plus rapide à une seconde infection. De manière générale, les résultats ouvrent la porte à l'utilisation des régions de méthylation de l'ADN comme bio-marqueur prédictifs d'infections passées et présentes.

Mots-clés: Epigénétique, méthylation de l'ADN, infection, régulation génique

Abstract

The importance of epigenetic modifications in the control of gene expression is widely accepted. Yet, it often remains unclear whether altered epigenetic patterns themselves invoke transcriptional modulation or are instead downstream consequences of regulatory factors. During my thesis, I investigated the role of DNA methylation in the regulation of innate immune responses. Specifically, I performed an integrated analysis of data on genome-wide DNA methylation, histone modifications, chromatin accessibility, and gene expression, in dendritic cells (DCs), before and after infection with *Mycobacterium tuberculosis* (MTB). I demonstrate that the immune response to infection is associated with loss of methylation at thousands of CpG sites, independent of cell proliferation. Active demethylation was specifically targeted to distal enhancer elements and was strongly associated with induction of nearby genes. However, time course analysis further indicates that most changes in gene expression in response to infection occur prior to detectable changes in DNA methylation. Footprinting analysis revealed that the recruitment of immune-related transcription factors, such as NF-κB/Rel, to these regions preceded the observed loss in methylation. Interestingly, levels of methylation at differentially methylated CpG sites never reverted back to higher levels during the course of infection, even among genes for which expression levels return to basal state. Collectively, these results show that DNA demethylation is likely not crucial for the establishment of the core regulatory program engaged by innate immune cells in response to a pathogen. Instead, it might play a role in the establishment of epigenetic memory, which allows for a faster response to a secondary infection. More generally, the results from this thesis opens the door for using DNA methylation marks as a predictive biomarker of past or present infection.

Keywords: Epigenetics, DNA methylation, infection, gene regulation

Table of contents

List of figures

1 Introduction

2 Article I

3 Article II

4 Discussion

List of tables

3 Article II

List of abbreviations

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
AM	Active modification
AR	Active restoration
ATAC-seq	Assay for transposase-accessible chromatin using sequencing
BER	Base excision repair
bp	Base pair
ChIP-seq	Chromatin immunoprecipitation sequencing
cDNA	Complimentary DNA
CGI	CpG island
CpG	Cytosine-phosphate-guanine dinucleotide
CRISPR	Clustered regularly interspaced short palindromic repeats
DC	Dendritic cell
DM	Differentially methylated
DMR	Differentially methylated regions
DNA	Deoxyribonucleic acid
TDG	Thymine DNA glycosylase
DNMT	DNA methyltransferase
eRNA	Enhancer RNA
FDR	False discovery rate

GO	Gene Ontology
H3K4me1	Histone 3 lysine 4 monomethylation
H3K4me3	Histone 3 lysine 4 trimethylation
H3K9me3	Histone 3 lysine 9 trimethylation
H3K27ac	Histone 3 lysine 27 acetylation
H3K27me3	Histone 3 Lysine 27 trimethylation
H3K36me3	Histone 3 Lysine 36 trimethylation
HDAC	Histone deacetylase
kb	Kilobase
mRNA	Messenger RNA
MethylC-seq	Whole genome shotgun bisulfite sequencing
MTB	*Mycobacterium tuberculosis*
PD	Passive dilution
PCR	Polymerase chain reaction
Pyro-seq	Pyrosequencing
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
scRNA-seq	single-cell RNA sequencing
SeqCap Epi	Capture-based bisulfite sequencing
TAB-seq	TET-assisted bisulfite sequencing
TB	Tuberculosis
TET	TET methylcytosine dioxygenase
tRNA	Transfer ribonucleic acid
TF	Transcription factor
TSS	Transcription start site

Acknowledgments

I would like to thank my supervisor, Luis Barreiro, for giving me the opportunity to work on an exciting project and for his invaluable assistance, support and guidance throughout the thesis.

I also wish to acknowledge my collaborators for their excellent experimental contribution and scientific interest.

I am thankful to all past and present members of the Barreiro lab for their help in different parts of the work and for creating a scientifically motivating environment. I am fortunate to be a part of an amazing group who made this period as enjoyable as possible.

1 Introduction

1.1 Epigenetics and chromatin

The term 'epigenetics' was coined by Waddington in 1942 to refer to 'heritable changes in genome function that occur without changes in the DNA sequence' (Waddington 2012). Despite the fact that every cell in a given multicellular organism contains the same genetic information, each cell exhibits different functions and morphologies. Knowing the nucleotide sequence alone is only a small part of the puzzle and the answer lies in the epigenetic regulation of genes. To understand epigenetics requires an understanding of chromatin structure. In eukaryotes, chromatin comprises of DNA wrapped ~147 bp around histone octamers (H2A, H2B, H3 and H4), which constitutes the nucleosome. Chromatin structure can either be loosely packed into euchromatin (open chromatin) or more densely packed into heterochromatin (closed chromatin) (Bell et al. 2011). Epigenetic mechanisms, such as DNA methylation and post-translational modifications of core histone tails, cooperatively determine chromatin configuration and the accessibility of the DNA to the transcription machinery and thus, govern the transcriptional regulation of the expression of genes (Berger 2007; Bernstein et al. 2007) (**Figure 1**).
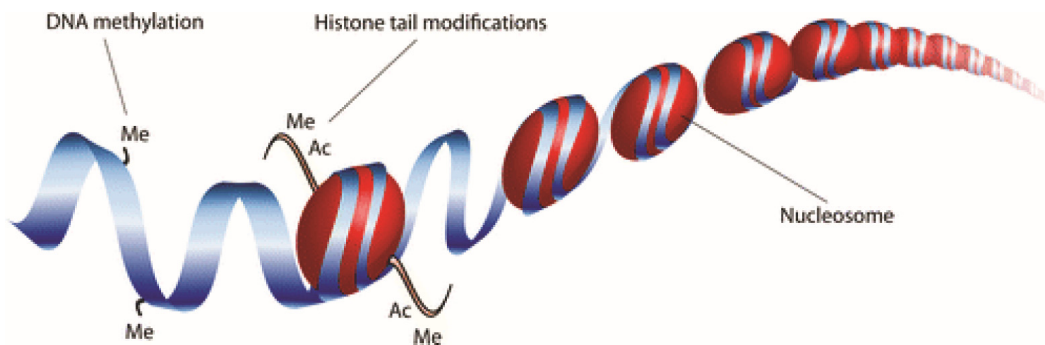
Figure 1. Epigenetic mechanisms of gene regulation. DNA is wrapped around histones and the combined loop of DNA and histone proteins is called a nucleosome. Epigenetic modifications can occur at the histone tails (including acetylation 'Ac' and methylation 'Me') or directly at the DNA (methylation). *Image taken from (Hoeksema and de Winther 2016).*

1.2 Histone modifications

There is a wide range of different modifications in the carboxy- and amino-terminal tails of histone proteins that can alter the structure and function of chromatin, including acetylation, methylation, phosphorylation, ubiquitylation, and SUMOylation (Berger 2007; Kouzarides 2007). The recent development of the chromatin immunoprecipitation technique (ChIP) using modification-specific antibodies and its adaptation sequencing (ChIP-seq) has revolutionized mapping of DNA-protein interactions (Furey 2012). It has provided new insights into the genome-wide distribution of histone modifications in many normal and disease-related processes (Barski et al. 2007; Consortium 2012; Roadmap Epigenomics et al. 2015). Although an in-depth review of histone modifications is outside the scope of this thesis, a few relevant modifications (i.e., H3K4me1, H3K4me3, H3K27me3, H3K36me3, H3K9me3, and H3K27ac) are described in more details below.

Histone methylation is associated with either transcriptional activation or inactivation (Kouzarides 2007). The effect of histone methylation on chromatin state is dependent not only on the specific lysine residue modified, but also on its degree of methylation, with the potential addition of one (me1), two (me2), or three methyl groups (me3). For example, histone H3 lysine 4 trimethylation (H3K4me3) shows increased signals in promoters of genes whereas H3K4me1 are associated with enhancer elements located far away from their target genes. It is also well-documented that H3K9me3 and H3K27me3 are associated with transcriptional repression or heterochromatin formation. In some cases, H3K4me3 and H3K27me3 co-exist as "bivalent domains" in genes that regulate development of stem cells, keeping these key genes in poised states for later activation (Bernstein et al. 2006). In the majority of cases, methylation at the aminoterminal domains of H3 and H4 is catalyzed by members of the SET-domain protein methyltransferase family (Dillon et al. 2005), and its removal is carried out by JmjC-containing lysine demethylases (Shi et al. 2004; Klose et al. 2006; Xiang et al. 2007).

Lysine residues at the N-terminal of histone tails are also subjected to either acetylation by histone acetyltransferase enzymes (HATs), or deacetylation by histone deacetylases (HDACs) (Jenuwein and Allis 2001; Kouzarides 2007). Histone acetylation, including histone 3 lysine 27

acetylation (H3K27ac), is positively correlated with activation in general. Acetylation removes positive charges and reduces the affinity between histones and DNA leading to an open chromatin structure, which is accessible to transcription factors. Co-occurrence H3K27ac and H3K4me1 marks has been widely used to classify active from inactive/primed (H3K4me1 only) enhancers (Creyghton et al. 2010; Calo and Wysocka 2013).

1.3 DNA methylation

In mammals, DNA methylation occurs primarily at the fifth carbon of cytosine residues that are followed by a guanine (CpG dinucleotides). Overall, mammalian genomes are very rich in DNA methylation with the exception of regions called CpG islands – CpG-rich regions found at promoters or near transcription start sites (TSS) of genes (Meissner et al. 2008; Lister et al. 2009). More recently, genome-wide DNA methylation analysis have identified lowly-methylated regions beyond CpG islands, that correspond to distal regulatory elements (enhancers, silencers and insulators) (Stadler et al. 2011). DNA methylation is often associated with transcriptional repression – for example during differentiation, X chromosome inactivation, and imprinting (Suzuki and Bird 2008). Aberrant patterns of DNA methylation can also have striking effects on individual health, including well-known links to cancer susceptibility and autoimmune disorders (Robertson 2005). Thanks to decades of research, many of the proteins and mechanisms involved in DNA methylation have already been identified. Processes that regulate DNA methylation are essentially broken down into three classes: "writers" are the enzymes that catalyze the addition of methyl groups onto cytosine residues, "erasers" modify and remove methyl groups, and "readers" recognize and bind to methyl groups to ultimately mediate changes in gene expression.

1.3.1 Writers

DNA methylation occurs by the addition of a methyl group from S-adenosylmethionine to cytosine with the help of DNA methyltransferases (DNMTs) (Goll and Bestor 2005;

Schermelleh et al. 2005). There are two types of DNA methyltransferase activities in mammals: *de novo* and maintenance methylation, which are achieved by DNMT3 and DNMT1 respectively (**Figure 2**). DNMT1 was the first eukaryotic DNA methyltransferase to be discovered (Bestor et al. 1988). DNMT1 seemed to be responsible only for maintaining methylation after each round DNA replication, which led to the assignment of DNMT1 as a maintenance DNA methyltransferase (Bestor 2000; Goyal et al. 2006). This is supported by findings showing that that DNMT1 co-localizes with the replication machinery (Leonhardt et al. 1992; Schermelleh et al. 2007). At replication sites, hemimethylated DNA is formed when the newly synthesized unmethylated strand pairs with the methylated template strand. Strikingly, while virtually all of methylated CpG sites are methylated in both strands, 98% of methylated cytosines in non-CpG context are highly asymmetrical with only one of strands being methylated (Lister et al. 2009). This suggests that DNMT1 recognizes its substrate cytosine residue only if a guanine residue is beside it.

DNMT3A and DNMT3B are *de novo* methyltransferases that are responsible for establishing cytosine methylation patterns at unmethylated DNA (Okano et al. 1999). Although DNMT3A and DNMT3B show considerable functional redundancy in early developmental stages, they have different expression profiles in distinct cell types. Moreover, DNMT3B appears to be specialized in particular parts of the chromosome as it engages methylation only at the centromeric region (Xu et al. 1999). It has also been proposed that there is a possible cooperation between the *de novo* and the maintenance DNMTs (Siedlecki and Zielenkiewicz 2006). DNMT3A and DNMT3B may also participate in the maintenance of methylation by restoring methylation at cytosine residues which have been overlooked by DNMT1 during replication. There is a third homolog in the DNMT3 family found only in germ cells, called DNMT3L (DNA methyltransferase 3-like). Although this protein has been shown to not possess methyltransferase activity, it is essential as a regulatory cofactor of DNMT3A and DNMT3B (Goll and Bestor 2005).

The family DNMT2 cytosine methyltransferases is the most strongly conserved among all known cytosine methyltransferases and it is ubiquitously expressed in most human and mouse tissues. What makes DNMT2 of enigmatic nature is that it lacks methyltransferase activity and

does not seem to act as a regulatory factor like DNMT3L – mice, flies and plants deficient of DNMT2 do not display any overt phenotype (Goll et al. 2006; Jeltsch et al. 2006). Interestingly, DNMT2 acts as a transfer RNA (tRNA) methyltransferase that specifically catalyzes the methylation of position 38 in tRNA^{Asp}, tRNA^{Gly} and tRNA^{Val} (Goll et al. 2006; Tuorto et al. 2012). Several reports have further demonstrated that DNMT2-mediated methylation contributes to the secondary structure of tRNAs and differential codon usage (Goll et al. 2006; Tuorto et al. 2012; Tuorto et al. 2015; Jeltsch et al. 2017; Zhang et al. 2018), suggesting its role as a modulator of protein translation.
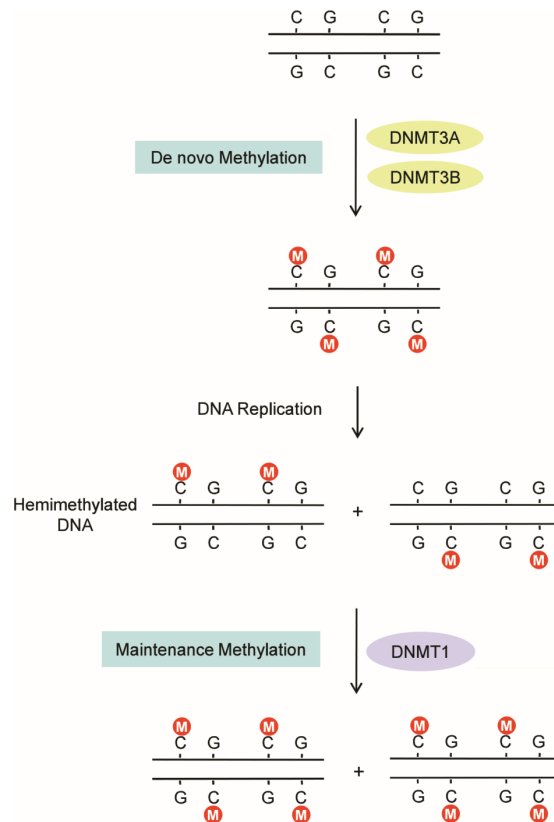
Figure 2. Establishment of DNA methylation patterns in mammals. *De novo* methyltransferases DNMT3A and DNMT3B introduce methyl groups (red circles) to the cytosine of previously unmethylated CpG dinucleotides on both strands. Replication of methylated DNA results in hemimethylated DNA in which the parent strand is methylated while the daughter strand is unmethylated. DNMT1 functions as maintenance methyltransferase by methylating the hemimethylated form of CpG sites. *Image adapted from (Yu et al. 2011).*

1.3.2 Erasers

DNA methylation is relatively stable compared with other epigenetic marks such as histone modifications. Nevertheless, loss of DNA methylation, or DNA demethylation, has been observed in different biological contexts. DNA demethylation is the process of removal of a methyl group from nucleotides in DNA and it may take place in a passive or active fashion. Passive DNA demethylation takes place in dividing cells. As DNMT1 maintains DNA methylation during cell replication, its absence allows newly synthesized DNA strands to be devoid of methylation. Active DNA demethylation occurs via direct removal of a methyl group independently of DNA replication and therefore can take place in both dividing and non-dividing cells. So far, there is no known mechanism in mammalian cells that can cleave the strong covalent carbon-to-carbon bond that connects cytosine to a methyl group. Instead, active demethylation occurs through a series of chemical reactions that revert 5mC back to C. A series of recent discoveries has brought clarity on our understanding of active DNA demethylation.

Until recently, the only known covalent epigenetic modification on DNA was methylation at position 5' of cytosine. A landmark discovery by Tahiliani et al. was made showing that 5mC is oxidized to 5-hydroxymethylcytosine (5hmC) by the enzyme ten-eleven translocation (TET) family proteins (Tahiliani et al. 2009). More importantly, work from the same group have shown that TET proteins and 5hmC may be involved in DNA demethylation – overexpression of TET1 leads to a decrease in 5mC levels. Since the discovery of TET, 5hmC has taken on a new central role in DNA demethylation. Given our current understanding, active demethylation involving TET fall into two groups, which both initially involve active modification (AM) of 5mC to generate 5hmC (Kohli and Zhang 2013) (**Figure 3**). In the process of passive dilution (PD), unmodified C is regenerated through DNA replication since DNMT1 does not recognize 5hmC and therefore cannot maintain it (Inoue et al. 2011; Inoue and Zhang 2011). Alternatively, active restoration (AR) requires further enzymatic modification of 5hmC to regenerate unmodified C. Specifically, TET can further oxidize 5hmC, yielding 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (He et al. 2011; Ito et al. 2011). 5fC or 5caC is subsequently excised by thymine DNA glycosylase (TDG) generating an abasic site as part of the base excision repair (BER) process (Fromme and Verdine 2004). Ideally, AM-AR has the advantage of achieving

rapid conversion of 5mC to unmodified C and therefore seems particularly well suited to locus-specific demethylation processes that require a rapid response to environmental stimuli.
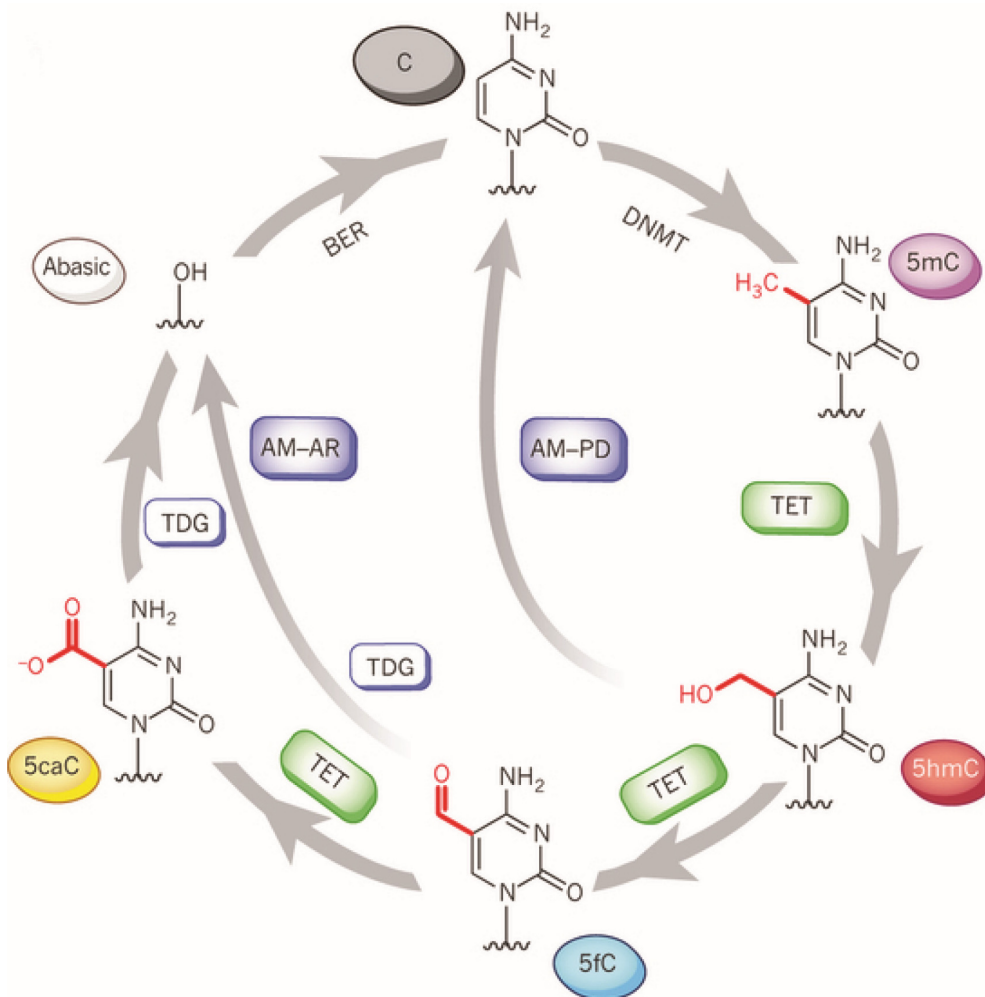
Figure 3. A complete pathway for dynamic cytosine modifications. 5mC bases are introduced by DNA methyltransferase (DNMT) enzymes. 5mC is further oxidized to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) during active DNA demethylation by TET family proteins. In the pathway of active modification followed by passive dilution (AM-PD), 5hmC is diluted in a replication-dependent manner to regenerate unmodified C. In the pathway of AM followed by active restoration (AM-AR), 5hmC is further oxidized by TET proteins to 5fC and then 5caC, which is then excised by TDG and repaired by the BER pathway into an unmodified cytosine, generating an abasic site. *Image taken from (Kohli and Zhang 2013).*

1.3.3 Readers

There are potentially multiple ways in which DNA methylation can decrease transcription levels or completely turn off genes. One simple way this is accomplished is by directly preventing the binding of the transcriptional machinery or transcription factors to their respective regions by the methyl groups themselves (Hark et al. 2000; Jaenisch and Bird 2003)**.** Alternatively, repression can be achieved via proteins that specifically recognize and bind 5mC. The identification of proteins that can read the methylation signal has shed light on how DNA methylation play a repressive role in gene expression. The first characterized methyl-binding proteins are the methyl-CpG binding domain (MBD) family which consists of five members, namely MBD1, MBD2, MBD3, MBD4 and MeCP2 (methyl-CpG-binding protein 2). These proteins each contain a conserved MBD domain that confers a high affinity for methylated CpG sites (Zhang et al. 1990; Nan et al. 1993; Hendrich and Bird 1998; Wade 2001; Hendrich and Tweedie 2003; Jaenisch and Bird 2003). Three of these proteins (MeCP2, MBD1 and MBD2) also contain a transcriptional repression domain (TRD) that allows these proteins to interact with a variety of repressor complexes and participate in methylation-dependent repression of transcription (Meehan et al. 1989; Nan et al. 1993) (**Figure 4**). The second family of methyl-binding proteins binds to methylated DNA via a zinc-finger domain and is composed of Kaiso, ZBTB4, and ZBTB38 (Prokhortchouk et al. 2001; Filion et al. 2006). Finally, there is the family of UHRF (ubiquitin-like, containing PHD and RING finger domain) proteins that includes, UHRF1 and UHRF2. These are multidomain proteins that bind methylated cytosines via a SET- and RING-associated DNA-binding domain (Hashimoto et al. 2008; Hashimoto et al. 2009).
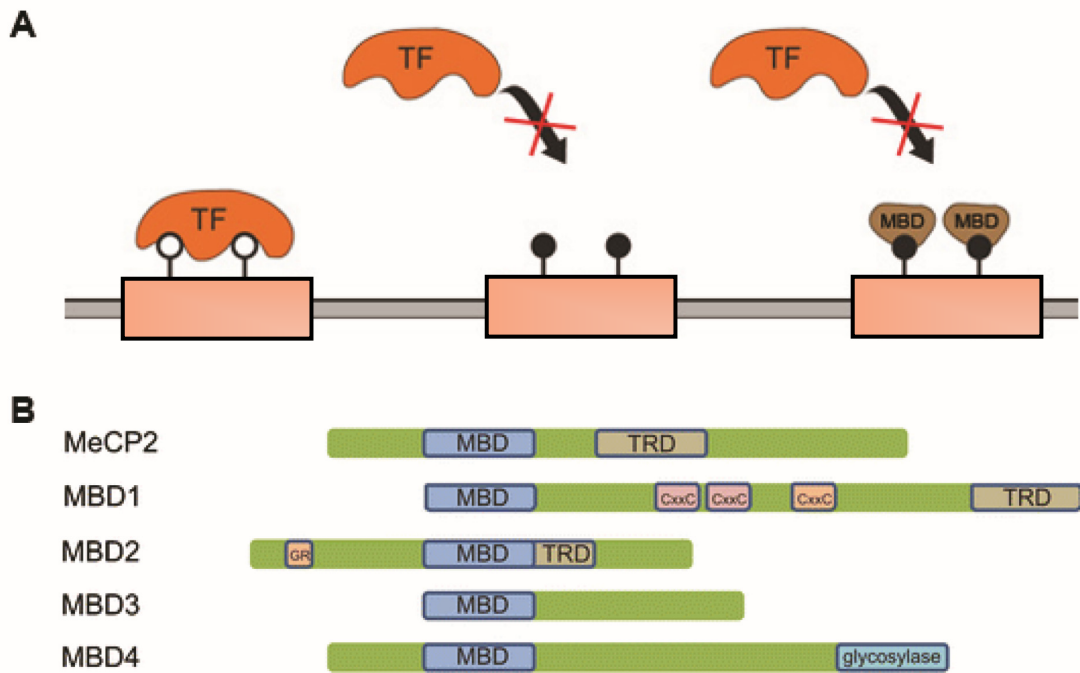
Figure 4. Readers of methylation signal and their potential mechanism in gene repression. (**A**) (Left to right) Transcription factor (TF) is able to bind to unmethylated DNA sequence (white circles); methylated CpG sites (black circles) or methyl-CpG-binding domain (MBD) proteins interfere with binding of TF. *Image adapted from (Trzyna et al. 2012).* (**B**) Graphical representation of the domains in the proteins that recognize and bind specifically methylcytosines. *Image taken from (Bogdanovic and Veenstra 2009).*

1.4 Mapping the epigenome

1.4.1 ChIP-seq

The increasing interest in the role of epigenetic processes in development and disease has coincided with the advancement of new methods to conduct large-scale and high-resolution epigenomic profiling. The most commonly used experimental approach to profile histone posttranslational modifications is chromatin immunoprecipitation sequencing (ChIP-seq) (Furey 2012). This method uses an antibody against a specific histone modification to immunoprecipitate chromatin regions bearing the corresponding modification. The key bioinformatics challenge in the analysis of ChIP-seq data is to accurately map thousands to millions of reads, corresponding to these regions, to the reference genome (Park 2009). Many sequence aligners for solving the problems of mapping sequence reads have been developed, such as Bowtie (Langmead et al. 2009) and BWA (Li and Durbin 2009). ChIP-seq is also a widely used approach to selectively enrich for DNA sequences bound by specific transcription factors, allowing for the generation of genome-wide binding site maps.

1.4.2 BS-seq

Although there are many DNA methylation analysis methods, bisulfite sequencing (BS-seq) is considered to be the gold standard method in DNA methylation studies (Harris et al. 2010; Laird 2010; Bock 2012). The direct examination of DNA methylation is hindered by the fact, that DNA methylation cannot be analyzed by standard DNA sequencing methods, since they are unable to distinguish 5-methylcytosine from unmodified cytosine. To overcome this, genomic DNA are first treated with sodium bisulfite prior to sequencing (Clark et al. 1994; Clark et al. 2006). Under appropriate conditions, sodium bisulfite causes the specific deamination of cytosine through a sulfonated intermediate, and its conversion to uracil. While unmethylated cytosine residues are converted to uracil, methylated cytosine residues are protected from this conversion and therefore remains intact. PCR amplification of converted DNA replaces the

uracil with thymine (**Figure 5**). Initially, bisulfite sequencing was used to assay individual loci with locus-specific PCR followed by Sanger sequencing. Recently, reduced representation bisulfite sequencing has extended the genomic coverage of bisulfite sequencing by using high-throughput sequencing technology. Reduced representation bisulfite sequencing (RRBS) combines restriction digestion with BS for analysis of high CpG density regions. Finally, whole-genome bisulfite sequencing (WGBS or MethylC-seq) provides single-base resolution and quantitative rates of methylation for all of the ~29 million CpG sites in the human genome. Mapping of BS-seq reads is performed using as a specific alignment algorithm that implements a bisulfite converted reference genome (C-to-T and a G-to-A) (Krueger and Andrews 2011).

Bisulfite treatment in combination with specially designed genotyping microarrays makes it possible to measure DNA methylation levels at a preselected fraction of CpG sites throughout the genome. The Illumina 450k Infinium methylation microarray, which contains over 450,000 CpG sites covering the majority of CpG islands, gene promoters and some enhancer regions, is the most commonly used array in human methylation research (Bibikova et al. 2011; Sandoval et al. 2011). Moreover, the recently developed MethylationEPIC 850k Infinium methylation microarray includes an additional 413,745 CpG positions that are enriched in human enhancer regions (Moran et al. 2016) provided by the ENCODE (Consortium 2012) and FANTOM5 (Lizio et al. 2015) consortia.

A potential issue with current bisulfite conversion-based methodologies is that they depend on the complete conversion of unmethylated cytosines (Wreczycka et al. 2017). Another limitation is that bisulfite conversion does not distinguish between 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC), the first derivative in the active DNA demethylation pathway (Huang et al. 2010; Jin et al. 2010) (**Figure 5**). Tet-assisted bisulfite sequencing (TAB-seq) overcomes this limitation and allows single-CpG resolution mapping of 5hmC (Yu et al. 2012).
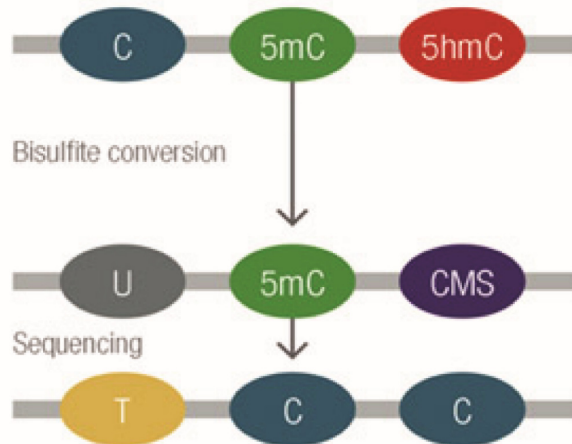
Figure 5. Detection of methylated DNA by bisulfite conversion. The methyl group covalently attached to the 5'position of cytosine protects against bisulfite conversion. Unmethylated cystosine (C) is converted to uracil (U; read as thymine (T) when sequenced), but not 5-methylcytosine (5mC). Bisulfite treatment converts 5-hydroxymethylcytosine (5hmC) to cytosine-5-methylenesulfonate (CMS), leaving both 5mC and 5hmC to be detected as C. *Image adapted from (Yu et al. 2012).*

1.4.3 ATAC-seq

The chromatin accessibility of genomic regions can be profiled with methodologies such as DNase I hypersensitive site sequencing (DNase-seq) (Song and Crawford 2010), formaldehyde-assisted isolation of regulatory elements followed by sequencing (FAIRE-seq) (Giresi et al. 2007), and assay for transposase-accessible chromatin sequencing (ATAC-seq) (Buenrostro et al. 2013). ATAC-seq is now becoming increasingly popular owing to its simple workflow involving substantially fewer cells as starting material. The procedure relies on a hyperactive Tn5 transposase to simultaneously fragment and insert sequencing adapters into nucleosome-free regions in the genome. In addition to its ability to capture open chromatin regions, ATAC-seq also allows the prediction of binding sites for hundreds of transcription factors simultaneously, by footprinting (Neph et al. 2012; Buenrostro et al. 2013). Ultimately, integration of multiple epigenome profiles (along with DNA sequence variation and gene expression data) will maximize the potential of these data to infer function.

1.5 Epigenetic regulation and innate immunity

1.5.1 Innate immunity

Classically, innate immunity is characterized by a rapid, nonspecific response to an invading pathogen (Medzhitov and Janeway 1997). Conversely, an adaptive immune response requires more time to mount, is highly specific to the invading pathogen, and forms memory cells that will respond faster and more robustly to a secondary challenge against an identical immune assault (Medzhitov and Janeway 1998). Understanding innate immune responses is fundamental as it provides the first line of defense against immune challenges (Medzhitov and Janeway 2000) and plays an important role in the activation of the adaptive system (Iwasaki and Medzhitov 2015). Innate immune cells, such as neutrophils, monocytes, macrophages and dendritic cells (DCs), are equipped with various pattern recognition receptors (PRRs), which recognize a wide array of conserved pathogen-associated molecular patterns (PAMPs) and discriminate between self and non-self molecules. Recognition of immune stimuli activates downstream molecular signaling pathways that culminate in the induction of sophisticated transcriptional programs involving the regulation of thousands of genes, which are coordinated with the help of signal-dependent transcription factors including NF-κB/Rel, AP-1, and interferon regulatory factors (IRFs) (Medzhitov 2001; Medzhitov and Horng 2009; Smale 2010; Smale 2011). Upon activation, these transcription factors bind to gene regulatory regions – promoters, enhancers, or silencers – where they function to initiate recruitment of various co-factors required for the activation of inflammatory and/or antiviral response signals. This cascade starts the process of pathogen clearance and the subsequent initiation of appropriate adaptive immune responses.

1.5.2 Dendritic cells: Linking innate and adaptive immunity

Dendritic cells (DCs) are professional antigen-presenting cells that have a central role in T cell activation and in initiation of adaptive immune responses. DCs express a number of pattern

recognition receptors, including Toll-like receptors (TLRs), Nod-like receptors (NLR) and RIG-I-like receptors (RLR), which recognize a wide array of pathogen-associated molecular patterns (Kapsenberg 2003). As immature cells specialized in antigen uptake and processing, DCs reside in and traffic through non-lymphoid peripheral tissues, continuously surveying the environment for invading microorganisms. As immature DCs capture antigens by endocytosis/phagocytosis, they undergo major changes in gene expression programs, evolving from immature, antigen-capturing cells to mature, antigen-presenting, T cell-priming cells. This process of DC maturation, in general, involves down-regulation of antigen internalization, a redistribution of major histocompatibility complex (MHC) molecules from intracellular endocytic compartments to the DC surface, an increase in the surface expression of costimulatory molecules, secretion of chemokines and cytokines, and surface expression of adhesion molecules and chemokine receptors (Tan and O'Neill 2005) (**Figure 6**).

To present foreign antigens to naïve T cells, DCs must migrate from inflamed or injured peripheral tissues to the closest draining lymph nodes through afferent lymphatic vessels. Migration of maturing DCs from the periphery into lymphoid tissues are coordinated by chemokines that interact with corresponding receptors on DCs (Alvarez et al. 2008). For example, immature DCs express CC-chemokine receptor 1 (CCR1), CCR2, CCR5 and CXC-chemokine receptor 1 (CXCR1) and are attracted to non-lymphoid tissues by their respective ligands, which are expressed constitutively or at inflammatory sites. DC maturation results in the downregulation of expression of these chemokine receptors and the upregulation of CCR7 expression. Expression of CCR7 switches DC responsiveness to its ligands, CC-chemokine ligand 19 (CCL19) and CCL21, that guide migration to secondary lymphoid organs. Maturation of DCs also induces the production of CCL22, CCL17 (i.e., chemokines that attract CCR4-expressing T cells), and CCL18. DC production of the chemokine CXCL16, in T cell-rich areas of lymphoid organs, may also function in promoting interaction between DCs and cytotoxic T cells.

DCs are capable of processing antigens and present peptide in the context of either MHC class I or II molecules, which interact with and stimulate cytotoxic T lymphocytes and T helper cells, respectively. As DCs mature, they acquire the properties necessary to form and transport

peptide-loaded MHC complexes to the cell surface. Antigen transport to the cell surface coincides with increased expression of costimulatory molecules, such as CD40, CD80 and CD86 (Tan and O'Neill 2005). These molecules amplify T cell receptor signaling and promote T cell activation. Moreover, the soluble cytokine profile secreted by DCs varies with the different stages of DC development and maturation thus influencing the different effector functions characteristic of immature vs. mature DCs (de Saint-Vis et al. 1998). A wide variety of cytokines may be expressed by mature DCs and the exact cytokine repertoire expressed will depend on the nature of the stimulus, maturation stage of the DC and the existing cytokine microenvironment. The distinct cytokine patterns released by mature DCs contribute to the commitment of naïve T cells into more specialized T cell subsets. For example, antigens that prime DCs to secrete IL-12 will typically induce Th1 differentiation (Heufler et al. 1996; Kalinski et al. 1999). Another example is the production of IL-10 and TGF-β, which leads to the generation of regulatory T (Treg) cells (Kushwah and Hu 2011).
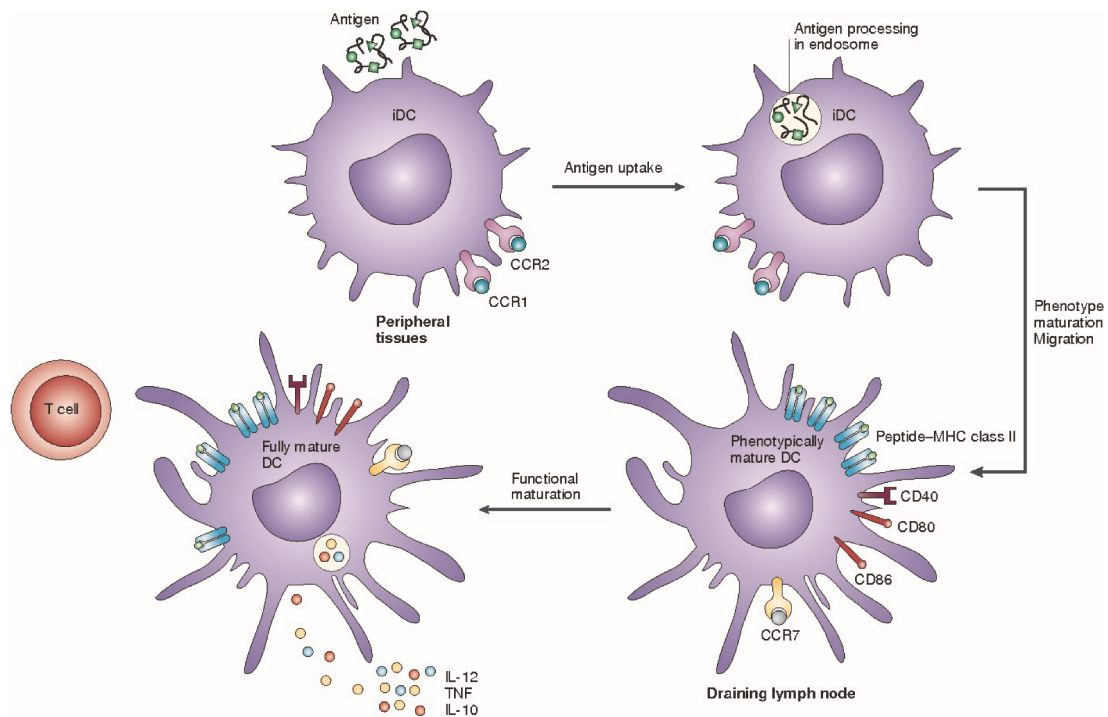
Figure 6. Dendritic cell maturation upon antigen encounter. Maturation of dendritic cells, in response to antigen, leads to the redistribution of MHC class II molecules and MHC-peptide

complexes from within the endocytic system to the cell surface, the production of several cytokines and membrane associated T cell stimulatory molecules, and the remodeling of expressed chemokine receptors. These changes allow dendritic cells to migrate to draining lymph nodes and induce antigen-specific immune response by activating T cells. *Image adapted from (Hackstein and Thomson 2004).*

1.5.3 Epigenetic control of the innate immune response

To elaborate an appropriate response to the threat, innate immune cells also undergo important epigenetic regulation. In response to immune stimuli, the most labile epigenetic changes involve the post-translational modifications of histone tails at promoter and enhancer regions (Hazzalin and Mahadevan 2005; Monticelli and Natoli 2013). At promoter regions, histone acetylation has been shown to be essential for the activation of many pro-inflammatory genes (Schmeck et al. 2005; Qiao et al. 2013). Genome-wide analysis, using ChIP-seq, have revealed that enhancer elements (marked by H3K4me1), also contribute to modulation of immune responses (Heintzman et al. 2007; Heintzman et al. 2009; Barish et al. 2010; Ghisletti et al. 2010; Garber et al. 2012; Calo and Wysocka 2013; Kaikkonen et al. 2013; Ostuni et al. 2013; Rogatsky and Adelman 2014). The canonical model is that binding of signal-dependent transcription factors (TFs) in response to stimuli occurs within cell type-specific repertoires of enhancers, that were already established by lineage-determining or "pioneer" TFs during cell differentiation (Barish et al. 2010; Ghisletti et al. 2010; Heinz et al. 2010; John et al. 2011; Mullen et al. 2011; Natoli et al. 2011; Trompouki et al. 2011; Garber et al. 2012; Kaikkonen et al. 2013; Ostuni et al. 2013; Rogatsky and Adelman 2014) (**Figure 7**). Interestingly, Ostuni et al. have reported a new class of distal regulatory elements, coined "latent enhancers" that appear after stimulation of mouse macrophages with different stimuli (Ostuni et al. 2013). These enhancers are inactive and unmarked at basal state but gain *de novo* H3K4me1 and H3K27ac marks during first encounter with a stimulus. Interestingly, H3K4me1 persists for a longer period of time even when stimulus has ceased and contributes to a faster and stronger transcriptional response of nearby genes to a secondary stimulus.

In contrast to histone modifications, little is known about the regulatory implications of DNA methylation in innate immune responses. DNA methylation has historically been considered to be a relatively stable epigenetic mark (Bernstein et al. 2007), and thus unlikely to respond to environmental perturbations on a short time scale. Despite its thermodynamic stability, there is increasing evidence that DNA methylation can rapidly respond to environmental perturbations, as exemplified in post-mitotic cells in the brain during neuronal activation (Weaver et al. 2004; Guo et al. 2011); or during monocyte differentiation into macrophages or dendritic cells (Klug et al. 2010). Likewise, studies on dividing cells or cell lines also argue for the involvement of an active enzymatic mechanism, as the kinetics of the demethylation procedure are too fast to be dependent on cell proliferation (Bruniquel and Schwartz 2003; Murayama et al. 2006; Niehrs and Schafer 2012). Indeed, the recent description of transitions from 5-methylcytosine (5mC) into more labile oxidized intermediates – such as 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) – by TET enzymes (Tahiliani et al. 2009) could provide a suitable mechanism to support rapid response genes.

Recent studies have reported altered DNA methylation patterns associated with activation of innate immune cells.(Marr et al. 2014; Zhang et al. 2014; Cizmeci et al. 2016; Wiencke et al. 2016). For instance, Marr et al. assessed epigenetic changes in macrophage DNA methylation in response to infection with an intracellular protozoa *Leishmania donovani* (Marr et al. 2014). Using 450K methylation array, they identified a set of 443 CpG sites with changes in methylation following live *L. donovani* infection. These epigenetic changes are linked to genes that play a critical role in host defense such as the JAK/STAT and Notch signaling pathway. Similarly, Sinclair et al. investigated DNA methylation dynamics in *Anaplasma phagocytophilum*-infected human neutrophils using methylated DNA binding domain (MBD) enrichment and next generation sequencing approach (MBD-seq) (Sinclair et al. 2015). Within 24 hours post-infection, marked increases in DNA methylation were observed genome-wide as compared with mock-infected controls. These studies, however, have focused exclusively on DNA methylation changes in promoter regions or at relatively few CpG sites in the genome at low-resolution, and that such changes are poorly predictive of changes in gene expression levels.
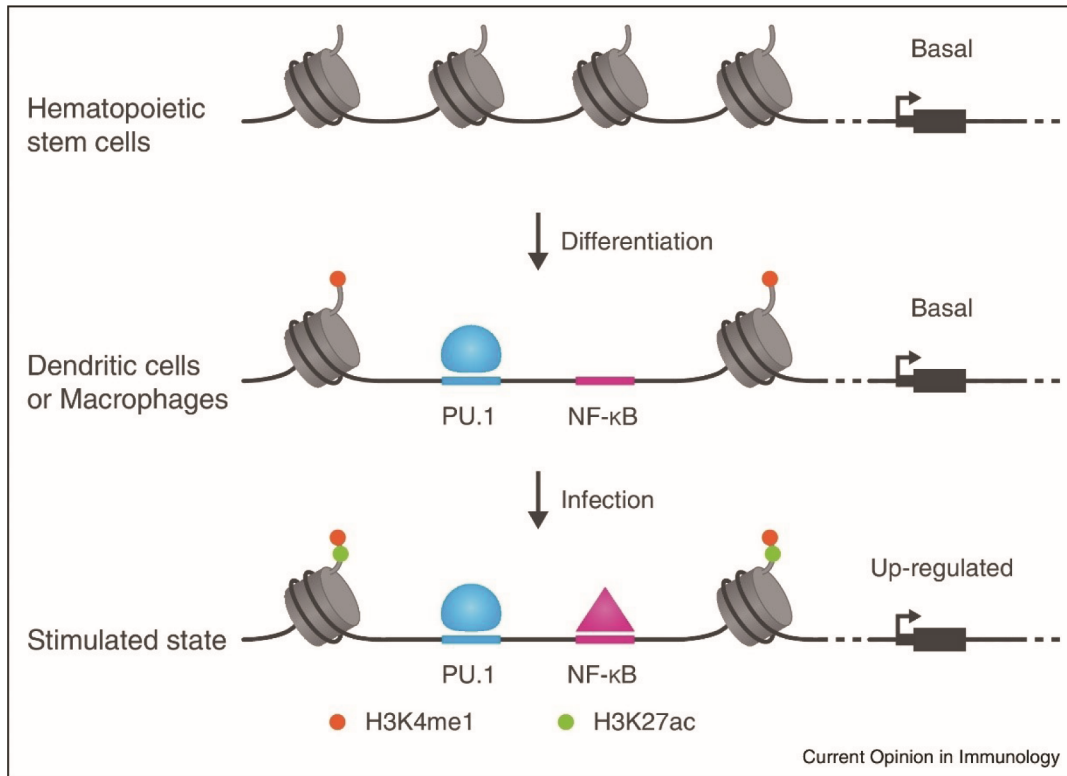
Figure 7. Pioneer transcription factors (TFs) organize the enhancer landscape required for stimulus-induced transcription in innate immune cells. During differentiation of hematopoietic stem cells to macrophages or DCs, lineage-determining or pioneer TFs, such as PU.1, open condensed chromatin and promote the deposition of H3K4me1 at enhancers. Immune stimuli (e.g., infection) trigger the recruitment of signal-dependent TFs, such as NF-κB within the cell type-specific enhancer repertoires already established by PU.1 prior to immune stimulation (i.e., primed enhancers). NF-κB binding leads to the deposition of the H3K27ac activating mark and the subsequent upregulation of stimulus-responsive genes. *Image taken from (Pacis et al. 2014).*

1.6 Research objectives

The main goal of this thesis was to elucidate the role of DNA methylation in the regulation of innate immune responses to infection. This work is divided into two chapters addressing different questions pertaining to the overarching goal.

Article I. Bacterial infection remodels the DNA methylation landscape of human dendritic cells

The first aim was to characterize changes in DNA methylation in innate immune cells during infection. Using high-throughput sequencing methods, I performed comprehensive transcriptional and epigenetic profiling of human dendritic cells, before and after infection with a pathogenic strain of *Mycobacterium tuberculosis* (MTB). A Carboxyfluorescein Diacetate Succinimidyl Ester (CFSE) proliferation assay was also performed to determine whether changes in methylation were independent of cell division.

Article II. DNA demethylation plays a limited role in the regulation of innate immune responses to infection

I next sought to assess the causal relationship between changes in DNA methylation and gene expression in response to infection. I generated paired data on DNA methylation, gene expression, and chromatin accessibility in non-infected and MTB-infected DCs at multiple time-points. These time-series datasets allowed the dissection of the relative order of regulatory events during infection.

2 Article I

Bacterial infection remodels the DNA methylation landscape of human dendritic cells

Pacis A, Tailleux L, Morin AM, Lambourne J, MacIsaac JL, Yotova V, Dumaine A, Danckaert A, Luca F, Grenier JC, Hansen KD, Gicquel B, Yu M, Pai A, He C, Tung J, Pastinen T, Kobor MS, Pique-Regi R, Gilad Y, Barreiro LB.

N.B. Please note that due to space limitations, this section does not include supplementary tables 1 to 8. Complete tables are available online within the supplementary information of the corresponding publication (https://genome.cshlp.org/content/25/12/1801/suppl/DC1).

Bacterial infection remodels the DNA methylation landscape of human dendritic cells

Alain Pacis[1,2], Ludovic Tailleux[3], Alexander M Morin[4], John Lambourne[5], Julia L MacIsaac[4], Vania Yotova[1], Anne Dumaine[1], Anne Danckaert[6], Francesca Luca[7], Jean-Christophe Grenier[1], Kasper D Hansen[8], Brigitte Gicquel[3], Miao Yu[9], Athma Pai[10], Chuan He[9], Jenny Tung[11], Tomi Pastinen[5], Michael S Kobor[4], Roger Pique-Regi[7], Yoav Gilad[12,*], Luis B Barreiro[1,13,*]

[1]CHU Sainte-Justine Research Center, Department of Genetics, Montreal, H3T1C5, Canada; [2]University of Montreal, Department of Biochemistry, Montreal, H3T1J4, Canada; [3]Institut Pasteur, Mycobacterial Genetics Unit, Paris, 75015, France; [4]Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, Department of Medical Genetics, University of British Columbia; [5]Génome Québec Innovation Centre, Department of Human Genetics, McGill University, Montréal, H3A0G1, Canada; [6]Institut Pasteur, Imagopole, Paris, 75015, France; [7]Wayne State University, Center for Molecular Medicine and Genetics and Department of Obstetrics and Gynecology, Detroit, MI, 48202; [8]Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics and McKusick-Nathans Institute for Genetic Medicine, Baltimore, MD, 21205; [9]University of Chicago, Department of Chemistry and Institute for Biophysical Dynamics, Chicago, IL, 60637; [10]Department of Biology, Massachusetts Institute of Technology, United States; [11]Duke University, Departments of Evolutionary Anthropology and Biology and Duke Population Research Institute, Durham, NC, USA 27708; [12]University of Chicago, Department of Human Genetics, Chicago, IL, 60637; [13]University of Montreal, Department of Pediatrics, Montreal, H3T1J4, Canada.

*Correspondence to: luis.barreiro@umontreal.ca or to gilad@uchicago.edu

Abstract

DNA methylation is an epigenetic mark thought to be robust to environmental perturbations on a short time scale. Here, we challenge that view by demonstrating that the infection of human dendritic cells (DCs) with live pathogenic bacteria is associated with rapid and active demethylation at thousands of loci, independent of cell division. We performed an integrated analysis of data on genome-wide DNA methylation, histone mark patterns, chromatin accessibility, and gene expression, before and after infection. We found that infection-induced demethylation rarely occurs at promoter regions and instead localizes to distal enhancer elements, including those that regulate the activation of key immune transcription factors. Active demethylation is associated with extensive epigenetic remodeling, including the gain of histone activation marks and increased chromatin accessibility, and is strongly predictive of changes in the expression levels of nearby genes. Collectively, our observations show that active, rapid changes in DNA methylation in enhancers play a previously unappreciated role in regulating the transcriptional response to infection, even in non-proliferating cells.

Introduction

The first immune mechanisms recruited to defend against invading pathogens are those associated with innate immune cells, such as dendritic cells (DCs) or macrophages. Once they sense an intruder, these cells induce sophisticated transcriptional programs involving the regulation of thousands of genes, which are coordinated with the help of signal-dependent transcription factors, including NF-κB/Rel, AP-1, and interferon regulatory factors (IRFs) (Medzhitov 2001; Smale 2010). The regulation of this program is achieved through a series of epigenetic changes, which are thought to modulate the access of transcription factors to specific DNA regulatory elements (Bierne et al. 2012).

The most well-studied epigenetic responses to immune stimuli involve the post-translational modification of histone tails at promoter and enhancer regions (Bierne et al. 2012; Monticelli and Natoli 2013). Histone acetylation has been shown to be essential for the activation of many pro-inflammatory genes (Ghisletti et al. 2010; Qiao et al. 2013), whereas increased activity of histone deacetylases is often associated with gene repression in the context of inflammation (Villagra et al. 2009). Moreover, recent studies suggest that the response of innate cells to different immune challenges can result in the appearance of histone marks associated with *de novo* enhancer elements (or latent enhancers) (Kaikkonen et al. 2013; Ostuni et al. 2013). These *de novo* enhancers have been postulated to contribute to a faster and stronger transcriptional response to a secondary stimulus (Ostuni et al. 2013).

In contrast, we still know remarkably little about the role of other epigenetic changes in controlling responses to infection. DNA methylation has been particularly understudied, as a consequence of the belief that methylation marks are highly stable, and unlikely to respond to environmental perturbations on a short time scale (Bierne et al. 2012; Monticelli and Natoli 2013). Recent work, however, suggests that DNA methylation patterns can rapidly change in response to certain environmental cues (Klug et al. 2010; Guo et al. 2011; Dowen et al. 2012; Marr et al. 2014), raising the possibility that rapid changes in DNA methylation might play a role in innate immune responses. To date, no studies have comprehensively investigated the contribution of rapid, active changes in methylation (in contrast to passive changes during cell

replication) to the regulatory programs induced by innate immune cells in response to an infectious agent. More broadly, the few studies in mammalian cells that demonstrate cell division-independent changes in DNA methylation have only focused on a small number of CpG sites and, surprisingly, have suggested that such changes are poorly predictive of changes in gene expression levels (Bruniquel and Schwartz 2003; Klug et al. 2010; Guo et al. 2011; Marr et al. 2014). Here, we report the first comprehensive epigenome and transcriptome analysis of monocyte-derived DCs – professional antigen presenting cells that play a central role in bridging innate and adaptive immunity – before and after *in vitro* infection with live pathogenic bacteria. All the data generated in this study are freely accessible via a custom web-based browser that enables easy querying and visualization of epigenetic profiles at any genomic region of interest (http://luis-barreirolab.org/EpigenomeBrowser).

Results

MTB infection induces active changes in DNA methylation in human DCs

We infected monocyte-derived DCs from six healthy donors with a live virulent strain of *Mycobacterium tuberculosis* (MTB), the causative agent of tuberculosis (TB) in humans. Monocyte-derived DCs are ideally suited to study active changes in methylation because they are post-mitotic and not expected to proliferate in response to infection (Pickl et al. 1996; Ardeshna et al. 2000). To experimentally confirm this assumption, we performed a Carboxyfluorescein Diacetate Succinimidyl Ester (CFSE) proliferation assay. This method relies on the ability of the highly fluorescent dye carboxyfluorescein to incorporate within cells. Following each cell division, the equal distribution of these fluorescent molecules to progeny cells results in a halving of per-cell fluorescence levels. We did not detect any decrease in per-cell fluorescence at 18 hours post-infection, which confirms that DCs do not proliferate after MTB infection **(Figure 1A)**. In contrast, we observed high rates of proliferation in our positive control, human monocytic THP-1 cells (**Figure 1A**).

At 18 hours after infection, we obtained paired data on single base-pair resolution DNA methylation levels (using whole genome shotgun bisulfite sequencing: i.e., MethylC-seq) and genome-wide gene expression data (using mRNA sequencing: i.e., mRNA-seq) in non-infected and MTB-infected DCs. For MethylC-seq data, we generated 8.6 billion single-end reads (mean of 648 ± 110 SD million reads per sample; **Supplementary Table 1**) resulting in an average coverage per CpG site of ~9X for each sample. We detected an average of 24 million CpG sites in each sample, corresponding to over 80% of CpG sites in the human genome. Genome-wide methylation data between biological replicates were strongly correlated, attesting to the high quality of the data (**Supplementary Figure 1**; mean r across all samples = 0.86).

As expected for mammalian cells, most CpG sites were highly methylated throughout the genome except near transcription start sites (TSSs), CpG islands, and putative enhancer elements (**Supplementary Figure 2A,B**). We found a significant negative correlation between gene expression levels and methylation levels around TSSs (r = -0.39; $P < 1 \times 10^{-16}$; **Supplementary Figure 2C,D**), highlighting the well-established role of proximal methylation

in the stable silencing of gene expression. Principal component analysis of our data along with MethylC-seq data from 21 other purified cell types and tissues revealed that the DC methylome is closely related to that of other blood-derived cells, particularly cells that share a common myeloid progenitor with DCs, such as neutrophils (**Supplementary Figure 2E**).

We next assessed the occurrence and the extent to which the response of DCs to a bacterial infection is accompanied by active changes in DNA methylation, using the BSmooth algorithm (Hansen et al. 2012). We defined MTB-induced differentially methylated regions (MTB-DMRs) as regions of 3 or more consecutive CpG sites exhibiting a significant difference in methylation between the two groups ($P < 0.01$) and an absolute mean methylation difference above 0.1 (Hansen et al. 2014). Using these criteria, we identified 3,271 MTB-DMRs, corresponding to both hypermethylated regions (48%) and hypomethylated regions (52%) (Figure 1B; Supplementary Table 2). To independently validate these changes, we generated methylation-sensitive pyrosequencing data on control versus MTB-infected DCs from 5 completely new individuals. We targeted 21 CpG sites that were differentially methylated in the MethylC-seq analysis, distributed across 4 hypermethylated (11 CpG sites) and 6 hypomethylated MTB-DMRs (10 CpG sites; **Supplementary Table 3**). We were able to validate 100% of the hypomethylated CpG sites, with effect sizes similar to or greater than those identified in the original bisulfite sequencing analysis (**Figure 1B,C**; **Supplementary Figure 3A**). In contrast, we were not able to validate any of the hypermethylated CpG sites (**Supplementary Figure 3B**), which indicates that most (if not all) active changes in methylation observed in response to infection are losses rather than gains in methylation, in accordance with previous findings (Klug et al. 2010).

We found that only 6% of hypomethylated regions overlapped with a promoter (**Figure 1D**) and that the vast majority of hypomethylated regions were located distal to TSSs (median distance of ~35 kb from the nearest TSS; **Figure 1E**; **Supplementary Table 2**). Hypomethylated regions occured in genomic regions that show increased levels of evolutionary conservation (**Supplementary Figure 4**), a finding that supports their functional importance. Moreover, gene ontology analysis revealed that these regions are significantly enriched (false discovery rate (FDR) < 0.05) near genes known to play a key role in the regulation of immune processes,

including the regulation of transcription, signal transduction, and cell apoptosis (**Figure 1F**; **Supplementary Table 4**). The set of genes near hypo-DMRs included virtually all of the "master-regulators" of innate immune responses, including *CREB5, REL*, *NFKB1*, *IRF2*, and *IRF4*. It also included key genes involved in DC-mediated activation of B and T cells (e.g., *CD83*) and the regulation of cell death (e.g., *BCL2*).
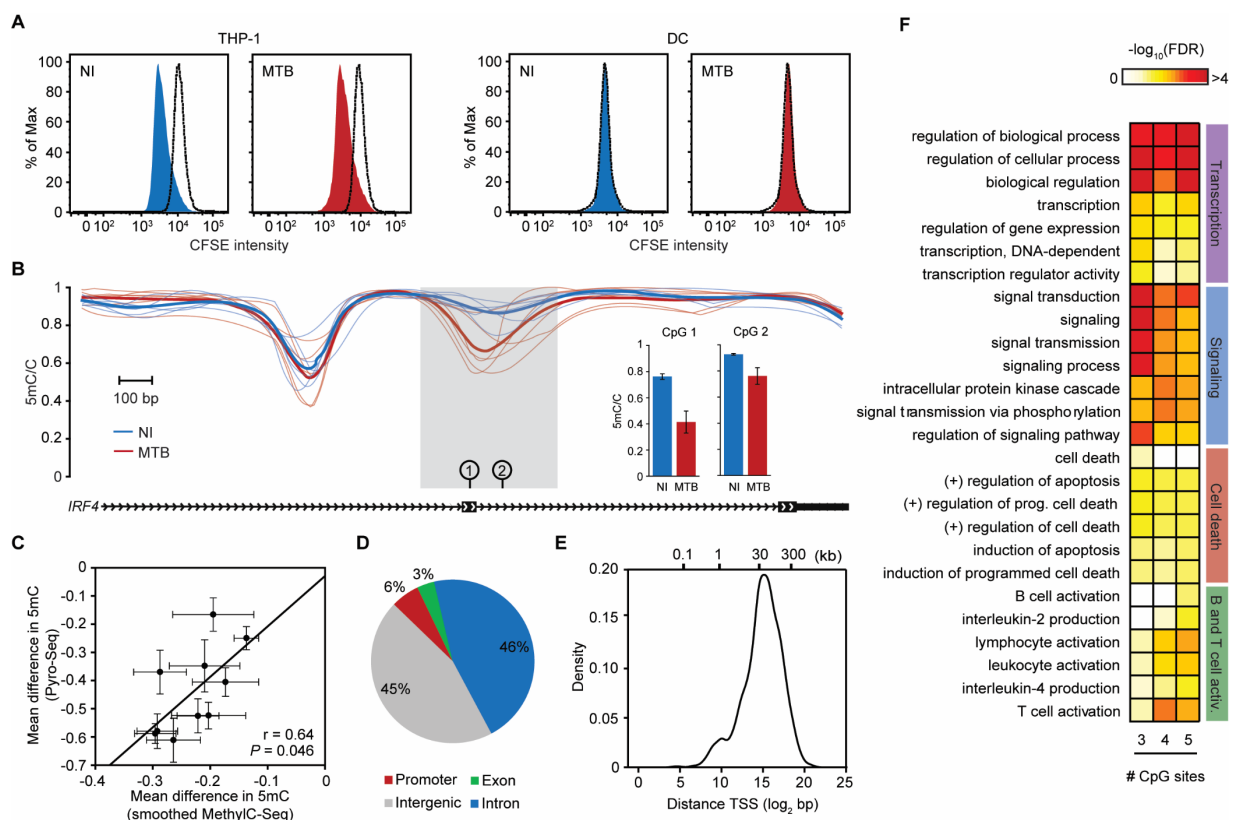
Figure 1. MTB-induced changes in methylation in post-mitotic human DCs. (**A**) CFSE-labeled THP-1 (left) and CFSE-labeled DCs (right). Proliferation was assayed in either non-infected cells or cells infected for 18 hours with MTB. Similar results were observed 48 hours post-infection (Supplementary Figure 16). (**B**) Example of a region showing active loss of DNA methylation in response to MTB infection (gray shading). The plot shows smoothed methylation values (y-axis) for six non-infected (blue) and six MTB-infected samples (red). Thick blue and red lines show average methylation levels for non-infected and infected cells, respectively. The inset on the right shows methylation levels at two individual CpG sites within the

hypomethylated region using bisulfite pyrosequencing as a validation method. (**C**) Scatter plot showing the correlation between MethylC-seq (x-axis; smoothed data) and pyrosequencing data (y-axis) for mean differences in methylation between infected and non-infected cells, at 10 CpG sites within hypomethylated DMRs. Data are represented as mean ± s.e.m., n = 6 for MethylC-seq and n = 5 for Pyro-seq. (**D**) Pie charts showing the distribution of hypomethylated regions in different genomic regions. Each MTB-DMR is counted only once: the overlap of a genomic region excludes all previously overlapped MTB-DMRs, starting clockwise from promoters (TSS ± 500 bp; red). (**E**) Distribution of distances of MTB-DMRs to the nearest TSS. (**F**) Representative gene ontology (GO) terms enriched among genes associated with hypomethylated regions. To demonstrate that the enriched biological processes are largely robust to the cutoff used to define MTB-DMRs, we show how these results differ depending on the number of differentially methylated CpG sites ($P < 0.01$) required to call an MTB-DMR (from at least three to at least 5 consecutive sites).

Active changes in methylation occur in regions enriched for 5-hydroxymethylcytosine

The TET family proteins catalyze the conversion of methylated cytosine (5mC) to 5-hydroxy-methylcytosine (5hmC), and are thus key players in the process of active demethylation. To evaluate if 5hmC levels dynamically change in response to MTB infection (as expected if 5mC sites must pass through the 5hmC state before demethylation), we generated single base-pair resolution maps of 5hmC across the genome using TET-assisted bisulfite sequencing (TAB-seq) (Yu et al. 2012) in one of the 5 original donor. As previously described for other cell populations (Song et al. 2011; Lister et al. 2013), we found markedly higher levels of 5hmC in gene bodies of highly expressed genes, consistent with a role for 5hmC in maintaining and/or promoting gene expression (**Figure 2A**) (Hahn et al. 2013; Hon et al. 2014).

Next, we evaluated if 5hmC marks were enriched within hypomethylated MTB-DMRs. We found that regions that became hypomethylated post-infection were already associated with significantly higher levels of 5hmC prior to infection (3.6-fold enrichment; Wilcoxon test; $P < 1 \times 10^{-16}$). Upon infection, 5hmC levels increased even further (Wilcoxon test; $P = 1.57 \times 10^{-11}$; **Figure 2B,C**), suggesting that 5hmC plays an important role in the cascade of events leading to active demethylation. The increase in 5hmC appears to be specific to hypomethylated regions since no enrichment was observed genome-wide, a result supported by quantitative immunocytochemistry data (**Figure 2D,E**). The striking enrichment of 5hmC within MTB-DMRs prior to infection strongly suggests that, in addition to its role as a transitory demethylation intermediate, 5hmC might also contribute to coordinating the gene expression program induced in response to a microbial stimulus.
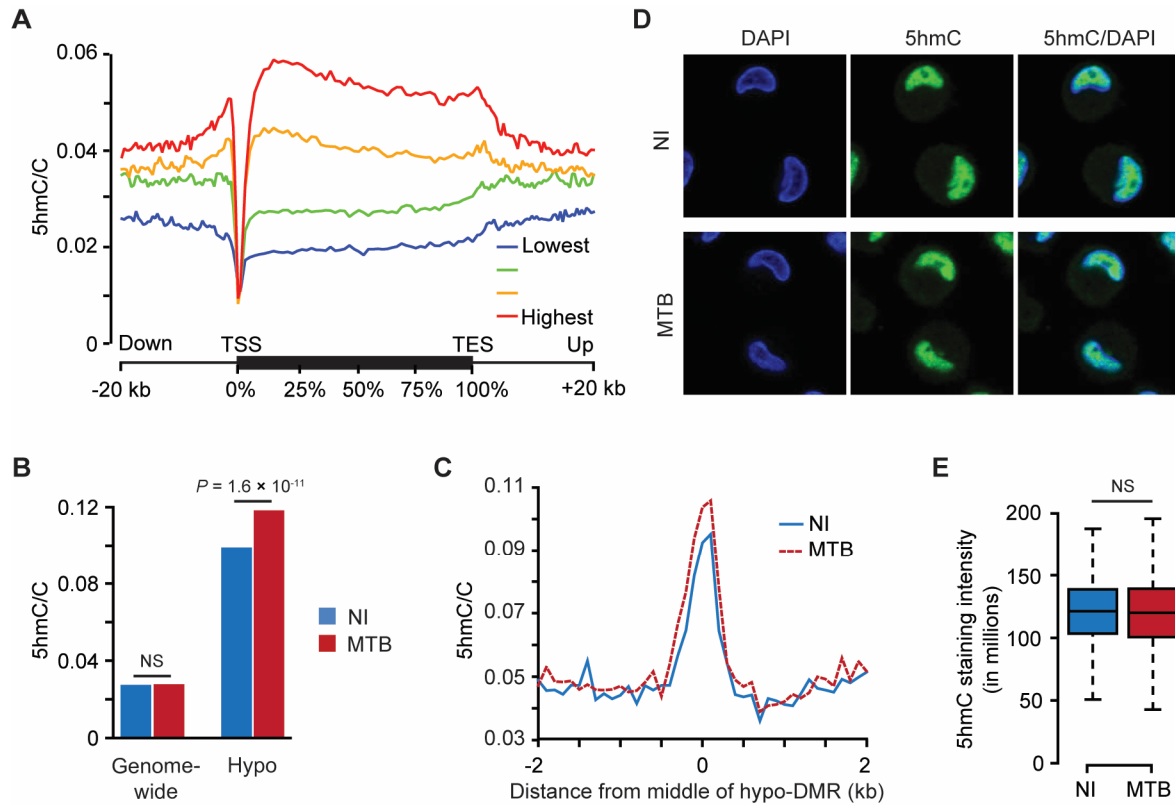
Figure 2. 5hmC is enriched in MTB-DMRs prior to infection. (**A**) Metagene profiles of 5hmC levels relative to Ensembl transcripts expressed at different levels in human DCs. We grouped genes in four quantiles based on their expression levels on non-infected DCs. (**B**) Bar plots showing mean 5hmC/C ratios within hypomethylated regions, before (blue) and after infection (red). (**C**) Composite plots of patterns of 5hmC before and after MTB infection ±3 kb around the midpoint of hypomethylated regions. (**D**) 5hmC staining in non-infected (top panel) and MTB-infected DCs (bottom panel). 5hmC levels are given by the levels of Alexa 488 (green: middle panel). Cells counterstained with DAPI to localize the nucleus are shown in the first panel. (**E**) Boxplots showing the distribution of 5hmC staining intensity. No significant differences were observed between the two groups.

MTB-DMRs overlap with enhancer elements that gain activation marks upon infection

Given that MTB-DMRs are primarily found distal to TSSs, we predicted that MTB-DMRs would overlap with enhancer regions. To test this hypothesis and evaluate how the chromatin states associated with MTB-DMRs dynamically change in response to infection, we collected ChIP-seq data for six histone marks (H3K4me1, H3K4me3, H3K27ac, H3K27me3, H3K36me3 and H3K9me3) in non-infected and infected DCs (Supplementary Table 1) from two additional donors. Using these data, we generated genome-wide, gene regulatory annotation maps for non-infected and MTB-infected DCs using the ChromHMM chromatin segmentation program (Figure 3A; Supplementary Figure 5) (Ernst and Kellis 2012). We found that 41% of hypomethylated regions overlapped with a ChromHMM-annotated enhancer region (defined by the presence of H3K4me1) already present in non-infected DCs, a 7.4-fold enrichment compared to genome-wide expectations ($\chi 2$-test; $P < 1 \times 10^{-16}$; **Figure 3B,C**; **Supplementary Table 2**). Slightly higher enrichments (8.1-fold; $P < 1 \times 10^{-16}$) were observed when defining chromatin states in MTB-infected DCs. Given the high-resolution of our histone maps, we could further distinguish between active and inactive/poised enhancer elements based on the presence or absence of the H3K27ac mark, respectively, in addition to H3K4me1 (Heintzman et al. 2007; Creyghton et al. 2010; Rada-Iglesias et al. 2011). Overall, we found that MTB infection leads to a significant increase of active enhancer elements (and decrease of inactive/poised enhancers) colocalizing with MTB-DMRs (**Figure 3B,C**).

We next extended our analysis by examining chromatin transition states at hypomethylated regions in response to MTB-infection. We found that 42% of hypomethylated regions occurred in regions that exhibited infection-dependent changes in chromatin state, a significantly higher proportion than expected compared to the rest of the genome ($P_{resampling} < 0.001$; Figure 3E). The chromatin state transitions observed within hypomethylated regions were primarily explained by the acquisition of histone activating marks (e.g., H3K27ac) in MTB-infected cells. For example, among hypomethylated regions that overlapped with predefined enhancers (i.e., enhancers observable in non-infected cells), 85% of those that exhibit a change in chromatin state gained an activation mark (H3K27ac or H3K27ac+H3K4me3; **Figure 3F,G**;

Supplementary Figure 6A). This proportion was markedly larger than that observed genome-wide (37%) ($\chi2$-test; $P = 1.1 \times 10^{-59}$; Figure 3F). Notably, we also found a large number of hypomethylated regions (n = 218; 12.7% of all hypomethylated regions) that overlapped with heterochromatin/repressed regions before infection but gained *de novo* enhancer marks upon MTB infection (H3K4me1 (+ H3K27ac + H3K4me3)). The number of *de novo* enhancers we observed among hypomethylated regions was significantly higher than expected by chance ($P_{resampling} < 0.001$; **Figure 3D,E,G**; **Supplementary Figure 6A**). The identification of enhancers only present in infected DCs resembles recent findings showing that, in response to different immune stimuli, mouse macrophages can gain *de novo* putative enhancer regions that were absent in naive cells (Kaikkonen et al. 2013; Ostuni et al. 2013). Interestingly, we observed that 5hmC was significantly enriched among *de novo* hypo-DMRs prior to infection (Wilcoxon test; $P = 5.27 \times 10^{-149}$), suggesting that 5hmC might be an early "pre-marking" mechanism of enhancer activation, even before the deposition of H3K4me1 marks (**Supplementary Figure 6A,B**).

Finally, we found that MTB-induced activation or *de novo* gain of enhancer elements at hypomethylated regions was associated with the induction of putative enhancer RNAs (eRNAs) (Wang et al. 2011) in these intergenic regions (as measured by whole-transcriptome RNA-seq) as well as with increased levels of histone marks associated with transcriptional activity (**Supplementary Figure 7**). Moreover, changes in eRNA levels in response to MTB infection show a striking positive correlation with changes in gene expression levels of nearby genes (r = 0.49, $P = 7.6 \times 10^{-13}$; **Supplementary Figure 7**), in support of a mechanistic link between demethylation, eRNA production and the regulation of proximal protein-coding genes (Lam et al. 2014).
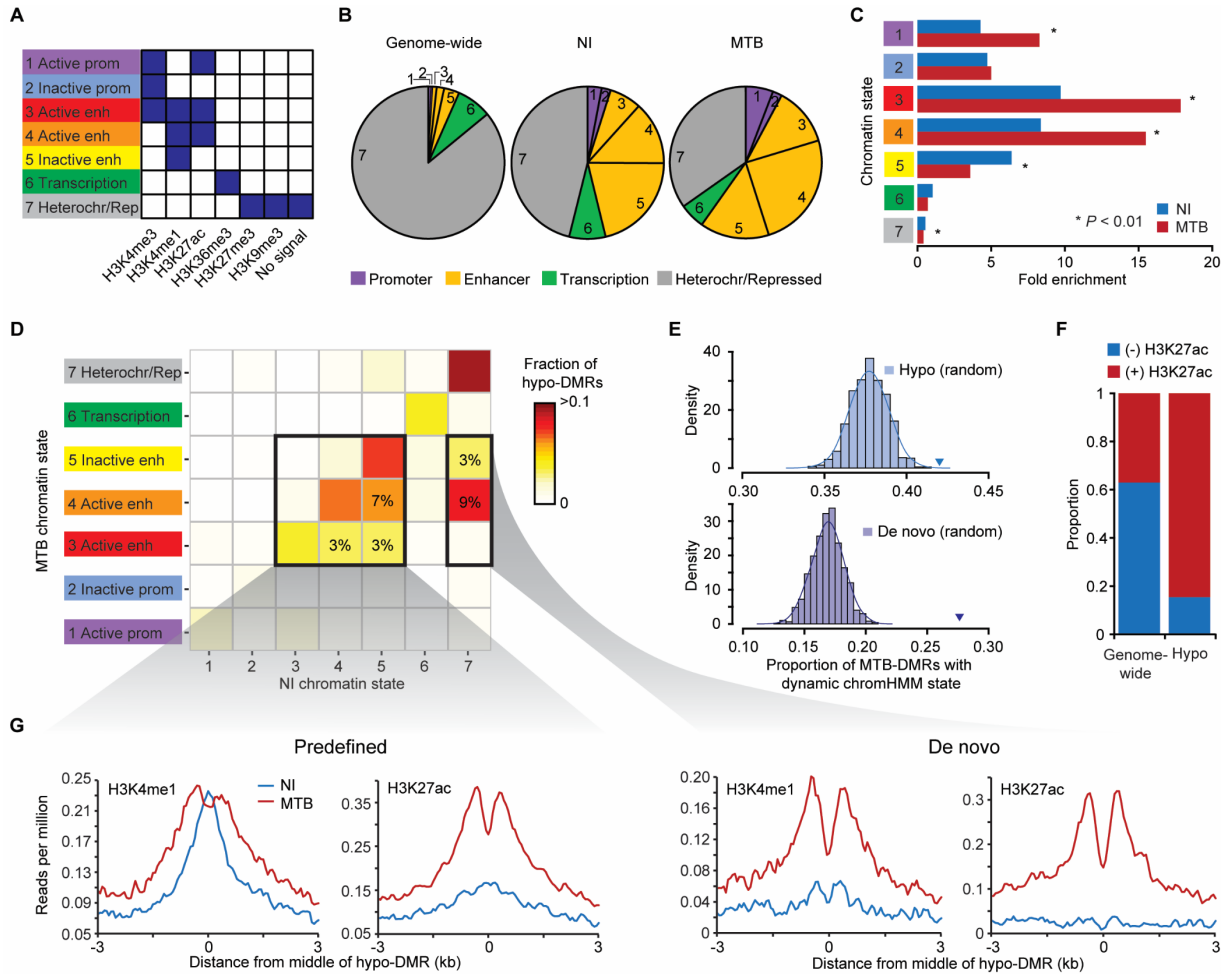
Figure 3. MTB-DMRs overlap with enhancer elements that become active upon infection in hypomethylated regions. (**A**) Combination of histone patterns used to define the 7 chromatin states. The precise relative contribution of each chromatin mark to each of the chromHMM-defined states can be found in Supplementary Figure 3. Note that state 7 was defined by either no signal *or* the presence of either H3K27me3/H3K9me3. (**B**) Pie charts showing the distribution of chromatin state annotations genome-wide (on non-infected cells) and within all MTB-DMRs in either non-infected (blue) or MTB-infected cells. The chromatin state codes are as defined in (A). (**C**) Fold enrichments of the different chromatin states within hypomethylated regions as compared to genome-wide expectations in non-infected (blue) and MTB-infected cells (red). (**D**) 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. The two inner boxes indicate two

subgroups of hypomethylated regions, predefined enhancer (detectable enhancer in non-infected DCs) and *de novo* enhancers (detectable enhancer only in MTB-infected DCs). The numbers inside the cells refer to the proportion of hypomethylated regions that undergo each of the highlighted transitions. (**E**) Top panel: Histogram showing the observed proportion of regions that change chromatin state after infection (any transition) when sampling 1000 random sets of regions matched to the chromatin states found in non-infected samples within hypomethylated regions. Each random set contains the same number of hypomethylated regions as those identified in the true data (n = 1,714). The blue triangle represents the observed proportion of hypomethylated regions that changed chromatin state in response to MTB infection. Bottom panel: Same as above but focusing on regions of the genome labeled as heterochromatin/repressed before infection (state 7; n= 790) that gain *de novo* enhancer marks upon MTB infection (states 3, 4, or 5). The blue triangle represents the proportion observed within the true set of hypomethylated regions. (**F**) Bar plots showing the proportion of hypomethylated regions that overlap with enhancers and show dynamic changes in chromatin state, as defined by the gain or loss of H3K27ac mark. (**G**) Composite plots of patterns of H3K4me1 and H3K27ac ChIP-seq signals ±3 kb around the midpoints of hypomethylated regions (x-axis) overlapping with predefined (right) and *de novo* (left) enhancers.

MTB-DMRs are bound by signal-dependent transcription factors

We next asked if MTB-infection was associated with changes in the levels of chromatin accessibility in MTB-DMRs. We mapped regions of open chromatin in non-infected and infected DCs based on genome-wide sequencing of regions showing high transposase (Tn5) sensitivity (using ATAC-seq in one additional donor) (Buenrostro et al. 2013). Overall, we observed that MTB-DMRs colocalize with regions of open chromatin, which further reinforces the regulatory potential of these regions (**Figure 4A**). Interestingly, we found that the response to MTB-infection was accompanied by a striking increase in Tn5 sensitivity levels in hypomethylated regions, which indicates that the chromatin in these regions became more accessible after infection (**Figure 4A**). This observation is commensurate with our data showing the acquisition of active histone marks in these regions, and further supports the idea that hypomethylated regions frequently reflect the presence of regulatory elements that become more active in response to infection.

An attractive feature of ATAC-seq data is the ability to identify motif instances occupied by transcription factors (TF) within regions of open chromatin (Neph et al. 2012; Buenrostro et al. 2013). We did so by using a modified version of the Centipede algorithm (Pique-Regi et al. 2011) specifically devised to test for aggregate differential binding of TFs between two experimental conditions. This method, which we call CentiDual, compares the intensity of the Tn5 sensitivity-based footprint across all matches to a given motif in the genome, between non-infected and infected samples (see Methods for details on the statistical model). We found compelling evidence for measurable, genome-wide transcription factor activity (i.e., binding to the genome; Bonferroni-corrected $P < 0.05$) in either non-infected or infected DCs for 264 TF binding motifs, representing over 200 unique transcription factors (some TFs can bind different motifs; **Supplementary Table 5**). Of these TF binding motifs, we found 55 that were differentially bound between non-infected and infected DCs (Bonferroni-corrected $P < 0.05$; 27 show increased binding and 28 show decreased binding; **Figure 4B**). Among TF binding motifs showing increased genome-wide binding after infection, we found several that are associated with NF-κB/Rel (e.g., NFKB1, REL) and IRFs (e.g., IRF1, IRF2) family members (**Figure 4B**; **Supplementary Table 5**), both of which play a primary role in the regulation of inflammatory

signals in response to infection (Smale 2010). Interestingly, several CTCF motifs showed significantly decreased binding in infected DCs (Bonferroni-corrected $P < 1.85 \times 10^{-14}$, **Supplementary Table 5**). CTCF is a well-established transcriptional insulator (Ong and Corces 2014), raising the possibility that the release of CTCF in response to infection might be an important mechanism for the regulation of efficient immune responses.

We next used CentiDual to test for differential binding within MTB-DMRs. Within hypomethylated regions we found increased binding (FDR < 0.1) at 8 TF binding motifs after infection. Strikingly, all of these motifs were associated with immune-induced TFs from the NF-κB/Rel (e.g., REL; FDR = 1.57×10^{-6}), AP-1 (FDR = 4.9×10^{-3}), or IRF (FDR = 3.97×10^{-3}) families (Figure 4C; Supplementary Table 5). This result demonstrates that hypomethyated regions correspond to places where immune-activated TFs are recruited after infection. In accordance with this argument, we found that, in infected DCs, TF binding motifs associated with NF-κB/Rel, AP-1, and IRF families were all significantly enriched within hypomethylated regions (up to 16-fold; **Supplementary Figure 8A**). Indeed, in MTB infected DCs, over 50% of the hypomethylated regions were bound by at least one of these signal-dependent TFs, which corresponds to an 3.8-fold increase relative to chance expectations (based on sampling random regions of the genome matched for length and GC content; $\chi2$-test; $P = 3.94 \times 10^{-63}$; **Supplementary Figure 8B**).
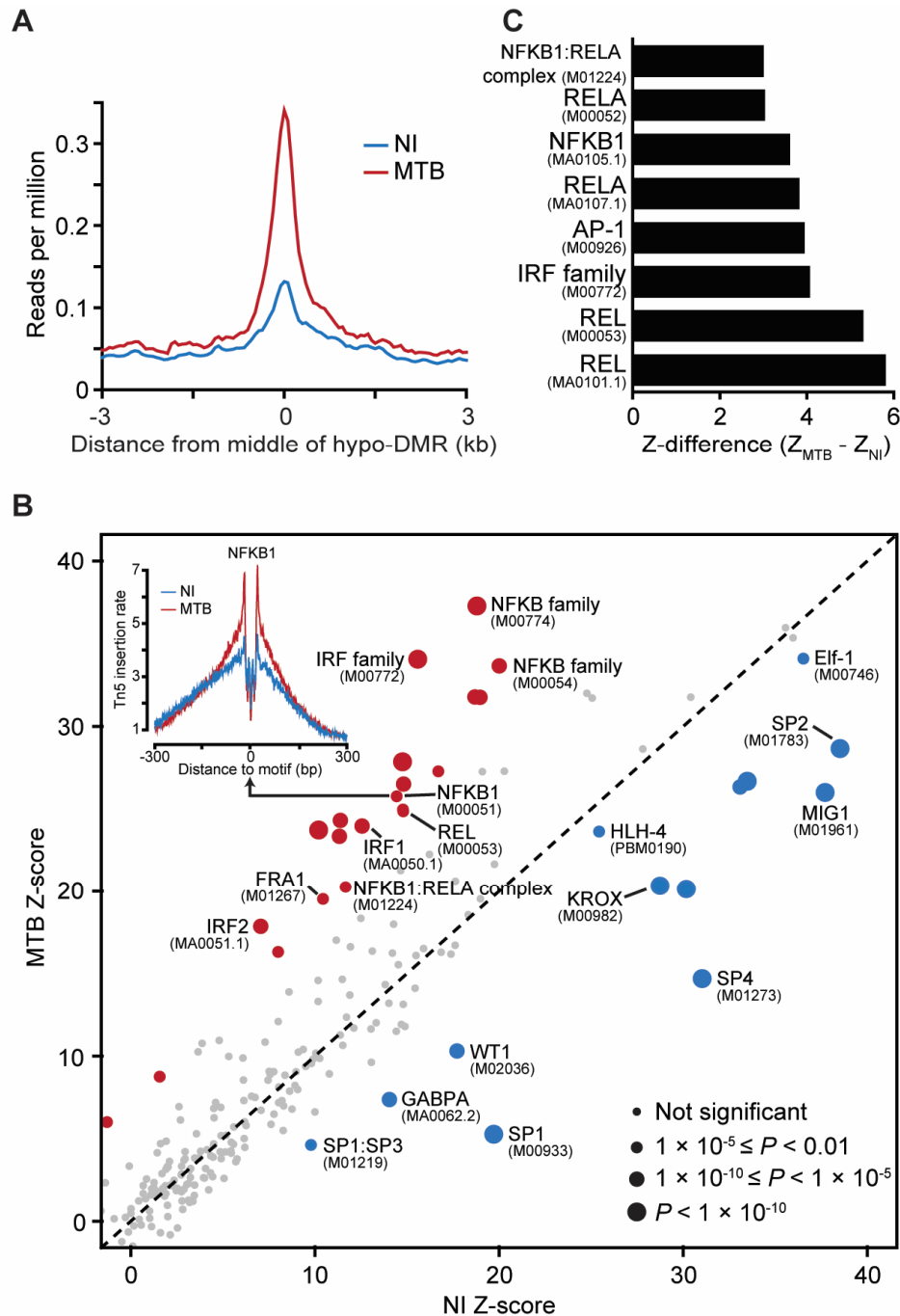
Figure 4. MTB-DMRs are bound by signal-dependent transcription factors. (**A**) Tn5-accessibility profiles before and after MTB infection, ±3 kb around the midpoints of hypomethylated regions. (**B**) Scatterplot comparing transcription factor occupancy score predictions between non-infected (y-axis) and MTB infected DCs (y-axis). The size of the dots is proportional to the level of statistical significance supporting differential binding in response

to MTB infection. Red dots represent TFs that show evidence for increased binding after MTB infection and blue dots represent TFs that show evidence for decreased binding after infection. The inset on the top right corner shows the genome-wide footprint on NF-kappaB (p50) motif (motif ID: M00051) in non-infected (blue) and MTB-infected DCs (red). In this example, the footprint in MTB-infected DCs is clearly stronger, which supports increased TF binding of at NF-kappaB (p50) motif genome-wide upon MTB infection. (**C**) TF motifs (motif IDs in parenthesis) that show significantly increased binding in hypomethylated regions after MTB infection.

MTB-DMRs are associated with genes differentially expressed in response to MTB infection

Finally, we asked if genes associated with hypomethylated regions were more likely to change expression levels in response to infection. We classified 2,051 and 1,947 genes as significantly up- or down-regulated post-infection, respectively (FDR < 1×10^{-4} & |log2 fold change| > 1; **Supplementary Table 6**). We next tested whether genes located near hypomethylated regions were more likely to be differentially expressed upon MTB infection relative to all genes in the genome. To do so, we first associated each hypomethylated region with a unique gene using the following criteria: if a hypomethylated region was located within a gene body, the region was assigned to that gene; otherwise, we assigned each hypomethylated region to the gene with the TSS closest to the midpoint of the MTB-DMR. Then, we tested for an enrichment of differentially expressed (DE) genes among three classes of genes: (*i*) "hypo-DMR-genes" corresponding to the set of genes associated with hypomethylated regions (n = 1,291); (*ii*) "predefined-DMR-genes" corresponding to the set of genes in hypomethylated regions that overlapped with predefined enhancer elements (n = 508, a subset of class *i*), and (*iii*) "*de novo*-DMR-genes" corresponding to the set of genes in hypomethylated regions that overlapped with *de novo* enhancer elements (n = 180, also a subset of class *i*).

We found that hypo-DMR-genes (class *i*) were significantly enriched among DE genes (1.6-fold, $\chi 2$-test; $P = 1.07 \times 10^{-17}$; **Figure 5A,B**) compared to all genes in the genome, consistent with the observation that changes in DNA methylation were globally correlated to changes in expression after infection (Supplementary Figure 9). This enrichment was noticeably stronger for predefined-DMR-genes (class *ii*; 1.9-fold, $P = 3.37 \times 10^{-14}$) and even more so for *de novo*-DMR-genes (class *iii*; 2.5-fold, $P = 6.52 \times 10^{-14}$). Indeed, among *de novo*-DMR-genes, 54% were DE, even at the very stringent cutoffs we used to define DE genes (**Figure 5A,B**). Among DE genes associated with hypomethylated regions, 74% were up-regulated after MTB infection – substantially more than the 51% of up-regulated genes observed genome-wide ($\chi 2$-test; $P = 4.4 \times 10^{-79}$, **Figure 5C,D**). This observation was even more pronounced when focusing specifically on predefined-DMR-genes (class *ii*) and *de novo*-DMR-genes (class *iii*), for which

78% ($P = 8.68 \times 10^{-36}$) and 94% ($P = 6.9 \times 10^{-23)}$, respectively, were associated with increased expression levels in response to infection (**Figure 5C**).

Finally, we performed a paired time course analysis of gene expression and DNA methylation levels for 6 hypo-DMRs (associated with 6 immune-related genes: *IRF4*, *REL*, *TRAFD1*, *CD83*, *BCL2*, *NFKB1*) aimed at defining the relative order of changes in DNA methylation versus changes in gene expression. We found that differential methylation was undetectable at 2 hours post-infection, even though half of the genes associated with these sites (3 out of the 6 genes tested) were already significantly up-regulated (**Figure 5E**). Thus, we speculate that demethylation is not required for TF binding at enhancer elements or for the subsequent up-regulation of the associated gene. However, for all CpG sites tested, the magnitude of DNA hypomethylation (relative to non-infected cells) increased at later time points until methyl marks were almost completely erased in infected cells, even if for most genes the largest fold changes in gene expression were observed at earlier time-points (**Figure 5E**).
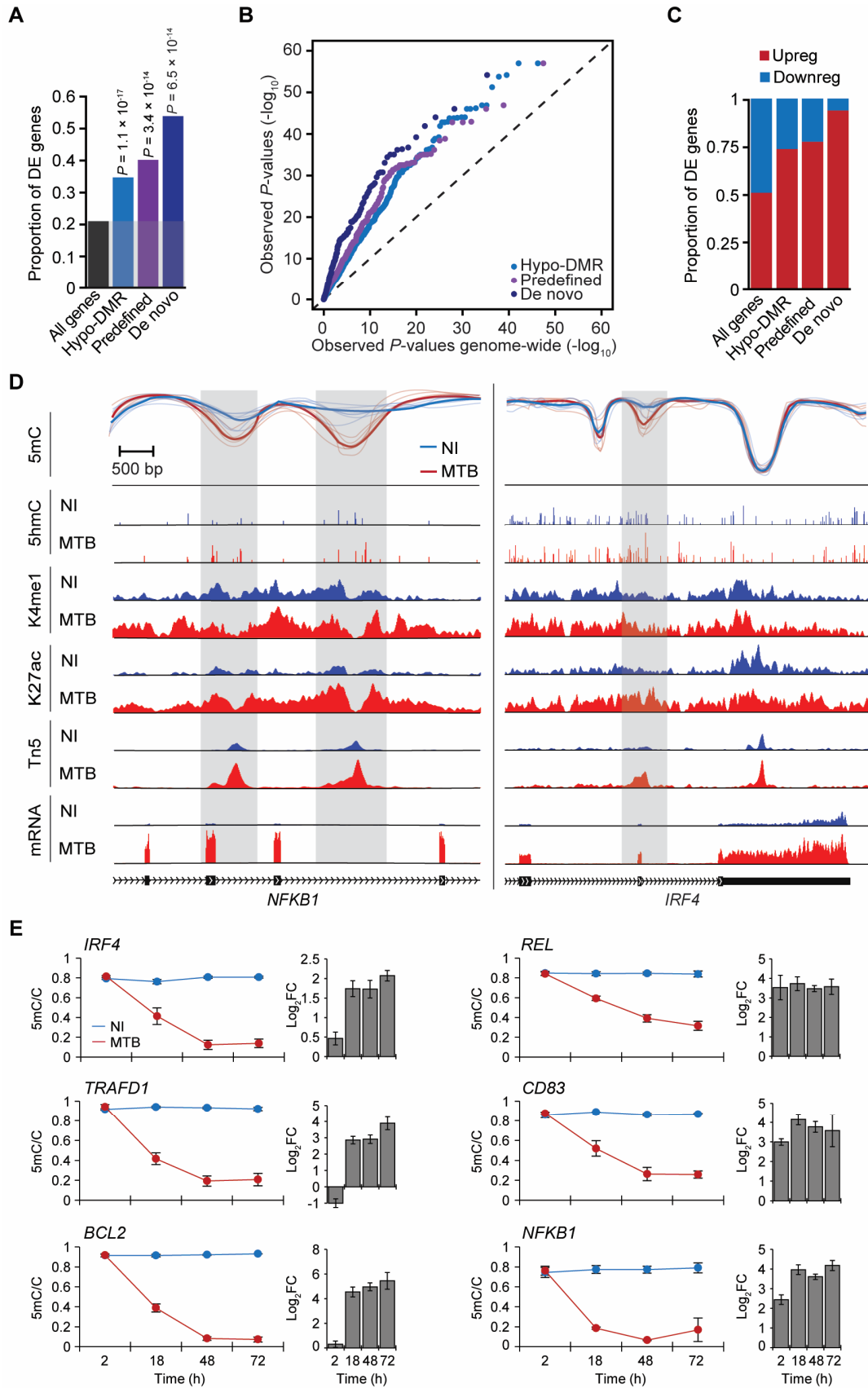
Figure 5. Differential methylation is coupled to differential gene expression. (**A**) Proportion of differentially expressed genes (y-axis) observed among all tested genes and among genes associated with different subgroups of hypo-DMRs. (**B**) QQ-plot showing that genes in the vicinity of hypo-DMRs show stronger statistical evidence for being differentially expressed in response to MTB infection (*P*-values on y-axis) compared to all genes tested (*P*-values on x-axis). (**C**) Proportion of up- and down-regulated genes among DE genes associated with the different subgroups of hypo-DMRs. (**D**) Examples of genes encoding for two key transcription factors, *NFKB1* (left panel) and *IRF4* (right panel) that are strongly upregulated in response to MTB infection and for which we identified one or more hypomethylated regions (gray shading) that overlap with putative enhancer elements. Normalized read signals for the indicated features are shown for non-infected (blue tracks) and infected conditions (red tracks). (K4me1) H3K4me1; (K27ac) H3K27ac; (Tn5) transposase-accessible chromatin; (mRNA) mRNA expression levels; (**E**) Changes in DNA methylation levels (y-axis) measured by pyrosequencing across four time points after MTB-infection (2, 18, 48, and 72 hours (h)) along with the corresponding fold changes in log2 scale (log$_2$FC) in normalized gene expression of the associated gene. Blue and red lines represent average methylation levels in non-infected and MTB-infected DCs, respectively. All data are represented as mean $\pm$ s.e.m., with a minimum of three biological replicates per group. PyroMark and real-time PCR data are reported in Supplementary Tables 7 and 8, respectively.

Discussion

The possibility that active changes in methylation, particularly demethylation, can occur in mammals has been a matter of debate for decades (Ooi and Bestor 2008; Pastor et al. 2013). Here, we provide compelling evidence that the response of human DCs to MTB infection is accompanied by widespread, rapid loss in DNA methylation. Although many possible mechanisms can account for this loss (Kohli and Zhang 2013), the observation that hypomethylated regions show increased levels of 5hmC in response to MTB infection strongly suggests that the family of TET proteins (TET1, TET 2, and TET3) are involved in this process. This possibility is further supported by recent studies showing that TET2 is required for active DNA demethylation in human monocytes (Klug et al. 2013) and during brain development (Lister et al. 2013). TET2 is also highly expressed in DCs at both the mRNA and protein levels (**Supplementary Figure 10**).

By integrating our methylation maps with ChIP-seq data on six histone marks we show that active demethylation occurs almost exclusively at distal regulatory elements, particularly enhancers. This observation, which is robust to the cutoffs used to call MTB-DMRs (**Supplementary Figure 11**), parallels what has been previously described in differentiating cells and during developmental processes (Ji et al. 2010; Stadler et al. 2011; Ziller et al. 2013) despite the fact that the mechanisms controlling active and passive changes in methylation are markedly different (Kohli and Zhang 2013). In contrast to previous studies that also reported active changes in methylation (e.g., in response to neuronal activation (Guo et al. 2011); or during monocyte differentiation into macrophages or dendritic cells (Klug et al. 2010)), we found a strong association between DMRs and changes in gene expression of nearby genes. The apparent discrepancy between our results and those previously reported is probably explained by the fact that past studies have only investigated active methylation changes in promoter regions, which our data suggest are infrequent, or only on a small subset of all CpG sites in the (mouse) genome (~1%). Moreover, we decided to focus on differentially methylated regions (3 or more consecutive differentially methylated CpGs) instead of methylation changes at individual CpG sites (Klug et al. 2010; Guo et al. 2011), a decision that likely enriched our analysis for DMRs involved in changes in gene expression. In support of this hypothesis, we

found that the enrichment for DE genes become stronger as we focus on MTB-DMRs with a larger number of differently methylated CpG sites (**Supplementary Figure 11**). More broadly, our results highlight the key importance of using single base-pair resolution maps of the DNA methylome in order to fully capture the relationship between changes in methylation and changes in gene expression.

We show that demethylation is often associated with the gain of histone activation marks and the recruitment of immune-activated TFs in response to infection. The recruitment of NF-κB and other master regulators to hypomethylated regions is likely associated with the opening of the chromatin in these regions, although it remains unclear whether the chromatin opens to allow the binding of these TFs (i.e., prior to binding) or if the observed increase in chromatin accessibility is a consequence of the binding itself. Our results shed some light on this problem. Specifically, we observed that changes in gene expression sometimes tend to occur *prior* to detectable changes in DNA methylation, at least in the hypo-DMRs we investigated in our time course experiment. These results support a model in which TF binding to enhancers leads to gene up-regulation followed by active demethylation, rather than vice-versa, consistent with the sequence of events proposed for other cellular contexts (Stadler et al. 2011; Schubeler 2015). We note, however, that our pyrosequencing data does not allow us to distinguish between 5-mC and 5hmC. Thus, it is possible that 5hmC levels were increased 2 hours post-infection and that these changes preceded the activation of certain enhancers, as recently suggested in T cells (Ichiyama et al. 2015). More generally, we cannot completely exclude the possibility that demethylation occurred prior to changes in gene regulation, but only in a small proportion of the cells (1 – 2%, based on the sensitivity threshold of pyrosequencing assays: (Tost and Gut 2007), making it difficult to detect at the 2 hour time point. Under this scenario, demethylation in only a few cells could account for the observed changes in gene expression. However, we have previously shown that >30% of DCs uptake MTB bacteria after only 1 hour post-infection using the same protocol (Barreiro et al. 2012), and in single-cell RNA-seq data from dendritic cells, >50% are transcriptionally responsive to immune challenge (Shalek et al. 2014). Thus, it seems unlikely that our results reflect a response driven by only a small minority of cells. Importantly, even if TF binding instigates changes in methylation, binding alone is not

sufficient: the vast majority (>99%) of binding events induced by infection occur at regions that do not change methylation (**Supplementary Figure 12**).

Finally, there is increasing evidence that, after a first encounter with a pathogen or other immune stimulus, innate immune cells keep such attacks "in memory." As a result, they are able to mount faster and stronger gene transcriptional responses upon restimulation and exhibit increased resistance to secondary infection. This process, termed trained immunity (Monticelli and Natoli 2013; Quintin et al. 2014; Saeed et al. 2014), has been attributed to epigenetic reprograming at the level of histone H3 methylation based on the observation that distal regulatory elements that gain *de novo* H3K4me1 (i.e., *de novo* enhancer marks) in response to immune activation generally do not lose this mark after the stimulation has ceased (Ostuni et al. 2013). Although epigenetic programming through histone modifications might be an important factor in trained immunity, our results raise the possibility that changes in DNA methylation might also contribute to short-term memory in innate immune cells. Indeed, changes in DNA methylation might be ideally suited as a mechanism of epigenetic memory since these changes are expected to be thermodynamically more stable and longer lasting than changes in histone marks. In support of this idea, we observed that the magnitude of DNA hypomethylation gradually increased with time since infection, and never reverted back to higher levels during our 72-hour time course experiment. Moreover, we show that the gain of *de novo* enhancers – assumed to account for trained immunity – often occurs concomitantly with the loss of DNA methylation in the same regions. Our results thus raise the possibility that trained immunity might not only be due to post-transcriptional changes in histone marks but also, and possibly primarily, due to changes in DNA methylation.

Methods

Biological material and sequencing libraries

Details of the experimental and statistical procedures can be found in the Supplementary Methods section. Blood samples from healthy donors were obtained from *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). Blood mononuclear cells from each donor were isolated by Ficoll-Paque centrifugation and blood monocytes were purified from peripheral blood mononuclear cells (PBMCs) by positive selection with magnetic CD14 MicroBeads (Miltenyi Biotec). Monocytes were then derived into DCs as previously described (Barreiro et al. 2012) and subsequently infected with MTB for 18 h at a multiplicity of infection of 1-to-1. For the ChIP-seq and ATAC-seq experiments, we used heat-killed bacteria (5-to-1 ratio), which we show to lead to virtually the same transcriptional response at 18 hours to that observed with live MTB (r = 0.91; Supplementary Figure 13). RNA-seq libraries were prepared using the TruSeq RNA Sample Prep Kit v2 or the Illumina Total Stranded RNA Library kit, as per the manufacturer's instructions. MethylC-seq libraries were generated by ligation of methylated sequencing adapters to fragmented genomic DNA followed by gel purification, sodium bisulfite conversion and six cycles of PCR amplification. TAB-seq libraries were generated as previously described (Yu et al. 2012) and ChIP-seq libraries for the six histone marks were prepared following the Illumina protocols, with minor modifications (see Supplementary Methods). ChIP-seq signals from the two biological replicates were highly concordant (mean r = 0.94 and range = 0.87-0.99; Supplementary Figure 14) and were combined for all analysis. Finally, ATAC-seq libraries were generated from 100,000 cells, as previously described (Buenrostro et al. 2013). Sequencing was performed using the Illumina HiSeq 2000 or 2500, as per the manufacturer's instructions

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). 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.

5hmC staining

The protocol was adapted from Santos et al. (Santos et al. 2003). 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.

Read processing and alignment

Sequencing data was processed using the Illumina analysis pipeline and FASTQ format reads were aligned to the human reference genome (GRCh37/hg19) using Bowtie 2 (Langmead and Salzberg 2012). Methylation levels for each CpG site were estimated by counting the number of sequenced C ('methylated' reads) divided by the total number of reported C and T ('unmethylated' reads) at the same position of the reference genome using the bismark package (Krueger and Andrews 2011). The summarized methylation estimates of strand-merged CpG sites were used to identify differences in methylation between non-infected and infected samples using the R package BSmooth (Hansen et al. 2012). RNA-seq based gene expression levels were estimated using HTSeq and differently expressed genes following MTB infection of DCs using the R package DESeq2 (Anders et al. 2013). ChromHMM (Ernst and Kellis 2012) was used to segment the genome into different chromatin states based on six histone modifications and the ChIP input. The ChromHMM model was learned separately for both infected and non-infected

DCs (main text) or using a unified model that learns and defines chromatin states in both infected and non-infected DCs at the same time (**Supplementary Figure 15**).

ATAC-seq data processing and footprinting analysis

Footprinting analyses were performed using a modified version of the Centipede algorithm (Pique-Regi et al. 2011) specifically devised to test for differential binding between two experimental conditions. 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 β parameter in:

$$\log\left(\frac{\pi_l}{1-\pi_l}\right) = \alpha + \beta \, \text{PWMscore}_l$$

where π_l represent the prior probability of binding in Centipede's model in motif location l. 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 π_{lt} for each condition t separately and for each site l. 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 (α and δ), the PWM effect β, and PWM times the treatment effect γ:

$$\log\left(\frac{\pi_{lt}}{1-\pi_{lt}}\right) = \alpha \times \left(1-I_t\right) + \beta \times \text{PWMscore}_l + \delta \times I_t + \gamma \times \left(I_t \times \text{PWMscore}_l\right)$$

where I_t 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 γ, corresponding to the evidence for condition specific binding.

Data visualization in the Immune Epigenome Browser

The browser, implemented using the WashU Epigenome Browser web interface (Zhou and Wang 2012), 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*.

Data access

Data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE64173 (ATAC-seq), GSE64175 (ChIP-seq), GSE64181 (TAB-seq), GSE64182 (wtRNA-seq), GSE64179 (mRNA-seq), and GSE64177 (MethylC-seq). PyroMark and real-time PCR data are reported in **Supplementary Tables 7** and **8**, respectively.

Acknowledgments

Supplementary Methods

In this section we repeat some of the text from the Methods 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).

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). 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; **Supplementary Figure 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). 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 manufacturer's 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.

TAB-seq library preparation and sequencing

TAB-seq libraries were performed as previously described (Yu et al. 2012). 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 $MgCl_2$, 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). 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). 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) and sequenced on an Illumina HiSeq 2500 using 100-bp paired-end reads.

MethylC-seq data processing

We used 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) (with the options -p 12 -N 1), which uses Bowtie 2 (Langmead and Salzberg 2012) 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 (**Supplementary 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; **Supplementary Figure 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) 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) 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). Using a paired design, we considered a gene as differentially expressed if statistically supported at a Benjamini and Hochberg (Hochberg 1995) false discovery rate (FDR) $< 1 \times 10^{-4}$ 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). 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):

$$\text{mean\_me} \ = \ 1/n \sum_{i=1}^{n} C_i/(C_i + T_i)$$

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). 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) 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 cutoff 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) 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 (Supplementary Figure 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 (Supplementary Figure 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) 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), 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) 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). 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) 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 β parameter in:

$$\log\left(\frac{\pi_l}{1-\pi_l}\right) = \alpha + \beta\, \mathrm{PWMscore}_l$$

where π_l represent the prior probability of binding in Centipede's model in motif location l. 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 π_{lt} for each condition t separately and for each site l. 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 (α and δ), the PWM effect β, and PWM times the treatment effect γ:

$$\log\left(\frac{\pi_{lt}}{1-\pi_{lt}}\right) = \alpha \times (1 - I_t) + \beta \times \text{PWMscore}_l + \delta \times I_t + \gamma \times (I_t \times \text{PWMscore}_l)$$

where I_t 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 γ, 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) with default parameters.

Gene Set Enrichment Analysis

We used Genetrail (Backes et al. 2007) 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) to control the False Discovery Rate (FDR).

Supplementary Figures
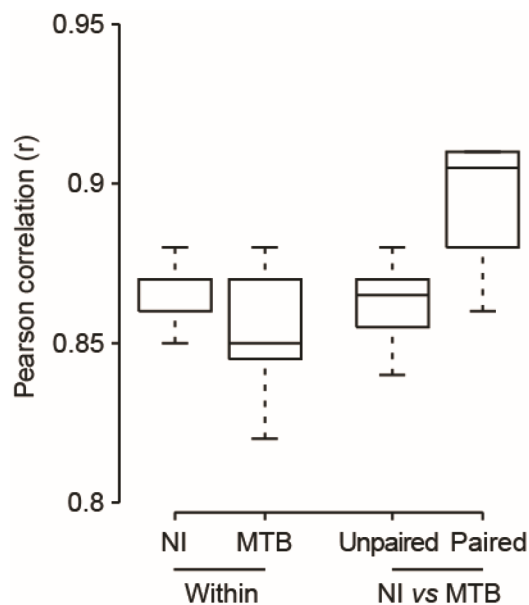
Supplementary Figure 1. Correlation between DNA methylation levels among replicates. Boxplots show pairwise comparisons of genome-wide methylation data (CpG sites with ≥ 5 coverage in all samples) between different individuals *within* either non-infected (NI) and MTB-infected (MTB) groups. Also shown 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).
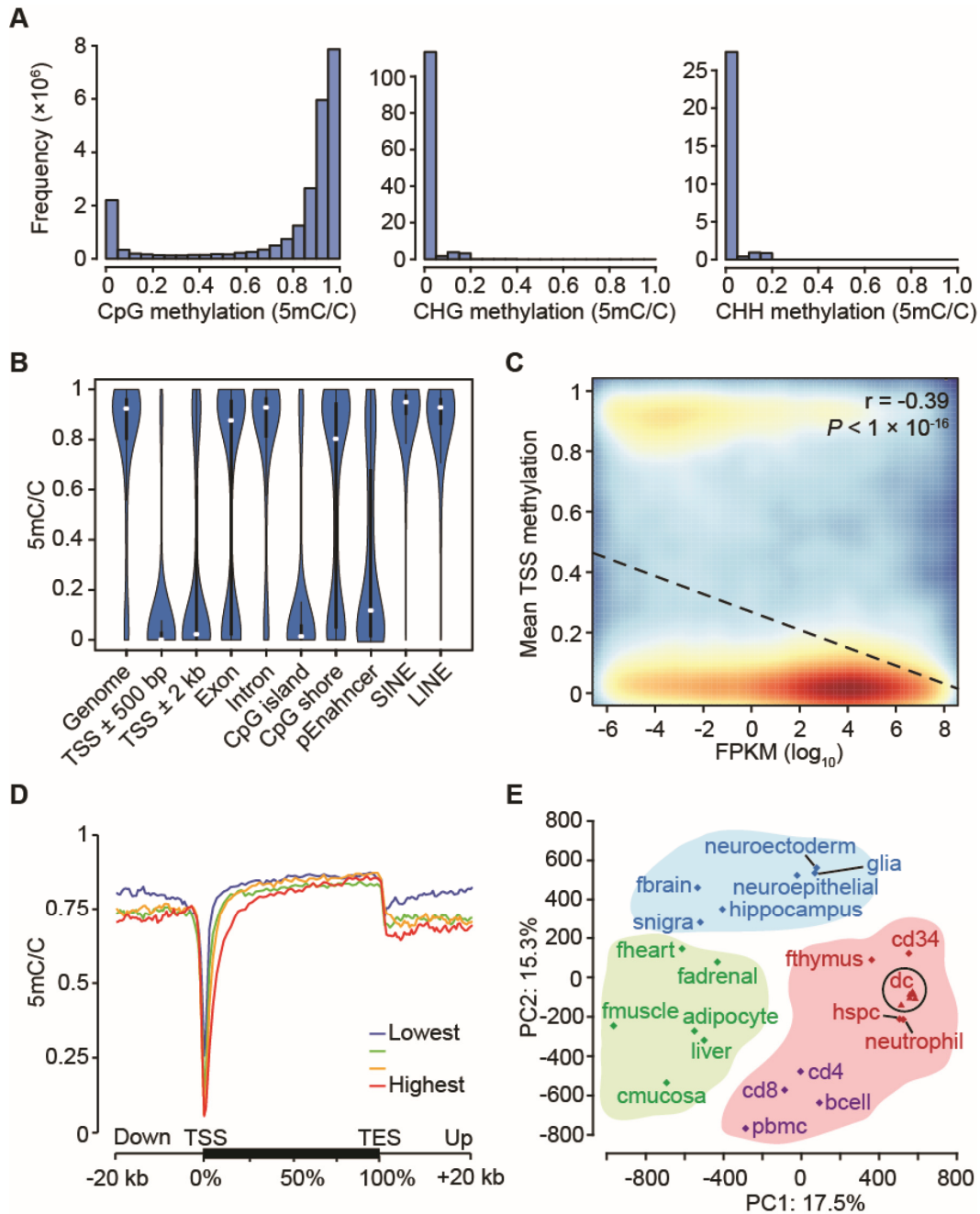
Supplementary Figure 2. Global patterns of DNA methylation in human DCs. (**A**) Distribution of genome-wide cytosine methylation in CpG, CHG and CHH (chromosome 1 only) contexts. (**B**) Violin plots showing the distribution of DNA methylation levels at CpG sites located in different genomic regions. 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 supplementary 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.
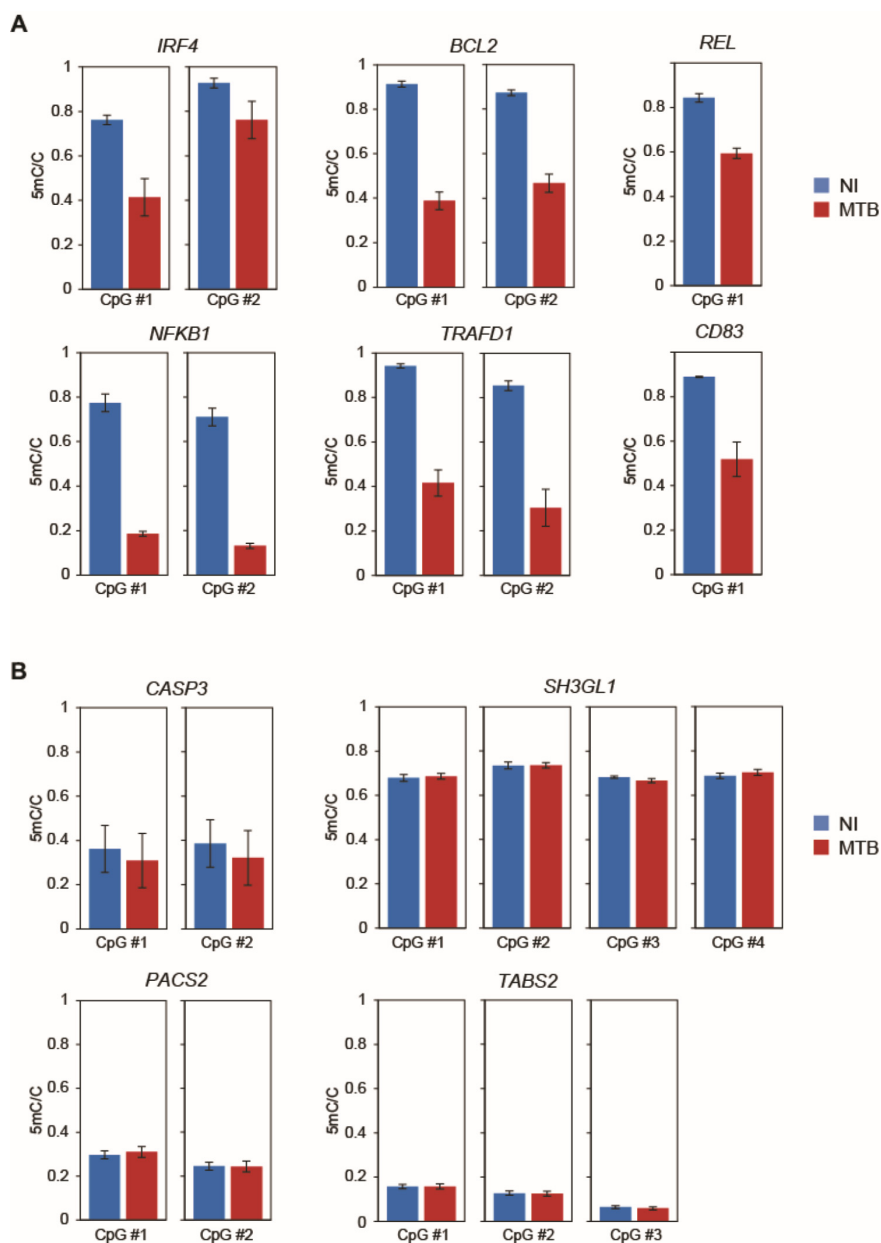
Supplementary Figure 3. Validation of differential methylation results by bisulfite pyrosequencing. (**A**) Bar plots of the average methylation levels in 10 CpG sites located within 6 hypo-DMRs and (**B**) 11 CpG sites within 4 hyper-DMRs (see Supplementary Table 3 for details). Blue and red boxes represent 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.
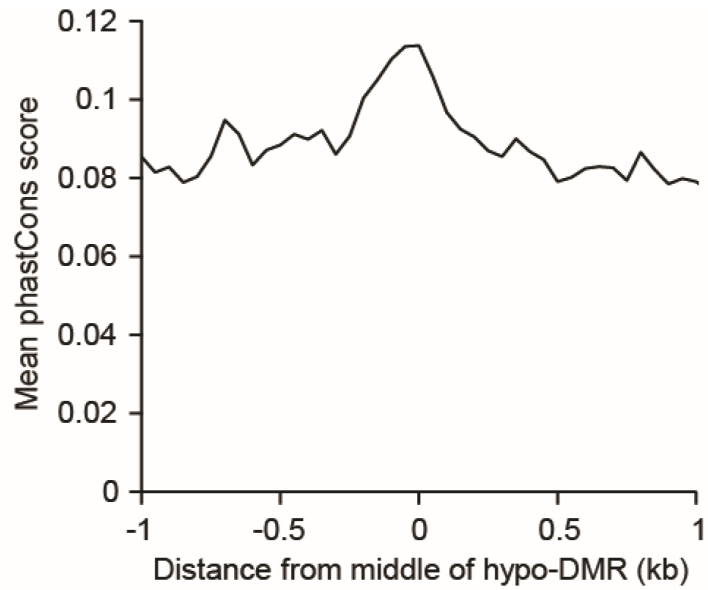
Supplementary Figure 4. Sequence conservation scores at MTB-DMRs. Average phastCons conservation score (Siepel et al. 2005) within 50-bp sliding windows, around the center of hypo-DMRs not associated with promoter regions (i.e., >3kb away from any known TSS).
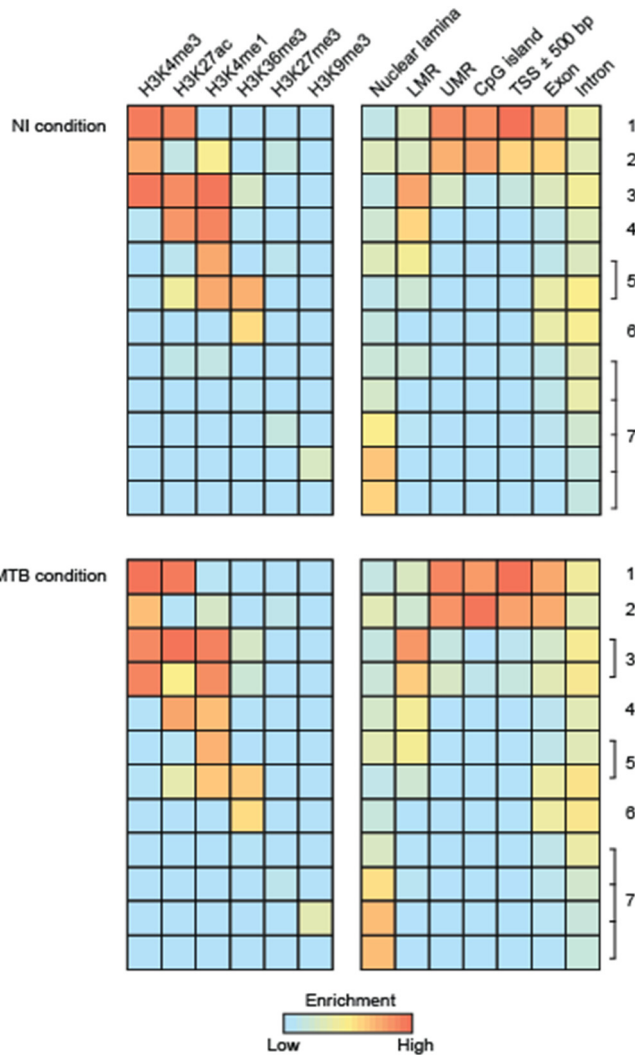
Supplementary Figure 5. Chromatin state annotation of infected and non-infected DC genomes. Six histone marks were used to learn a 12-state chromatin model separately for infected and non-infected DCs. Enrichment for emission probabilities for each mark (left panel) and different gene elements (right panel; scaled by column) within each state are 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).
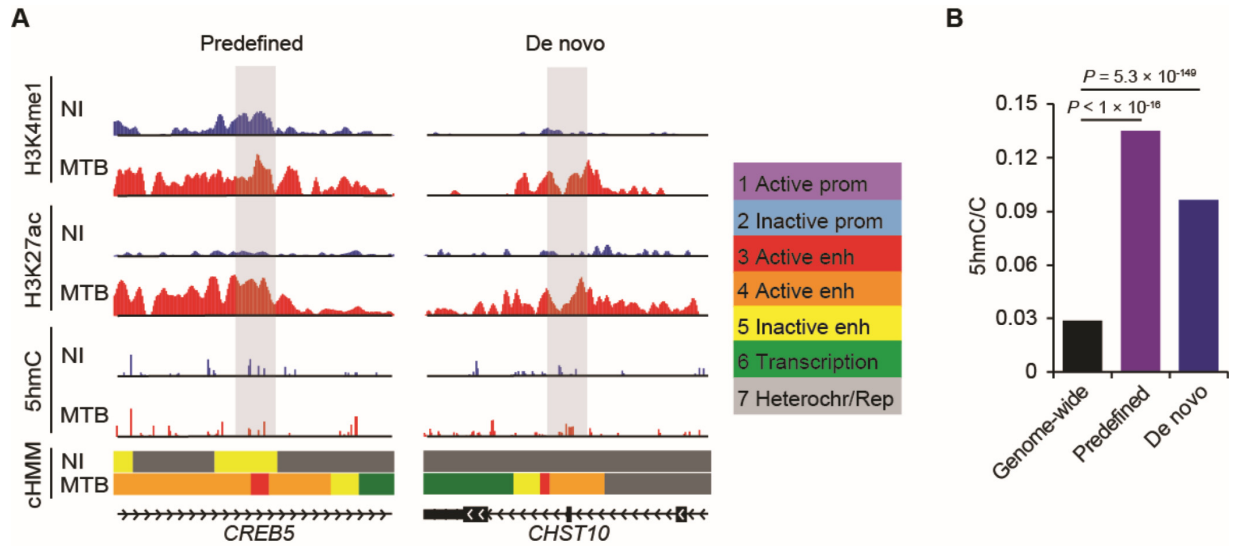
Supplementary Figure 6. Representative examples of a predefined and *de novo* enhancer at regions exhibiting loss in methylation (gray area). (**A**) Left panel: a predefined H3K4me1-marked enhancer region that subsequently gains activation marks (H3K27ac) after MTB infection. Right panel: a region that 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**) Bar plots of the hydroxymethylation levels in non-infected DCs showing that both predefined and *de novo* enhancers are enriched in 5hmC prior to MTB-infection.
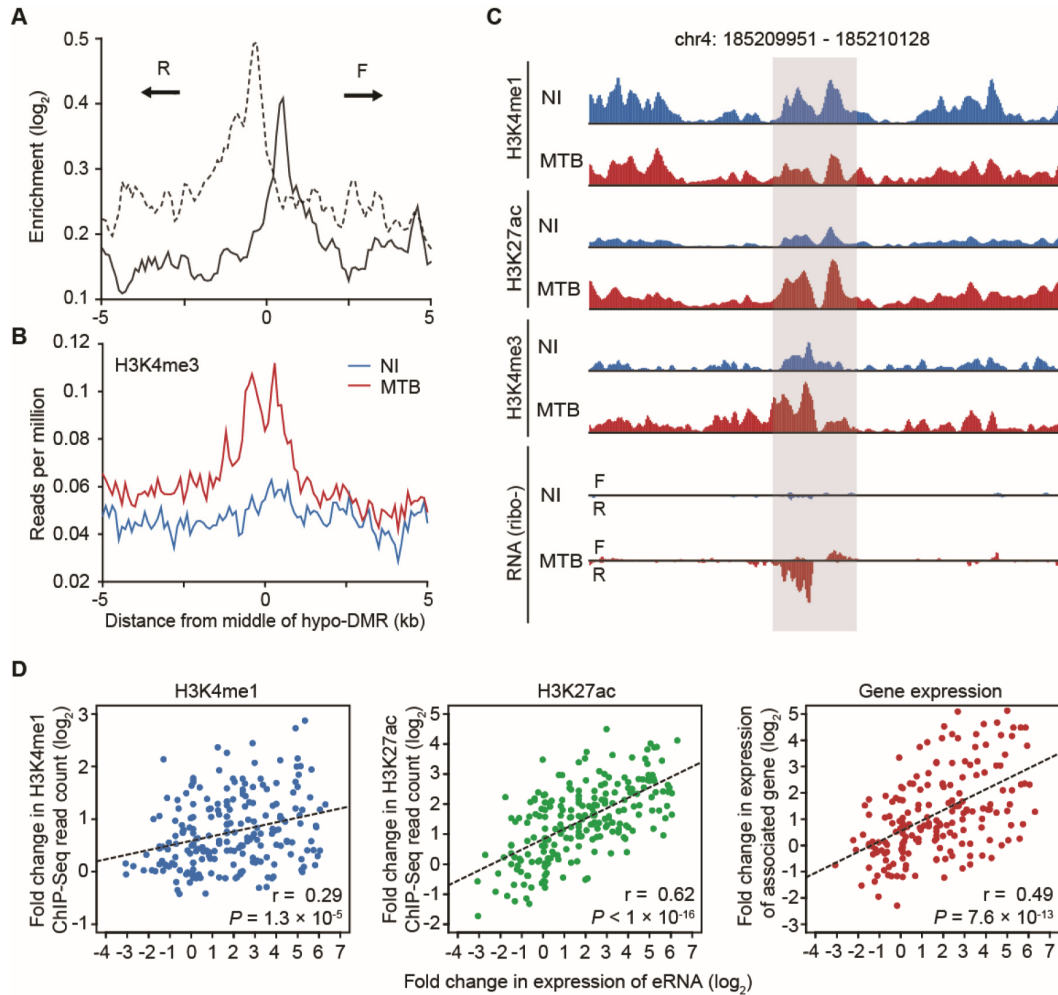
Supplementary Figure 7. Relationship between eRNA expression at hypomethylated regions and deposition of histone marks. (**A**) Fold changes (log2) in enhancer-associated RNA expression levels after MTB infection as a function of eRNA location relative to the center of the hypomethylated regions. (F) 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.
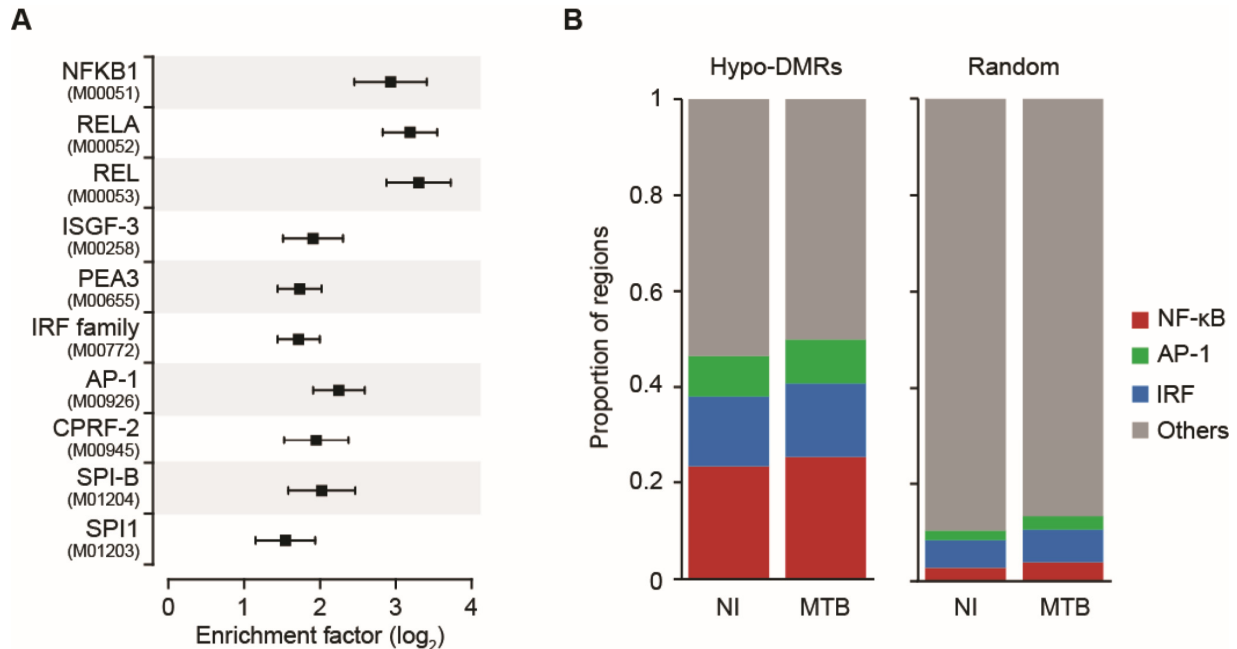
Supplementary Figure 8. MTB-DMRs are enriched for signal-dependent TF footprints.
(**A**) TF binding motifs for which the number of well-supported footprints (posterior Pr > 0.95) within hypomethylated regions were enriched relative to the genomic background, in MTB infected DCs. The enrichment factors 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 Supplementary 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.
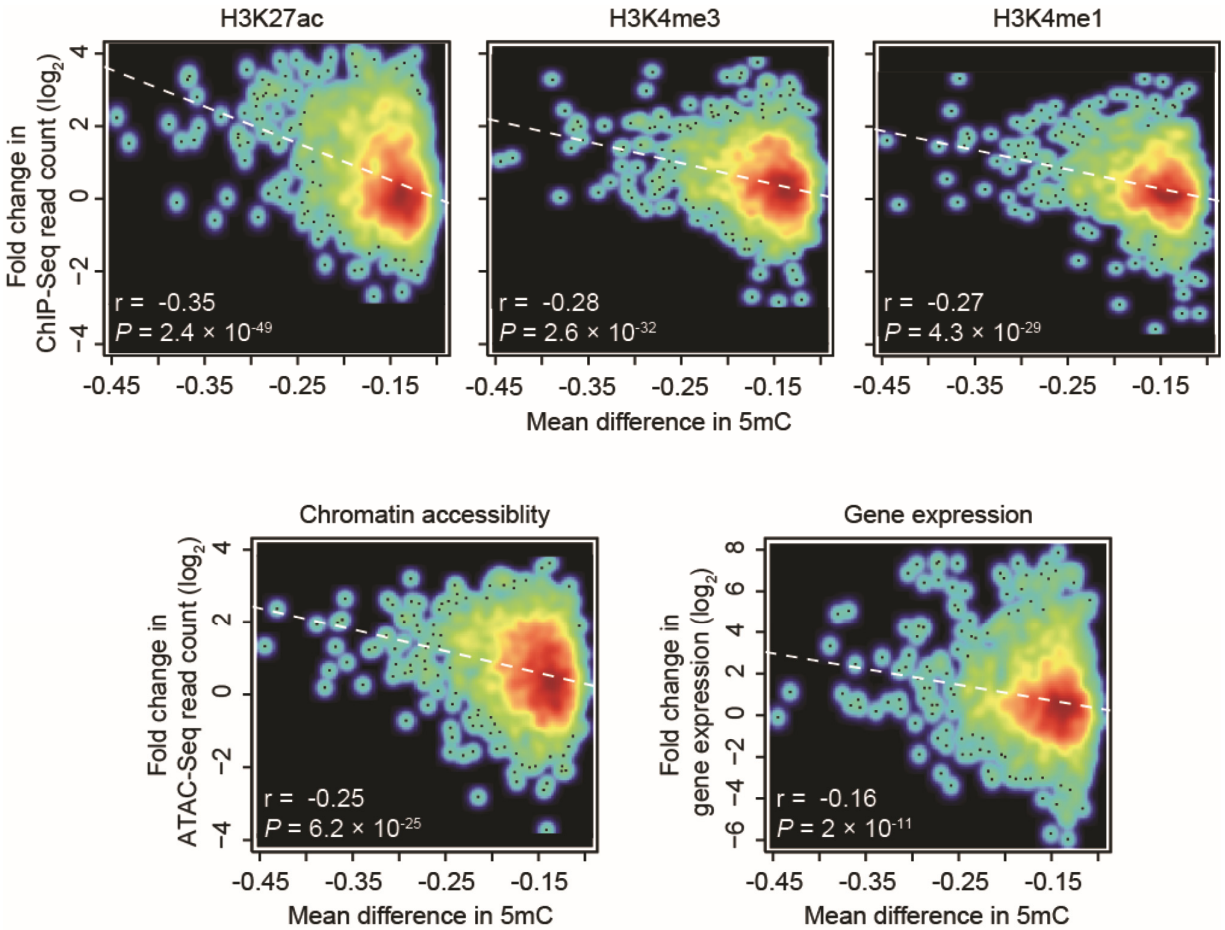
Supplementary Figure 9. Correlation between changes in DNA methylation and gene expression, and other epigenetic marks. Smooth scatterplots depicting the relationship between the magnitude of loss in DNA methylation level after infection and changes (in log2 scale) in H3K27ac, H3K4me3 and H3K4me1 histone marks, chromatin accessibility, and mRNA expression levels of nearby genes.
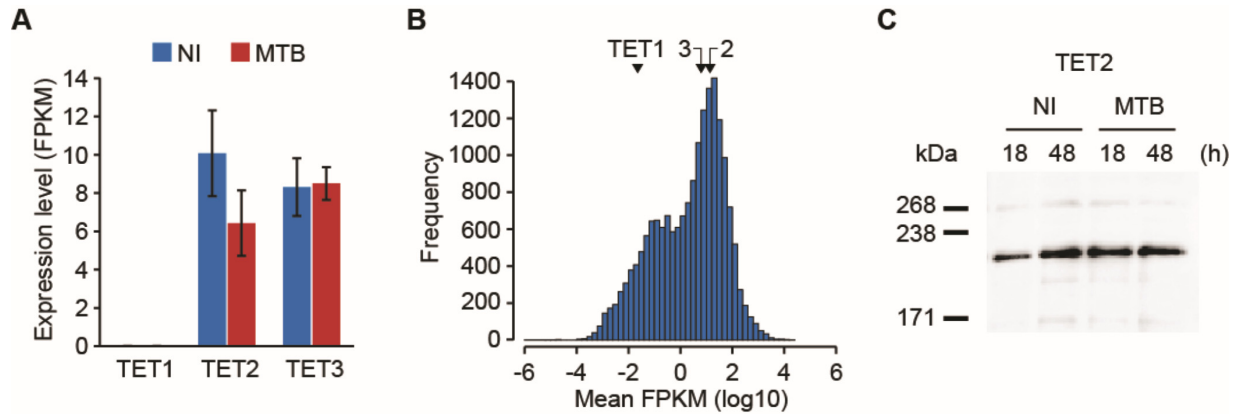
Supplementary Figure 10. Expression profiles of TET family of enzymes. (A) Bar plots showing the transcript levels (in FPKM) of *TET1*, *TET2* and *TET3* before (NI) and 18 hours post-infection (MTB) using RNA-seq. Data are represented as mean ± s.e.m., n = 6. **(B)** Distribution of mean baseline gene 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).

Enhancer					
t-stat n	2.131	2.776	3.365	4.032	4.773
3	5.9	6.1	6.3	7.4	8.7
4	6.4	6.5	6.8	8.4	10.0
5	6.7	7.0	7.9	9.8	11.1
6	6.9	7.7	8.5	10.3	11.9
7	7.1	9.0	9.3	12.0	12.3

DE gene					
t-stat n	2.131	2.776	3.365	4.032	4.773
3	1.3	1.3	1.4	1.6	1.8
4	1.4	1.6	1.6	1.7	2.0
5	1.5	1.4	1.7	2.0	2.3
6	1.7	1.7	1.8	2.3	2.6
7	1.8	1.7	2.0	2.5	2.8

Supplementary 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.
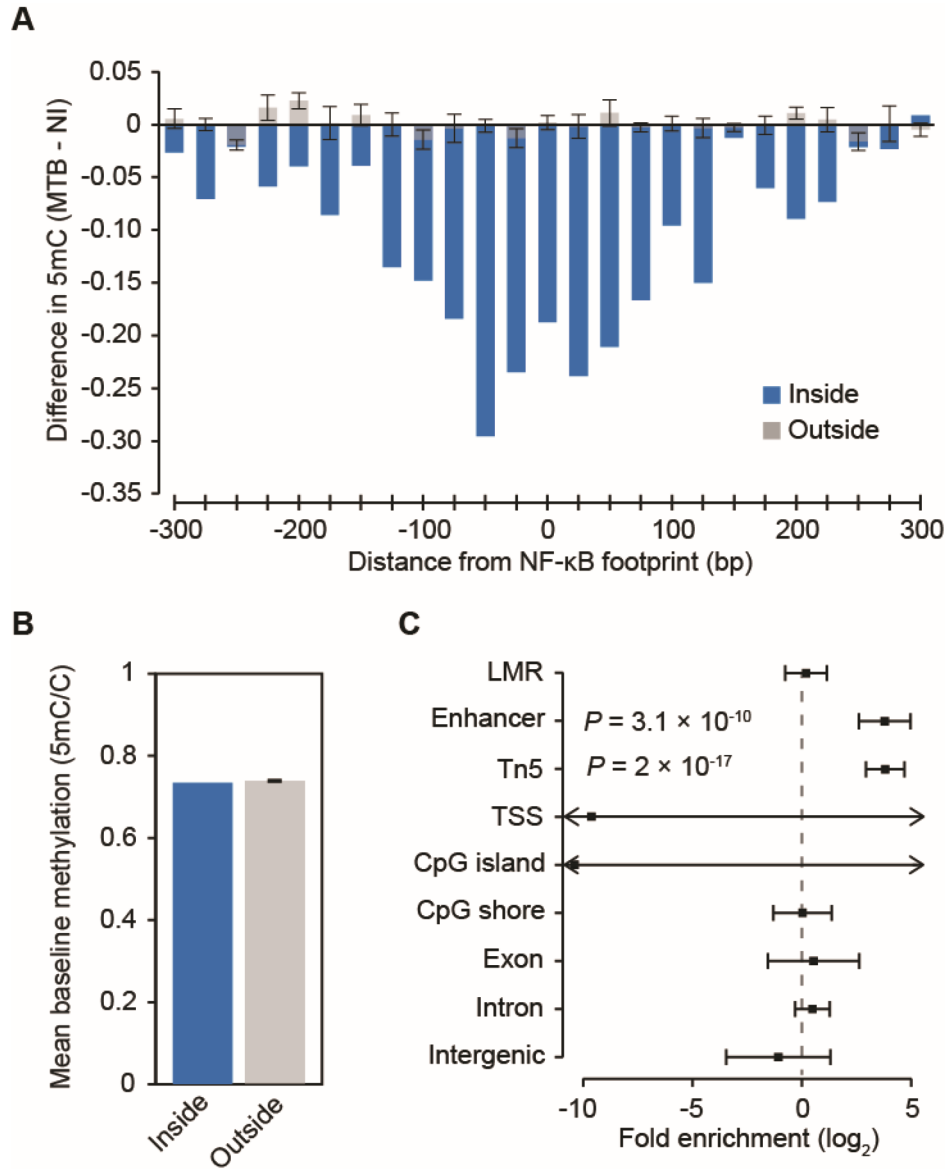
Supplementary Figure 12. TF binding alone is not sufficient to induce loss in methylation levels. (**A**) Changes in methylation ±300 bp around sites (25-bp sliding window), inside (blue) and outside (gray) hypomethylated regions, where NF-κB binding events occurred after infection (i.e., posterior probability 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.
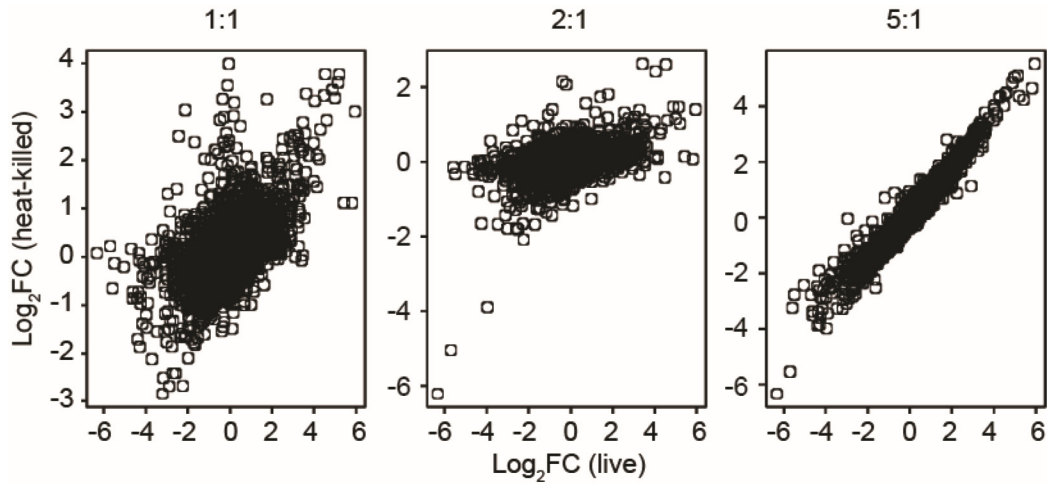
Supplementary Figure 13. Transcriptional responses using live MTB and heat-killed bacteria at different ratios. Scatterplots depicting the correlation between the log2 fold changes (log$_2$FC) in gene expression levels using live MTB and heat-killed bacteria at different ratios.
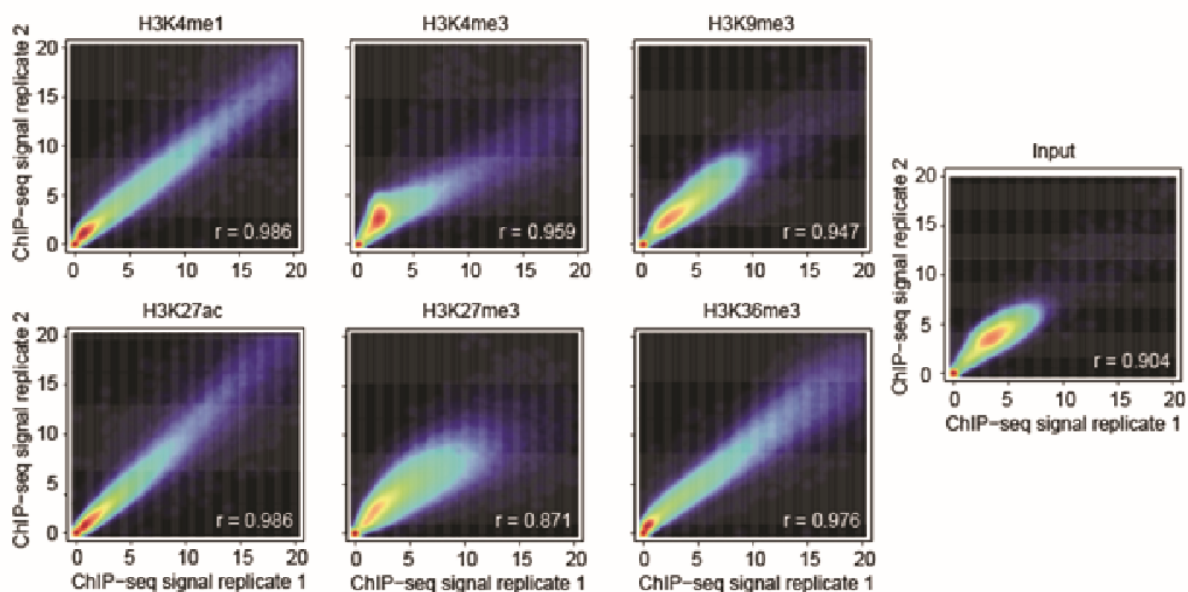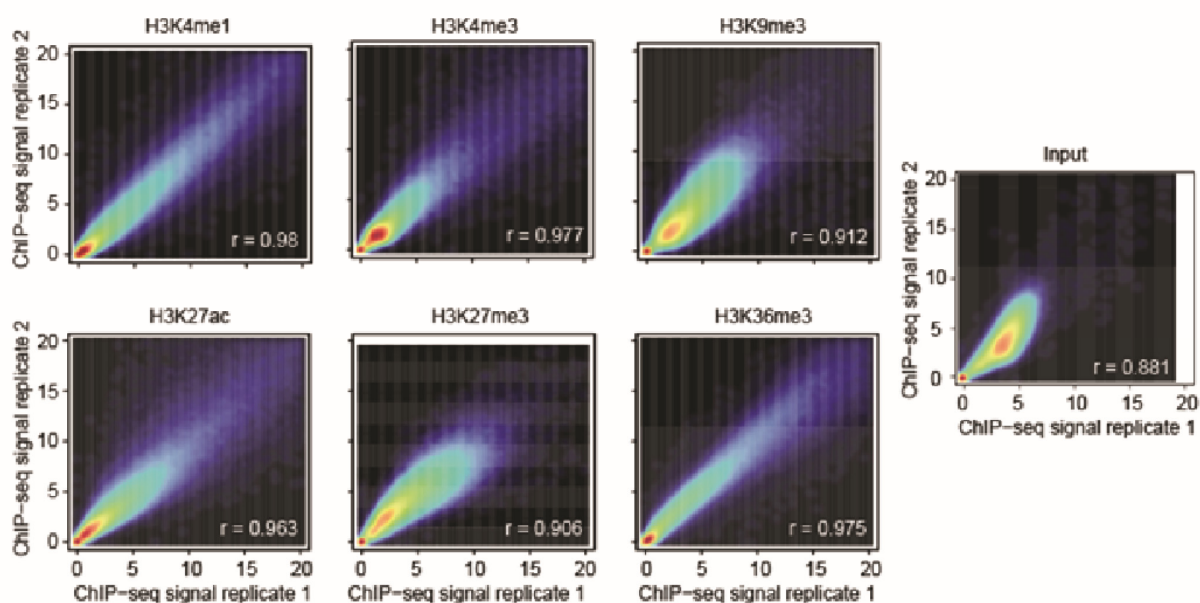
Supplementary Figure 14. Correlation of ChIP-seq signals for each histone mark between biological replicates. Smoothed scatterplots showing the correlation of ChIP-seq signals within 10000-bp sliding windows for each histone mark between 2 biological replicates.
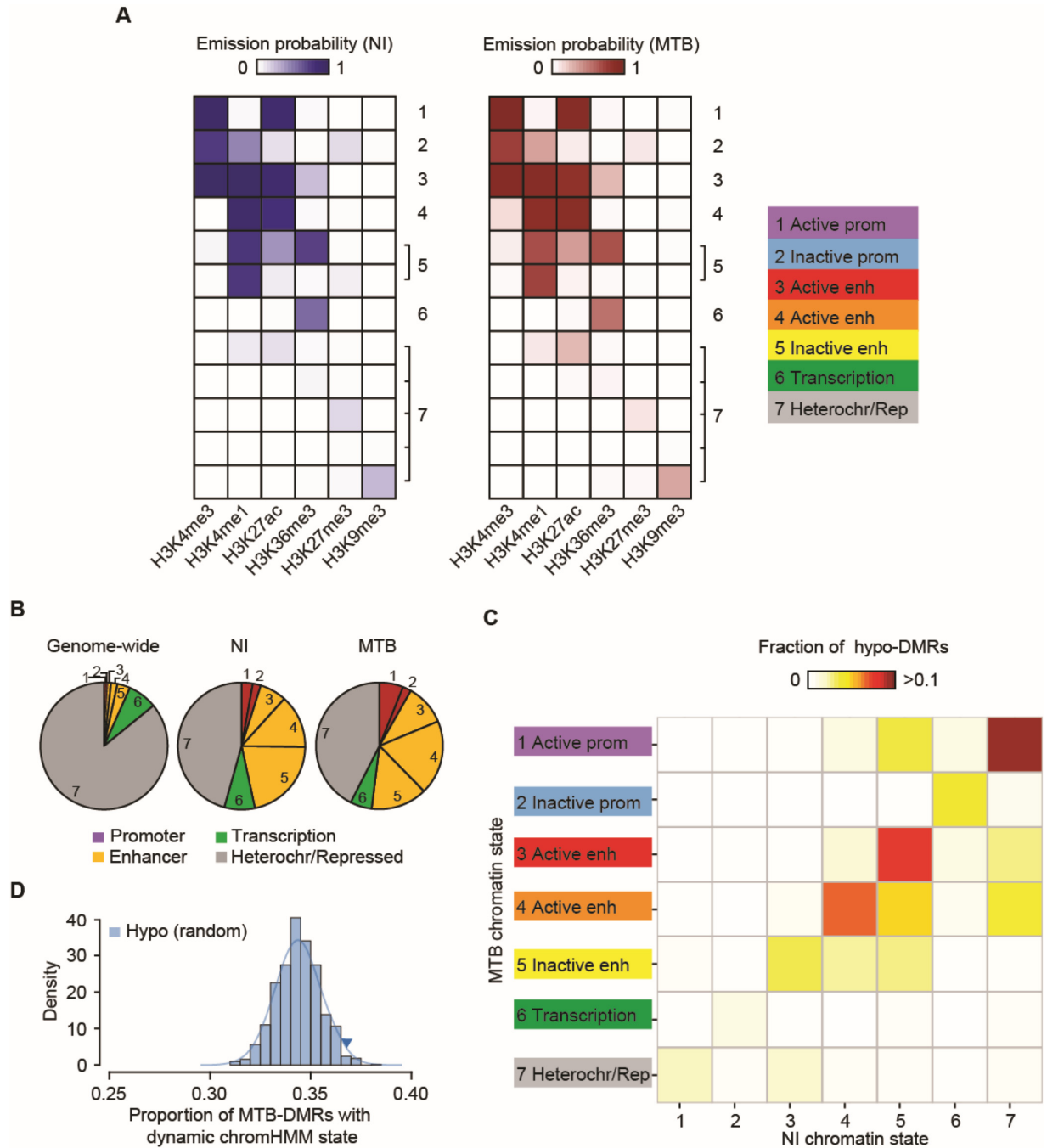
Supplementary Figure 15. Chromatin state emission probabilities and characterization of dynamic changes in histone marks in MTB-DMRs, based on a unified ChromHMM model. (**A**) ChromHMM emission probabilities for each chromatin state using a unified ChromHMM model that was learned from both infected 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.
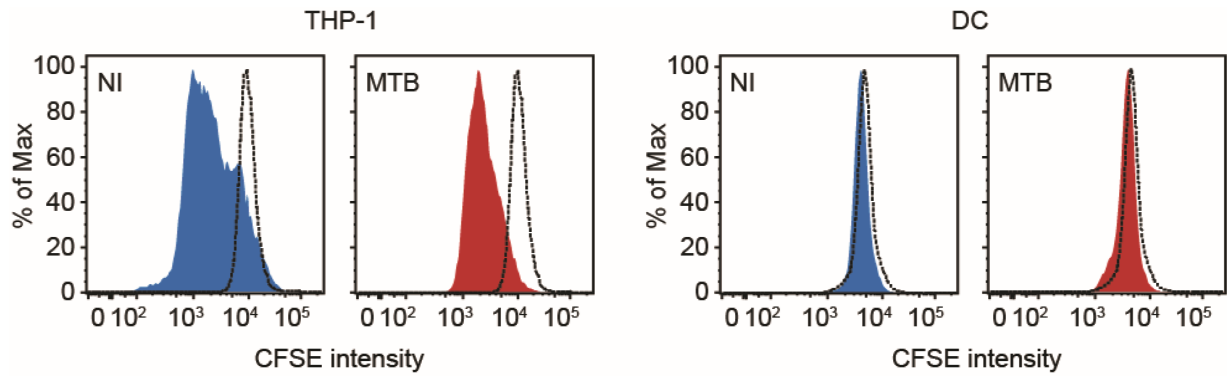
Supplementary Figure 16. Proliferation assays in non-infected cells or cells infected for 48 hours with MTB. CFSE-labeled THP-1 (left) and CFSE-labeled DCs (right).

3 Article II

DNA demethylation plays a limited role in the regulation of innate immune responses to infection

Pacis A, Mailhot-Léonard F, Tailleux L, Randolph HE, Yotova V, Dumaine A, Grenier JC, Barreiro LB.

Manuscript in preparation

N.B. Please note that due to space limitations, this section does not include supplementary tables 1 to 5. Complete tables can be found under the following link: https://www.dropbox.com/home/Public/Article2_SuppTables.

DNA demethylation plays a limited role in the regulation of innate immune responses to infection

Alain Pacis[1,2‡], Florence Mailhot-Léonard[1,2‡], Ludovic Tailleux[3], Haley E Randolph[1,2], Vania Yotova[1], Anne Dumaine[1], Jean-Christophe Grenier[1], Luis B Barreiro[1,4*]

[1]CHU Sainte-Justine Research Center, Department of Genetics, Montreal, H3T1C5, Canada; [2]University of Montreal, Department of Biochemistry, Montreal, H3T1J4, Canada; [3]Institut Pasteur, Mycobacterial Genetics Unit, Paris, 75015, France; [4]University of Montreal, Department of Pediatrics, Montreal, H3T1J4, Canada.

‡ These authors equally contributed to this study.
*Correspondence to: luis.barreiro@umontreal.ca

Abstract

DNA methylation is considered to be a relatively stable epigenetic mark. Yet, a growing body of evidence indicates that DNA methylation levels can rapidly change, for example, in innate immune cells facing an infectious agent. Nevertheless, the causal relationship between changes in DNA methylation and gene expression during infection remains to be elucidated. Here, we generated time course data on DNA methylation, gene expression, and chromatin accessibility patterns during infection of human post-mitotic dendritic cells (DCs) with *Mycobacterium tuberculosis*. We found that the immune response to infection is accompanied by active demethylation of thousands of CpG sites overlapping distal enhancer elements. However, virtually all changes in gene expression in response to infection occur prior to detectable changes in DNA methylation, indicating that the observed loss in methylation is a downstream consequence of transcriptional activation. Footprinting analysis revealed that immune-related transcription factors (TF), such as NF-κB/Rel, are recruited to enhancer elements prior to the observed loss in methylation, suggesting that DNA demethylation is mediated by TF binding to cis-acting elements. Collectively, our results show that DNA demethylation is likely not required for the establishment of the core regulatory program engaged upon infection.

Introduction

Innate immune cells, such as dendritic cells (DCs) and macrophages, are the first mediators recruited in response to an invading pathogen. Upon stimulation, these cells undergo pervasive changes in their transcriptional program, activating hundreds of genes involved in immune-related processes in a rapid and highly choreographed fashion. This is achieved through the binding of signal-dependent transcription factors (TFs), including NF-κB/Rel, AP-1, and interferon regulatory factors (IRFs), to gene regulatory regions of the genome where recruitment of various co-activators is initiated (Medzhitov and Horng 2009; Smale 2010). Alterations to the epigenome, such as histone modifications and DNA methylation, are recognized as important permissive or suppressive factors that play an integral role in modulating access of TFs to cis-acting DNA regulatory elements via the regulation of chromatin dynamics. Consequently, changes to the epigenetic landscape are expected to have a significant impact on gene expression.

Many studies have highlighted the paramount importance of histone modifications concerning the regulation of complex gene expression programs underlying immune responses (Bierne et al. 2012; Smale et al. 2014). However, the exact role that DNA methylation plays in innate immune response regulation remains ambiguous. We have previously shown that infection of post-mitotic DCs is associated with an active loss of methylation at enhancers and that such demethylation is strongly predictive of changes in expression levels of nearby genes (Pacis et al. 2015). To date, many other studies correlate these two processes (Bruniquel and Schwartz 2003; Murayama et al. 2006; Marr et al. 2014; Zhang et al. 2014; Ichiyama et al. 2015; Sinclair et al. 2015; Cizmeci et al. 2016; Wiencke et al. 2016), but it remains unclear whether altered methylation patterns directly invoke transcriptional modulation or whether such patterns are the downstream consequence of TF binding to regulatory regions. Thus, the causal relationship between changes in DNA methylation and gene expression during infection remains unresolved. To address this question, we characterized in parallel genome-wide patterns of DNA methylation, gene expression, and chromatin accessibility in non-infected and MTB-infected DCs at multiple time points. Our results show that the loss of DNA methylation observed in response to infection is not functionally required for the activation of most enhancer elements

but that, instead, the demethylation seen is a downstream consequence of TF binding followed by transcriptional induction.

Results

Bacterial infection induces stable loss of DNA methylation at enhancers of dendritic cells

To investigate the relationship between changes in DNA methylation and gene expression and in response to infection, we infected monocyte-derived DCs from 4 healthy individuals with a live virulent strain of *Mycobacterium tuberculosis* (MTB) for 2-, 18-, 48-, and 72-hours. At each time-point, we obtained single base-pair resolution DNA methylation levels for over 130,000 CpG sites using a customized capture-based bisulfite sequencing panel (SeqCap Epi, see Methods), in matched non-infected and MTB-infected DCs. Our customized SeqCap Epi panel interrogates 33,059 regions highly enriched among putative enhancer elements (58% are associated with the H3K4me1 enhancer mark (Heintzman et al. 2009); **Supplementary Figure 1A**), which are the main targets of methylation changes in response to infection (Pacis et al. 2015). In total, we generated ~717 million single-end reads (mean = 17.5 million reads per sample; **Supplementary Table 1**), resulting in an average coverage of ~70X per CpG site (**Supplementary Figure 1B**). Methylation values between samples were strongly correlated, attesting to the high quality of the data (**Supplementary Figure 1C**; median r across all samples = 0.94).

We next assessed temporal changes in methylation levels in response to infection using the DSS software (Feng et al. 2014). We defined differentially methylated (DM) CpG sites as those showing a significant difference of methylation between infected and non-infected samples at a False Discovery Rate (FDR) < 0.01 and an absolute mean methylation difference above 10%. Using these criteria, we identified 6,174 DM CpG sites across the time course of infection. Consistent with previous findings (Pacis et al. 2015), the vast majority of changes in methylation (87%) were associated with the loss of DNA methylation in infected cells (**Figure 1A,B**).

To test if live bacteria were required to induce the observed changes in DNA methylation, we collected similar data on DCs exposed to heat-killed MTB in addition to the live MTB experiments. Changes in methylation in response to live and heat-killed MTB were strikingly correlated, particularly at the later time-points post-infection (r ≥ 0.84 at 18h and above;

Supplementary Figure 2). These results show that DCs do not require exposure to a live pathogen to elicit the overall loss in methylation detected in response to infection. Simply, the engagement of innate immune receptors and activation of pathways involved in pathogen sensing and elimination is sufficient to induce shifts in methylation. Hierarchical clustering analysis of the DM sites observed when considering samples exposed to either live or heat-killed bacteria shows that >80% of the sites exhibited gradual loss of methylation over the time course of infection until methylation marks were almost completely erased, and that very few changes are detectable at 2 hours post-infection (DM Cluster 3; **Figure 1C,D**; **Supplementary Table 2**).

Monocyte-derived DCs do not proliferate in response to infection (Pacis et al. 2015) and, therefore, any observed loss in methylation must occur through an active mechanism involving the ten-eleven translocation (Tet) enzymes, a family of enzymes that converts 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) (Wu and Zhang 2017). Thus, we used Tet-assisted bisulfite sequencing (TAB-seq) data collected from non-infected DCs (Pacis et al. 2015) to assess if DM sites had significantly different levels of 5hmC as compared to non-DM sites. We found that DM sites (Cluster 3) show strikingly high levels of 5hmC even prior to infection (**Figure 1E**; 3.2-fold enrichment compared to non-DM sites; Wilcoxon test; $P < 1 \times 10^{-16}$), suggesting that DM sites are likely pre-bound by TET enzymes (likely TET2 (Klug et al. 2013; Alvarez-Errico et al. 2015), the most expressed Tet enzyme in DCs (**Supplementary Figure 3**)), and that 5hmC may serve as stable mark that allows priming of these enhancers (Yu et al. 2012; Calo and Wysocka 2013; Mahe et al. 2017).
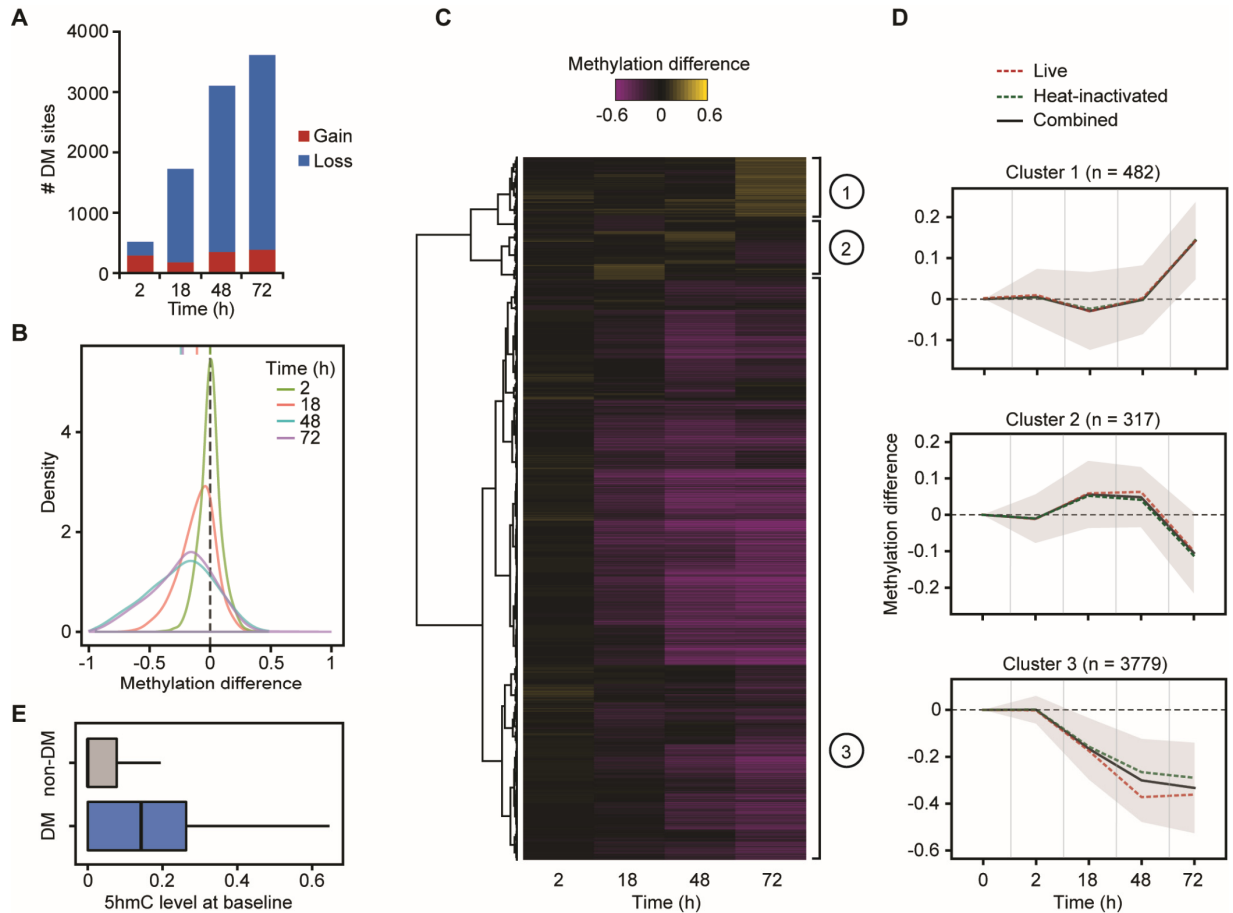
Figure 1. DNA methylation dynamics in DCs during MTB infection. (**A**) Bar plots showing the number of differentially methylated (DM) CpG sites identified at a |methylation difference| > 10% and FDR < 0.01 (y-axis) at each time point after MTB infection (2, 18, 48 and 72 hours (h)) (x-axis). (**B**) Distribution of differences in methylation between infected and non-infected cells at DM sites, at each time point. (**C**) Heatmap of differences in methylation constructed using unsupervised hierarchical clustering of the 4,578 DM sites (identified at any time point using live and heat-inactivated MTB-infected samples combined; y-axis) across four time points after infection, shows three distinct patterns of changes in methylation. (**D**) Mean differences in methylation of CpG sites in each cluster across all time points; shading denotes ±1 standard deviation. For visualization purposes, we also show the '0h' time point, where we expect no changes in methylation. (**E**) Boxplots comparing the distribution of 5hmC levels in non-infected DCs between non-DM and DM sites (Cluster 3).

Up-regulation of inflammatory genes precedes DNA demethylation

We collected RNA-seq data from matched non-infected and infected samples at each time point, for a total of 34 RNA-seq profiles across time-treatment combinations (mean = 42.2 million reads per sample; **Supplementary Table 1**). The first principal component of the resulting gene expression data accounted for 63% of the variance in our dataset and separated infected and non-infected DCs (**Supplementary Figure 4A**). We found extensive differences in gene expression levels between infected and non-infected DCs: of the 13,956 genes analyzed, 1,987 (14%), 4,371 (31%), 4,591 (33%), and 5,189 (37%) were differentially expressed (DE) at 2, 18, 48 and 72 hours post-infection, respectively (FDR < 0.01 and absolute log2(fold change) > 1; **Supplementary Table 3**). We also collected RNA-seq data in samples stimulated with heat-inactivated MTB and found that, similar to changes in methylation, changes in gene expression in response to live and heat-inactivated MTB were strongly correlated ($r \geq 0.94$; **Supplementary Figure 4B**). We next grouped the set of DE genes across the time course (7,457 in total) into 6 distinct temporal expression clusters (**Figure 2A,B**; **Supplementary Table 3**). These clusters cover a variety of differential expression patterns, including genes which show increasing up-regulation over time (DE Cluster 5: Persistent induced; n = 2,091) to genes in which the highest levels of expression occur at 2 or 18 hours followed by a decrease towards basal levels (DE Cluster 4: Early induced (n = 765), and DE Cluster 6: Intermediate induced (n = 839), respectively) (**Figure 2B**). Gene ontology (GO) enrichment analysis revealed that induced genes were strongly enriched among GO terms directly related to immune function, including defense response (FDR = 1.2×10^{-11}) and response to cytokine (FDR = 8.2×10^{-12}), whereas repressed genes were primarily enriched for gene sets associated with metabolic processes (**Figure 2C**; **Supplementary Table 4**).

We next tested whether genes located near DM sites – particularly focusing on those sites exhibiting stable loss of methylation (i.e., Cluster 3 in Figure 1C,D) – were more likely to be differentially expressed upon MTB infection relative to all genes in the genome. We found that genes associated with one or more DM sites were strongly enriched among genes that were up-regulated in response to infection, regardless of the time point at which expression levels started

to change: early (2.5-fold, $P = 3.23 \times 10^{-11}$), intermediate (3.5-fold, $P = 3.59 \times 10^{-25}$) and persistent (3.1-fold, $P = 3.80 \times 10^{-33}$) (**Figure 2D,E**).

If demethylation is required for the activation of enhancer elements and the subsequent up-regulation of their target genes, we would expect demethylation to occur *prior* to changes in gene expression; instead, we found the opposite pattern. Among up-regulated genes associated with DM sites (n = 593), 37% exhibited at least a two-fold increase in gene expression levels at 2-hours post-infection, although differential methylation did not begin to be detectable until 18-hours post-infection (**Figure 2E**). For only 17 genes (less than 3% of all up-regulated genes associated with DM sites), DNA demethylation occurred prior to gene induction (**Supplementary Figure 5**). Thus, DNA demethylation doesn't appear to be necessary for gene activation.

To confirm that our findings were generalizable to other innate immune cell types and pathogenic infections, we performed a separate time course analysis of differential methylation in Salmonella-infected macrophages from one additional donor over six time-points (**Supplementary Table 1**). We discovered hundreds of CpG sites that exhibited a progressive loss of methylation over the time course of infection, corroborating our findings in MTB-infected DCs (**Figure 3A**). To assess whether loss of methylation arises after the activation of associated enhancers, we collected ChIP-seq data for acetylation of histone 3 lysine 27 (H3K27ac) at 2-hours post-infection, as changes in DNA methylation have yet to occur at this point. We found that the deposition of activating H3K27ac marks preceded demethylation at these CpG sites (**Figure 3B**). Moreover, using previously published RNA-seq expression data from Salmonella-infected macrophages (Nedelec et al. 2016), we found that most genes associated with these sites were up-regulated at 2-hours post-infection (**Figure 3C**) prior to any changes in methylation. Collectively, these findings indicate that DNA demethylation is not required for the activation of most enhancer elements and that the vast majority of methylation changes induced by infection are a downstream consequence of transcriptional activation.
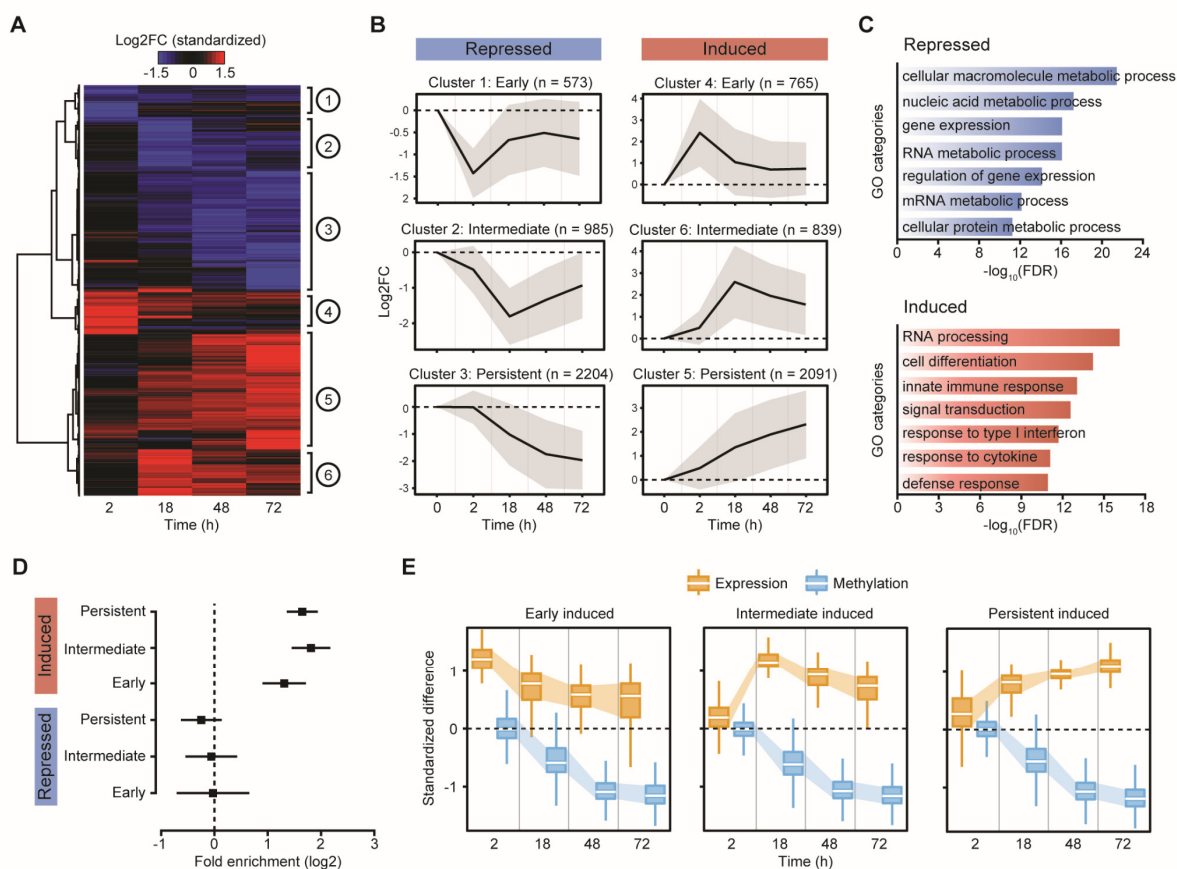
Figure 2. Time course analysis of changes in DNA methylation and gene expression in DCs in response MTB infection. (**A**) Heatmap of differences in expression (standardized log2 fold changes) constructed using unsupervised hierarchical clustering of the 7,457 differentially expressed genes (identified at any time point using cutoffs of |log2FC| > 1 and FDR < 0.01; y-axis) across four time points after MTB infection results in six distinct patterns of changes in expression. (**B**) Mean log2 fold expression changes of genes in each cluster across all time points; shading denotes ±1 standard deviation. For visualization purposes, we also show the '0h' time point, where we expect no changes in expression. (**C**) Gene ontology enrichment analyses among genes that are repressed or induced in response to MTB infection. (**D**) Enrichment (in log2; x-axis) of differentially expressed genes associated with Cluster 3 DM CpG sites. Error bars show 95% confidence intervals for the enrichment estimates. (**E**) Boxplots showing the distribution of standardized differences in methylation of Cluster 3 DM sites (blue) along with

the corresponding standardized differences in expression of the associated genes (orange), across all time points.
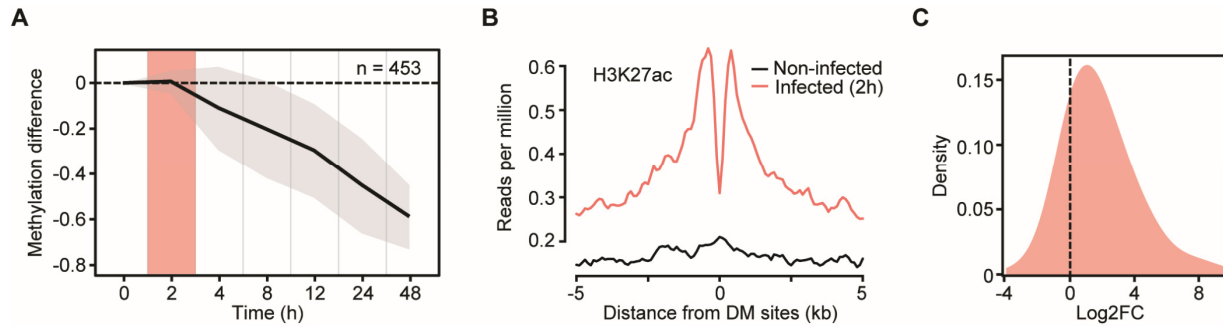
Figure 3. Relationship between changes in DNA methylation and gene expression in macrophages in response to *Salmonella* infection. (**A**) Mean differences in methylation (y-axis) in CpG sites that show stable loss of methylation (similar to Cluster 3 DM sites in Figure 1C,D; n = 453) in Salmonella-infected macrophages, across six time points after infection (2, 4, 8, 12, 24 and 48 hours (h); x-axis). Shading denotes ±1 standard deviation. For visualization purposes, we also show the '0h' time point, where we expect no changes in methylation. (**B**) Composite plots of patterns of H3K27ac ChIP-seq signals ±5 kb around the midpoints of hypomethylated sites (x-axis) in macrophages at 2 hours post-infection with Salmonella. (**C**) Distribution of log2 fold expression changes (between non-infected and Salmonella-infected macrophages at 2 hours (Nedelec et al. 2016)) for genes associated with DM sites in Figure 3A (n = 269).

The binding of most infection-induced TFs does not require active demethylation

We next asked whether MTB-induced changes in gene expression were associated with changes in chromatin accessibility. To do so, we profiled regions of open chromatin in non-infected and infected DCs at the same time-points (plus one additional time-point at 24 hours) using ATAC-seq (Buenrostro et al. 2013). Overall, we found that the response to MTB infection was accompanied by an increase in chromatin accessibility across regulatory regions associated with genes up-regulated upon MTB infection, regardless of their expression profiles (**Figure 4A**). Interestingly, most increases in chromatin accessibility were observed at later stages of infection, suggesting that the activation of early response genes does not require significant changes to the chromatin structure.

To investigate the relationship between DNA methylation and TF occupancy, we performed TF footprinting analysis on our target regions (i.e., the set of putative enhancers tested for dynamic DNA methylation). We classified target regions as "hypomethylated regions" (n = 1,877) or "non-differentially methylated regions" (non-DMRs) (n = 31,182) according to whether or not these regions overlap DM CpG sites (from differential methylation Cluster 3, specifically). We found that hypomethylated regions were significantly enriched for the binding of immune-related TFs relative to regions exhibiting consistent methylation levels. These immune-related TFs include several master regulators of the innate immune response such as RelA (NF-κB/Rel family member; up to 8.3-fold enrichment across the time course; FDR $\leq 3.14 \times 10^{-20}$) and Stat5 (up to 2.2-fold enrichment across the time course; FDR ≤ 0.09) **Figure 4B**; **Supplementary Table 5**).

We next used CentiDual (Pacis et al. 2015) to test for differential binding of TFs between non-infected and infected samples, specifically focusing on the set of TF family members known to orchestrate innate immune responses to infection (i.e., NF-κB, AP-1, STATs and IRFs). We found increased binding at NF-κB/Rel binding motifs starting at 2-hours post-infection, despite the fact that no changes in methylation were observed at such early time points ($P = 0.03$; **Figure 4C**; **Supplementary Table 5**; see Methods). A similar trend, although non-significant, was

observed for AP-1 (**Supplementary Figure 6**). These data show that, while demethylated regions overlap areas bound by immune-induced TFs, the binding of these TFs occurs prior to DNA demethylation.

Although demethylation does not appear to be required for the binding of key TFs involved in regulation of innate immune responses, it is plausible that the removal of methylation marks at DM sites might enable occupancy of methylation-sensitive factors at later time points (Domcke et al. 2015; Zhu et al. 2016; Yin et al. 2017). In support of this hypothesis, we found at later time-points (18 hours and above) a stronger enrichment for the binding of TFs that preferentially bind to unmethylated motifs (or "methyl-minus" as defined by Yin et al. (Yin et al. 2017)) within hypomethylated regions (up to 1.8-fold enrichment; $\chi2$-test; $P \leq 2.90 \times 10^{-22}$; **Figure 4D**; see Methods). Collectively, these results suggest that, although demethylation is likely not required for the establishment of the core regulatory program engaged rapidly upon infection, it might play a role in fine-tuning the innate immune response by facilitating the binding of salient methyl-sensitive TFs that play a role in later responses to infection.
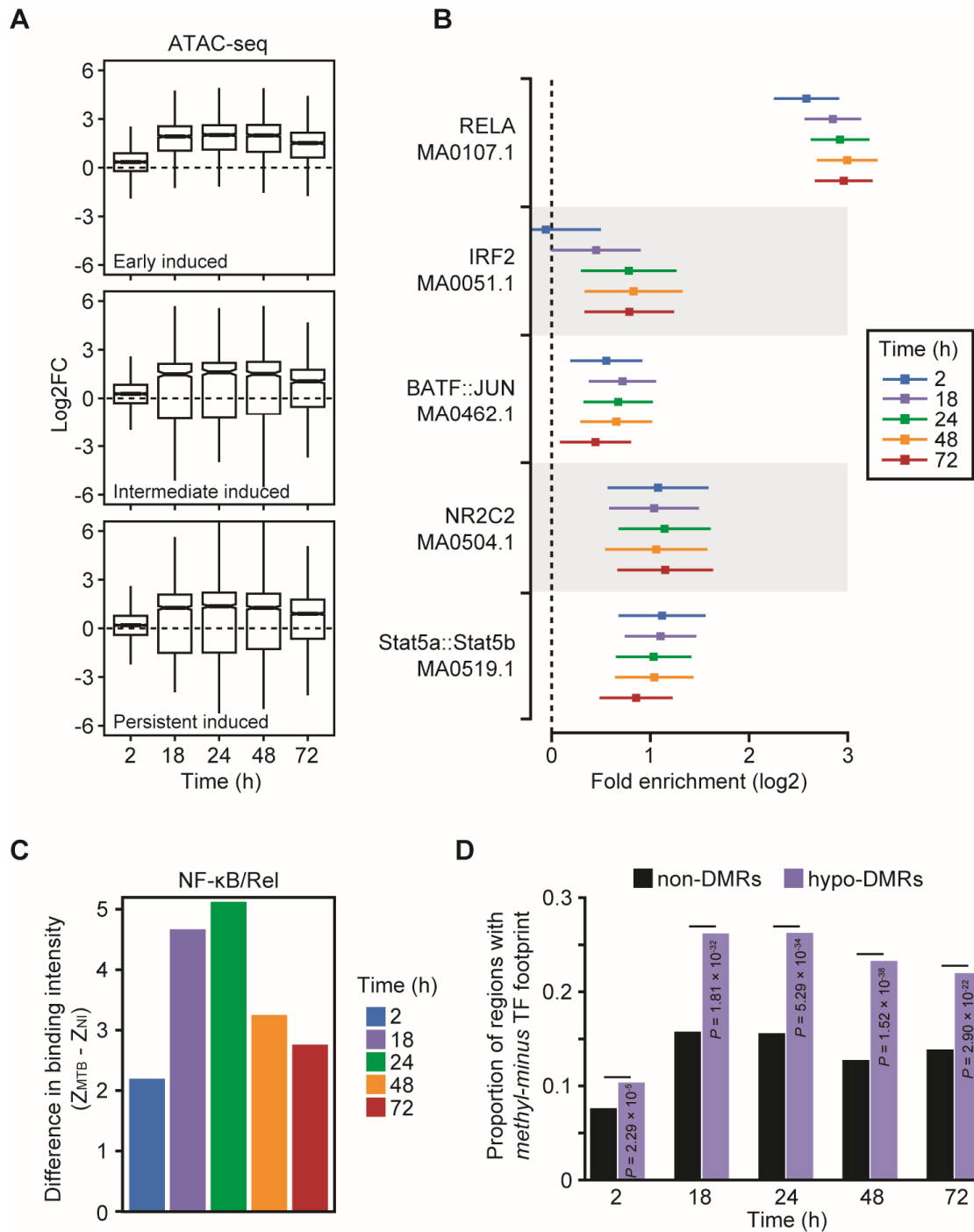
Figure 4. Relationship between changes in DNA methylation and transcription factor binding in DCs in response to MTB infection. (**A**) Boxplots showing the distribution of log2 fold changes in chromatin accessibility (between non-infected and MTB-infected DCs across the five time points of infection (2, 4, 18, 24, 48 and 72 hours) for open chromatin regions

associated with the three classes of induced genes described in Figure 2A,B. (**B**) TF binding motifs for which the number of well-supported footprints (posterior probability > 0.99) are enriched within hypomethylated regions were enriched (FDR < 0.01) relative to non-DMRs, in MTB-infected DCs. The enrichment factors (x-axis) are shown in a log2 scale and error bars reflect the 95% confidence intervals. A complete list of all TF binding motifs for which footprints are enriched within hypomethylated regions can be found in Supplementary Table 5. (**C**) Bar plots showing significant differences in TF occupancy score predictions for NF-κB/Rel motifs between MTB-infected and non-infected DCs (Z_{MTB} - Z_{NI}; y-axis; see Methods), across all time points (x-axis). A positive Z-score difference indicates increased TF binding in hypomethylated regions after MTB infection. (**D**) Proportion of regions that overlap a methylation-sensitive ("methyl-minus"; reported in Yin et al. (Yin et al. 2017)) TF footprint (y-axis) observed among non-DMRs and hypomethylated regions (hypo-DMRs; see Methods).

Discussion

In this study, we have generated paired data on DNA methylation, gene expression, and chromatin accessibility in non-infected and MTB-infected DCs at multiple time-points. Our results show that bacterial infection leads to marked remodeling of the methylome of phagocytic cells (both DCs and macrophages), with several thousand CpG sites showing stable loss of methylation via active DNA demethylation. Yet, only a small proportion of up-regulated genes associated with a DM site (less than 3%) exhibit a dynamic in which DNA demethylation precedes gene activation. Despite the limited resolution of our time course experiment, this observation strongly suggests that losses in methylation are a downstream consequence of TF binding and transcriptional activation, at least for early responsive genes.

The observed changes in methylation most likely occur via TET2-mediated active demethylation, as previously shown (Klug et al. 2013; Pacis et al. 2015; Vento-Tormo et al. 2016). Consistent with this hypothesis, we found that CpG sites that lose methylation upon infection display high levels of 5hmC at baseline, suggesting that these regions are actively bound by TET2 even prior to infection. Moreover, *TET2* is strongly upregulated 2 hours after infection (~2.5 fold, **Supplementary Figure 7**). 5hmC could be a stable intermediate and serve as an epigenetic priming mark, ensuring the rapid response of DCs against infection (Creyghton et al. 2010; Serandour et al. 2012; Yu et al. 2012; Calo and Wysocka 2013; Hon et al. 2014; Vento-Tormo et al. 2016; Mahe et al. 2017). Importantly, our SeqCap Epi data does not allow us to distinguish between 5mC and 5hmC. Thus, it is possible that DM sites that gain 5hmC, but do not show a loss of 5mC, at 2-hours post-infection may precede the activation of certain enhancers, as was recently suggested in T cells (Ichiyama et al. 2015). Further studies will be necessary to investigate the functional relevance of 5hmC in the induction of inflammatory genes during infection.

Using footprint analysis, we show that NF-κB/Rel, a master regulator of inflammation, is recruited to hypomethylated regions as soon as 2-hours post-infection. This finding is consistent with ChIP-seq data collected from macrophages stimulated with Kdo2-Lipid A (KLA), a highly specific TLR4 agonist, which shows that the NF-κB subunit p65 is rapidly recruited to enhancer

elements within one hour post-stimulation (Kaikkonen et al. 2013). We hypothesize that the rapid binding of NF-κB, and of other immune-induced TFs, likely instigates the opening of the chromatin, the recruitment of histone acetyltransferase p300, and the subsequent deposition of activating H3K27ac marks in these regions (Bhatt and Ghosh 2014). Interestingly, p300 can acetylate TET2, conferring enhanced enzyme activity (Zhang et al. 2017), which might account for the loss of DNA methylation in response to infection.

Our results indicate that most changes in gene expression that occur in response to infection are independent from DNA demethylation, further supporting the lack repressive capacity of DNA methylation (Ford et al. 2017). Notably, only a small proportion of up-regulated genes associated with a DM site exhibit a dynamic in which DNA demethylation precedes gene activation. Similar to previous findings (Han et al. 2001; Kress et al. 2006; Sato et al. 2006; Stadler et al. 2011; de la Rica et al. 2013; Schubeler 2015; Vento-Tormo et al. 2016), our results further reinforce the idea that site-specific regulation of DNA demethylation is mediated by TFs that bind to cis-acting sequences. Interestingly, several recent reports have shown that other epigenetic modifications, such as the H3K4me1 enhancer mark, have a similar passive regulatory function (Dorighi et al. 2017; Rickels et al. 2017; Vandenbon et al. 2017). However, our results do not exclude the possibility that demethylation might be necessary for the binding of a second wave of TFs that only play a role at later stages of infection (18 hours post-infection or later). In agreement with this hypothesis, we observed a significant enrichment for the binding of TFs that were shown to be sensitive to CpG methylation ("methyl-minus") in hypomethylated regions. Ultimately, this suggests that, although not a key regulator, DNA demethylation could play a role in fine-tuning immune responses by facilitating the binding of methylation-sensitive TFs at enhancers.

After an infection is cleared, TFs are expected to unbind and gene expression as well as DNA methylation levels are anticipated to return to basal state. However, our 72-hour time course study of DNA methylation showed that levels of methylation at DM sites gradually decreased with time post-infection but never reverted back to higher levels. Interestingly, this pattern was also observed for genes in which the largest fold changes in gene expression occurred at earlier time points. Thus, it is tempting to speculate that demethylation in response to infection could

have a specific biological role in trained immunity (or innate immune memory) (Ostuni et al. 2013; Quintin et al. 2014; Saeed et al. 2014; Kaufmann et al. 2018). Regions that lose methylation could act as "super-primed enhancers", potentially allowing for a faster response to a secondary infection. More generally, if DNA demethylation is seen as an epigenetic imprint of prior infection, DNA methylation levels in phagocytic cells could be used as a predictive biomarker for present or past infectious status.

Methods

Bacterial preparation and infection of phagocytes

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 sex-specific variation in DNA methylation levels. Only individuals self-reported as presently healthy, not currently taking medication, and without a history of diseases such as malaria, tuberculosis, cancer or hepatitis were included in the study.

Peripheral blood mononuclear cells (PBMCs) from each donor were isolated by Ficoll-Paque centrifugation and monocytes were further purified by positive selection with magnetic CD14 MicroBeads (Miltenyi Biotec). Monocytes were then derived into DCs as previously described (Barreiro et al. 2012) and subsequently infected with MTB for 2, 18, 48, and 72 hours at a multiplicity of infection of 1:1. For the infection with heat-inactivated bacteria, we used a multiplicity of infection of 5:1, which we show leads to virtually the same transcriptional response at all four time points compared to that observed with live MTB (**Supplementary Figure 2**).

We infected macrophages with *Salmonella typhimurium* as previously described (Nedelec et al. 2016). The day prior to infection, aliquots of *Salmonella typhimurium* were thawed and bacteria were grown overnight in Tryptic Soy Broth (TSB) media. Bacterial culture was diluted to mid-log phase prior to infection and supernatant density was checked at OD_{600}. Monocyte-derived macrophages were infected at a multiplicity of infection of 10:1 for 2 hours at 37°C. A control group of non-infected macrophages was treated the same way but with medium without bacteria. Macrophages were then washed and cultured for another hour in the presence of 50*μg/ml* gentamycin in order to kill all extracellular bacteria present in the medium. The cells were then washed a second time and cultured in complete medium with 3*μg/ml* gentamycin for an additional 2 hours, the time point we refer to in the main text. A control group of non-infected macrophages was treated the same way but with medium without bacteria.

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 of RNA degradation (RNA integrity number > 8) were kept for further experiments.

SeqCap Epi library preparation and sequencing

Libraries were generated with KAPA Library Preparation Kit for Illumina Platforms (KAPA Biosystems), as per the manufacturer's instructions. Briefly, genomic DNA was spiked with 30 ng of unmethylated cl857 *Sam7* Lambda DNA (Promega, Madison, WI) and fragmented to 100-300 bp with an S2 sonicator (Covaris). Fragments were then end-repaired, A-tailed, and ligated with methylated sequencing adapters. Between every enzymatic step, libraries were purified using AMPure beads (Agencourt). After ligation, in addition to the AMPure bead purification, a DUAL-SPRI size selection was performed to further select for fragments with adapters in the window of 200-400 bp. Sodium bisulfite conversion was performed with EZ DNA Methylation Lightning Kit (Zymo Research) and libraries were amplified using KAPA Hifi HotStart Uracil Tolerant Enzyme (KAPA Biosystems). Library quality was assessed by 2100 Bioanalyzer (Agilent Technologies). Samples showing the desired profile were pooled together in equal mass according to Qubit quantification. We then performed a hybridization using the SeqCap Epi kit (Roche NimbleGen). The sample pool, indexes corresponding to the sequences of the adapters used for library preparation, and repetitive DNA (C_0t) were desiccated and then incubated in hybridization buffer with a set of customized probes for 72 hours to select and sequence target regions only. Specifically, DNA methylation data was collected for 33,059 target regions spanning >130,000 CpG sites (mean length = 300 bp; mean number of CpG sites = 5), which is less than 1% of the ~28 million CpGs contained in the human genome. These regions are primarily comprised of MTB-induced differentially methylated regions identified at 18 hours post-infection using whole-genome bisulfite sequencing, as well as other distal regulatory elements in DCs where changes in DNA methylation have been shown to be most likely to occur (Pacis et al. 2015) (**Supplementary Figure 1A**). Moreover, these candidate regions are nearby differentially expressed genes in response to MTB at 18 hours. Probes targeting a two kilobase

region between coordinates 4500 and 6500 bp of the lambda genome (NC_001416.1) are included by default in the SeqCap Epi design, as a control for bisulfite conversion efficiency. Sequencing was performed using the Illumina HiSeq 2500, as per the manufacturer's instructions.

RNA-seq library preparation and sequencing

RNA-seq libraries were prepared using the TruSeq RNA Sample Prep Kit v2 or the Illumina Total Stranded RNA Library kit, as per the manufacturer's instructions. Sequencing was performed using the Illumina HiSeq 2500, as per the manufacturer's instructions.

ChIP-seq library preparation and sequencing

We collected ChIP-seq data for the H3K27ac histone mark in non-infected and Salmonella-infected macrophages as previously described (Pacis et al. 2015). Samples were crosslinked with 1% w/v formaldehyde for 20 minutes at RT and immediately quenched for 5 minutes with 1.25M Glycine at RT. The formaldehyde fixed samples were then sonicated to 150-500 bp using an S2 sonicator (Covaris), and subsequently ChIP-DNA was prepared using a manual chromatin immunoprecipitation method with Antibodies-Antigen incubation for 18 hours, followed by bead incubation for 135 minutes, and 6 5-minute washing steps. Approximately 1×106 cells were used for each ChIP and 500,000 cells for the input. The following antibody for H3K27ac was used: Abcam, Ab4729, Lot No.: GR200563-1. ChIP and Input libraries were prepared using the MicroPlex Library Preparation Kit (C05010010, Diagenode), with alterations including: PCR enrichment (13 to 14 cycles) prior to size selection and use of Ampure beads for size selection (250-350 bp). Libraries were sequenced on an Illumina HiSeq 2500.

ATAC-seq library preparation and sequencing

ATAC-seq libraries were generated from 50,000 cells, as previously described (Buenrostro et al. 2013) and sequencing was performed using the Illumina HiSeq 2500.

SeqCap Epi data processing and differential methylation analysis

Adaptor sequences and low-quality score bases (Phred score < 20) were first trimmed using Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). The resulting reads were mapped to the human reference genome (GRCh37/hg19) and lambda phage genome

using Bismark (Krueger and Andrews 2011), which uses Bowtie 2 (Langmead and Salzberg 2012) and a bisulfite converted reference genome for read mapping. Only reads that had a unique alignment were retained. Methylation levels for each CpG site were estimated by counting the number of sequenced C ('methylated' reads) divided by the total number of reported C and T ('unmethylated' reads) at the same position of the reference genome using Bismark's methylation extractor tool. 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 (**Supplementary Table 1**).

In DCs, differentially methylated (DM) CpG sites at each time point following MTB infection were identified using the R package DSS (Feng et al. 2014). We used a linear model with the following design: *DNA methylation ~ Donor + Infection*, which allowed us to consider the paired nature of the experiment and capture the effects of infection on DNA methylation observed within donors. We considered a CpG site as differentially methylated if statistically supported at a False Discovery Rate (FDR) < 0.01 and an absolute mean methylation difference above 10%. Only CpG sites that had a coverage of at least 5X in each of the samples were included in the analysis (103,649 in total).

To identify DM sites that show stable loss of methylation (as Cluster 3 DM sites in DCs) in Salmonella-infected macrophages using one individual, we performed a hierarchical clustering analysis on sites that specifically: (*i*) do not change methylation at 2 hours (|methylation difference| < 10%), and (*ii*) lose methylation at 48 hours (methylation difference < -40%).

5hmC enrichment at DM sites

To calculate the enrichment of 5-hydroxymethylcytosine (5hmC) at DM CpG sites (Clusters 1, 2 and 3), we compared the distribution of 5hmC levels between DM and non-DM sites, using previously published TAB-seq data from non-infected DCs (Pacis et al. 2015). Since non-DM sites have lower overall levels of baseline methylation than DM sites **(Supplementary Figure**

8A), we performed similar enrichment analysis by using a random set of non-DM sites that matches the distribution of methylation found in non-infected samples within each set of DM sites **(Supplementary Figure 8B)**. Each random set contains the same number of CpG sites as those identified within each DM cluster.

RNA-seq data processing and identification of differentially expressed genes

Read count estimates per gene were obtained using the alignment-free method Kallisto (Bray et al. 2016). For all downstream analyses, we excluded non-coding and lowly-expressed genes with an average read count lower than 10 in all of the samples, resulting in 13,955 genes in total. The R package DESeq2 (Anders et al. 2013) was used to identify differences in expression levels between non-infected and infected samples, at each time point. Nominal p-values were corrected for multiple testing using the Benjamini-Hochberg method (Benjamini and Hochberg 1995). The complete list of differentially expressed genes can be found in Supplementary Table 3.

Gene set enrichment analysis

We used ClueGO (Bindea et al. 2009) at default parameters to test for enrichment of functionally annotated gene sets among differentially expressed genes. The results for these enrichment analyses are reported in Supplementary Table 4. Enrichment p-values were based on a hypergeometric test using the set of 13,955 genes as background. Benjamini-Hochberg method was applied for multiple testing correction.

ChIP-seq data processing and tag density profiles

We started by trimming adapter sequences and low-quality score bases using Trim Galore. The resulting reads were mapped to the human reference genome using Bowtie 2 with the following option: -N 1. Only reads that had a unique alignment were retained and PCR duplicates were further removed using Picard tools (http://broadinstitute.github.io/picard/). Tag density profiles for chromatin modifications and genome accessibility patterns around regions of interest was accomplished with ngs.plot package (Shen et al. 2014) using default parameters.

ATAC-seq data processing and TF footprinting analysis

ATAC-seq reads were trimmed for adapter sequences and low-quality score bases and were mapped to the human reference genome. Mapping was performed using BWA-MEM (Li and Durbin 2009) in paired-end mode at default parameters. Only reads that had a unique alignment (mapping quality > 10) were retained. TF footprinting analyses were performed as previously described, using the Centidual algorithm (Pacis et al. 2015) which specifically tests for differential binding between two experimental conditions. For each of the actively bound TFs in DCs (FDR < 0.2 at 18 hours post-infection; **Supplementary Table 5**)), we first trained Centidual assuming that the footprint was bound in the two conditions. Then, we fixed the model parameters and generated a likelihood ratio and posterior probability π_{lt} for each condition t separately and for each site l. 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 (α and δ), the PWM effect β, and PWM times the treatment effect γ:

$$\log\left(\frac{\pi_{lt}}{1-\pi_{lt}}\right) = \alpha \times (1 - I_t) + \beta \times \text{PWMscore}_l + \delta \times I_t + \gamma \times (I_t \times \text{PWMscore}_l)$$

where I_t 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 γ, corresponding to the evidence for condition specific binding. ATAC-seq samples were down-sampled to obtain similar number of reads between NI and HI samples at each time-point. We used a window size of 250 bp on either side of the motif match, and reads with fragment lengths [40, 99] and [100, 139] bp for footprinting analyses. Motifs that belong to non-human organisms were excluded in downstream analyses.

To test for differential binding of immune-related TFs (NF-κB, AP-1, STATs and IRFs) between non-infected and infected samples, we compared the intensity of the Tn5 sensitivity-based footprint across all matches to motifs of TFs that belong to each family, in the hypomethylated regions. Specifically, the list of motif IDs (and corresponding names) that were aggregated to their respective TF family are shown in **Table 1**.

Table 1. List of motif IDs aggregated to their respective immune-related transcription factor family

NF-κB		AP-1		STATs		IRFs	
MA0101.1	REL	MA0490.1	JUNB	MA0517.1	STAT1::STAT2	MA0050.2	IRF1
MA0107.1	RELA	MA0491.1	JUND	MA0519.1	Stat5a::Stat5b	MA0051.1	IRF2
		MA0462.1	BATF::JUN			MA0772.1	IRF7
		MA0476.1	FOS			MA0652.1	IRF8
		MA0477.1	FOSL1			MA0653.1	IRF9
		MA0478.1	FOSL2				
		MA0656.1	JDP2(var.2)				
		MA0834.1	ATF7				
		MA0835.1	BATF3				

To test for enrichment of binding of methylation-sensitive ("methyl-minus") TFs in hypomethylated regions, we compared the proportion of regions that overlap well-supported footprints (posterior probability > 0.99) of "methyl-minus" TFs reported in Yin et al. (Yin et al. 2017)), among non-DMRs and hypomethylated regions. The list of motif IDs (and corresponding names) that were included in the analysis are shown in **Table 2**.

Table 2. List of motif IDs categorized as methylation-sensitive ("methyl-minus") transcription factors

MA0018.2	CREB1	MA0605.1	Atf3	MA0761.1	ETV1
MA0028.2	ELK1	MA0609.1	Crem	MA0762.1	ETV2
MA0038.1	Gfi1	MA0615.1	Gmeb1	MA0763.1	ETV3
MA0058.3	MAX	MA0616.1	Hes2	MA0764.1	ETV4
MA0059.1	MAX::MYC	MA0632.1	Tcfl5	MA0765.1	ETV5
MA0062.2	Gabpa	MA0635.1	BARHL2	MA0768.1	LEF1
MA0081.1	SPIB	MA0636.1	BHLHE41	MA0772.1	IRF7
MA0093.2	USF1	MA0638.1	CREB3	MA0777.1	MYBL2
MA0095.2	YY1	MA0640.1	ELF3	MA0780.1	PAX3
MA0099.2	FOS::JUN	MA0641.1	ELF4	MA0794.1	PROX1
MA0104.3	Mycn	MA0649.1	HEY2	MA0821.1	HES5
MA0136.2	ELF5	MA0663.1	MLX	MA0822.1	HES7
MA0149.1	EWSR1-FLI1	MA0679.1	ONECUT1	MA0823.1	HEY1
MA0156.2	FEV	MA0680.1	PAX7	MA0829.1	Srebf1(var.2)
MA0414.1	XBP1	MA0684.1	RUNX3	MA0830.1	TCF4
MA0463.1	Bcl6	MA0686.1	SPDEF	MA0831.1	TFE3
MA0464.2	BHLHE40	MA0694.1	ZBTB7B	MA0834.1	ATF7
MA0470.1	E2F4	MA0731.1	BCL6B	MA0839.1	CREB3L1
MA0473.2	ELF1	MA0735.1	GLIS1	MA0840.1	Creb5
MA0474.2	ERG	MA0736.1	GLIS2	MA0844.1	XBP1
MA0475.2	FLI1	MA0749.1	ZBED1	MA0850.1	FOXP3
MA0476.1	FOS	MA0750.1	ZBTB7A	MA0862.1	GMEB2
MA0495.1	MAFF	MA0754.1	CUX1	MA0864.1	E2F2
MA0511.2	RUNX2	MA0756.1	ONECUT2	MA0871.1	TFEC
MA0526.1	USF2	MA0757.1	ONECUT3	MA0876.1	BSX
MA0595.1	SREBF1	MA0758.1	E2F7	MA1099.1	Hes1
MA0603.1	Arntl	MA0759.1	ELK3		

Relationship between gene expression and chromatin accessibility

Peaks were first called on ATAC-seq using the MACS2 software suite (Zhang et al. 2008) with the added parameters: -g hs -q 0.05 --broad --nomodel --extsize 200 --nolambda. All peaks from each sample were then merged to provide one set of combined peaks. To count the number of reads overlapping peaks, we used featureCount (from the subread package version 1.4.4) (Liao et al., 2014) with the following option: -p -P -d 19 -D 1000. For all downstream analyses, we excluded low-count peaks with have an average read count lower than 10 across all samples, resulting in 79,282 peaks in total. We then plotted the distribution of changes in Tn5

accessibility (between non-infected and MTB-infected DCs across the five time points of infection (2, 4, 18, 24, 48 and 72 hours)) for the top 25% most variable peaks associated with DE genes in each cluster. The DE genes associated with the selected peaks represent ~50% of the total genes within each of the DE cluster: (*i*) Early induced: 418/765 = 55%; (*ii*) Intermediate induced: 418/839 = 49%; and (*iii*) Persistent induced: 1083/2091 = 52%.

Acknowledgments

Supplementary Figures
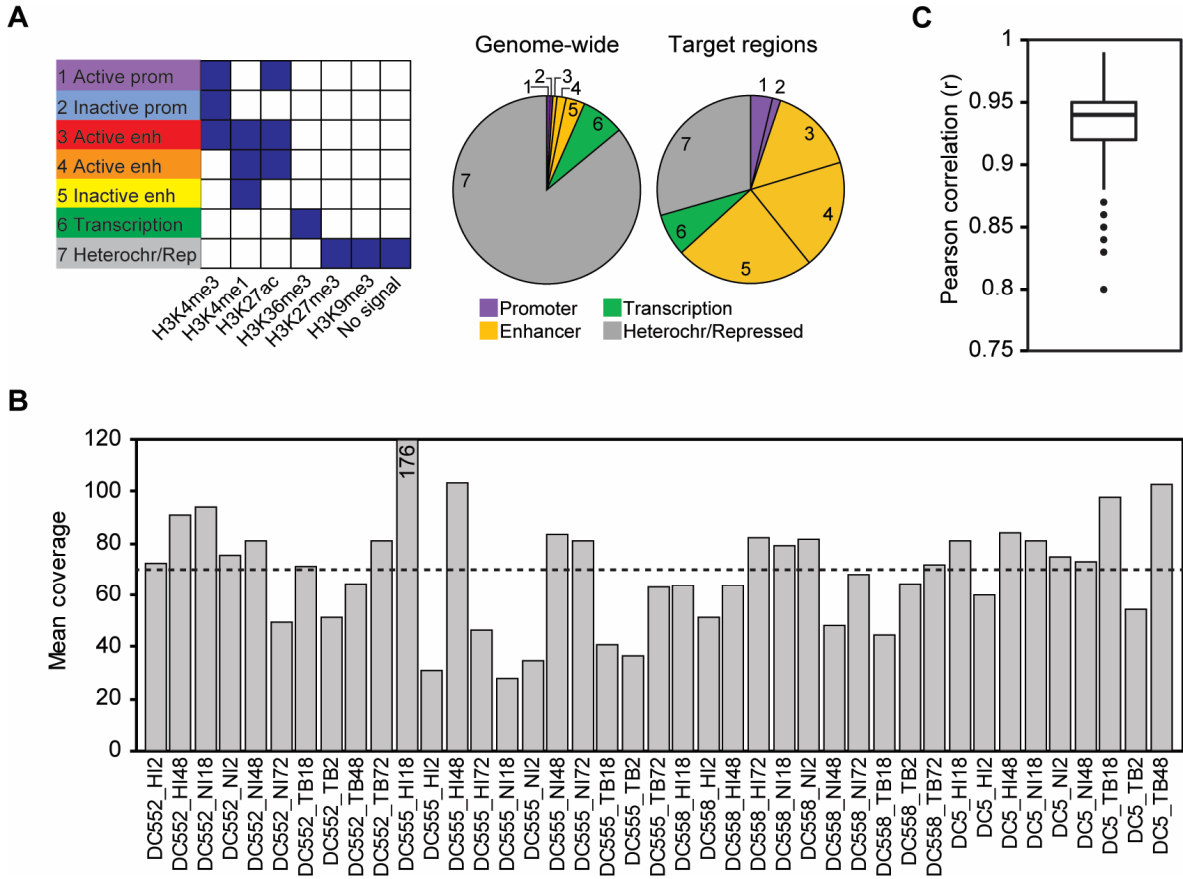
Supplementary Figure 1. Characteristics of SeqCap target regions. (**A**) Left: Combination of histone patterns used to define the 7 chromatin states (previously defined in (Pacis et al. 2015)). Right: Pie charts showing the distribution of chromatin state annotations genome-wide and within SeqCap target regions in MTB-infected DCs. (**B**) Bar plots showing mean coverages of CpG sites within the target regions for each sample. Dotted line denotes the average coverage across all samples. (**C**) Correlation between DNA methylation levels at target regions among replicates (using only CpG sites with ≥5 coverage in all samples).
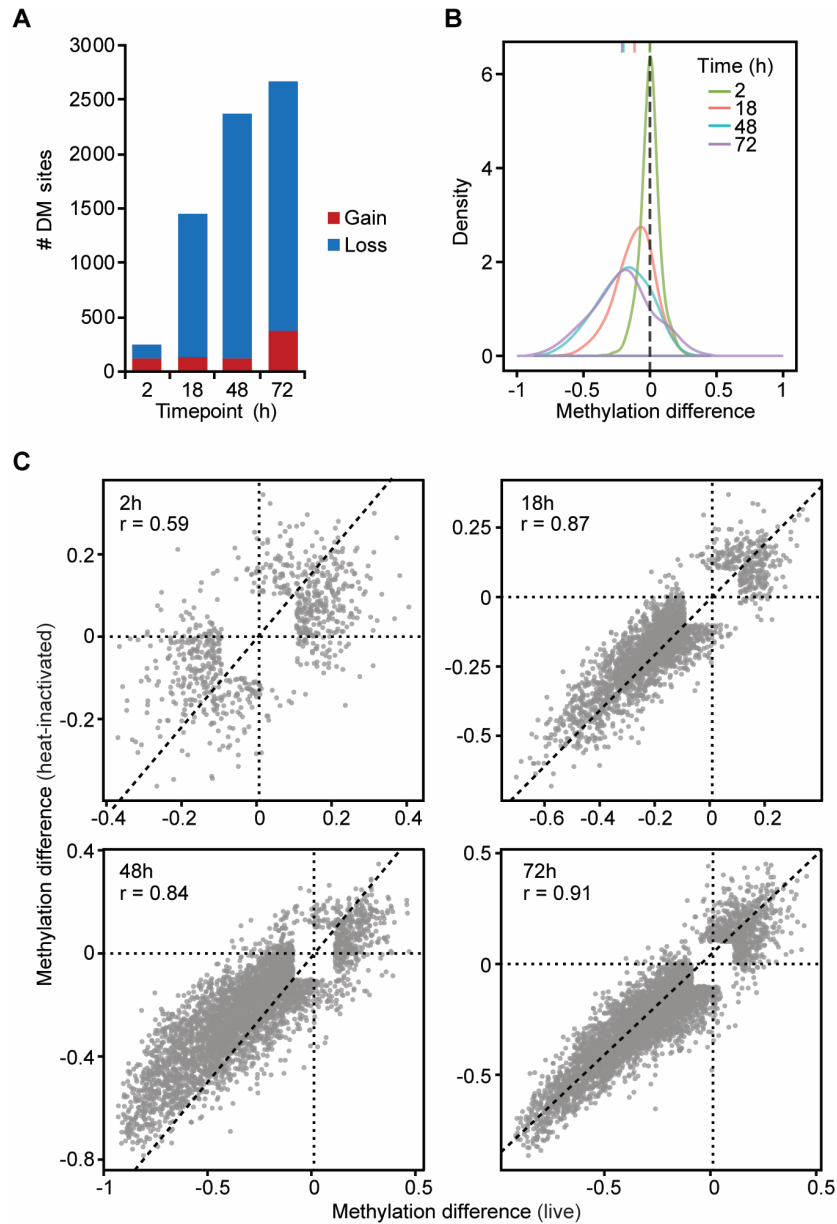
Supplementary Figure 2. DNA methylation dynamics in DCs in response to infection with heat-inactivated MTB. (**A**) Bar plots showing the number of differentially methylated sites identified at a |methylation difference| > 10% and FDR < 0.01 (y-axis) at each time point after infection with heat-inactivated MTB (2, 18, 48 and 72 hours (h); x-axis). (**B**) Distribution of differences in methylation between infected and non-infected cells at DM sites, at each time point. (**C**) Scatterplots depicting the correlation between differences in methylation in response to live (x-axis) or heat-inactivated MTB (y-axis) at each time point after infection (2, 18, 48 and 72 hours (h)).
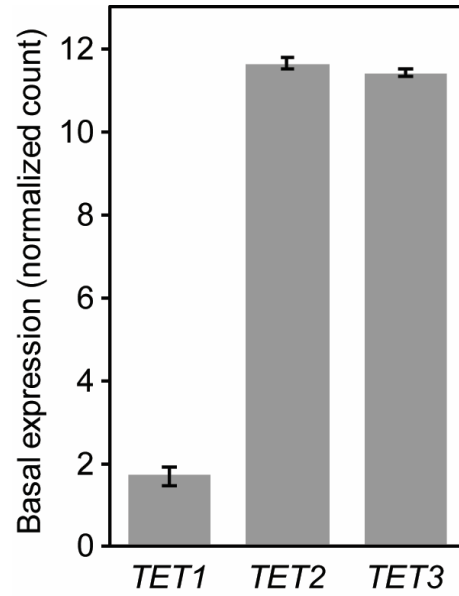
Supplementary Figure 3. Expression of *TET* genes in non-infected DCs. Bar plots showing the expression levels (in log2-normalized read counts) of *TET* genes, in non-infected DCs. All data are represented as mean ± s.e.m.
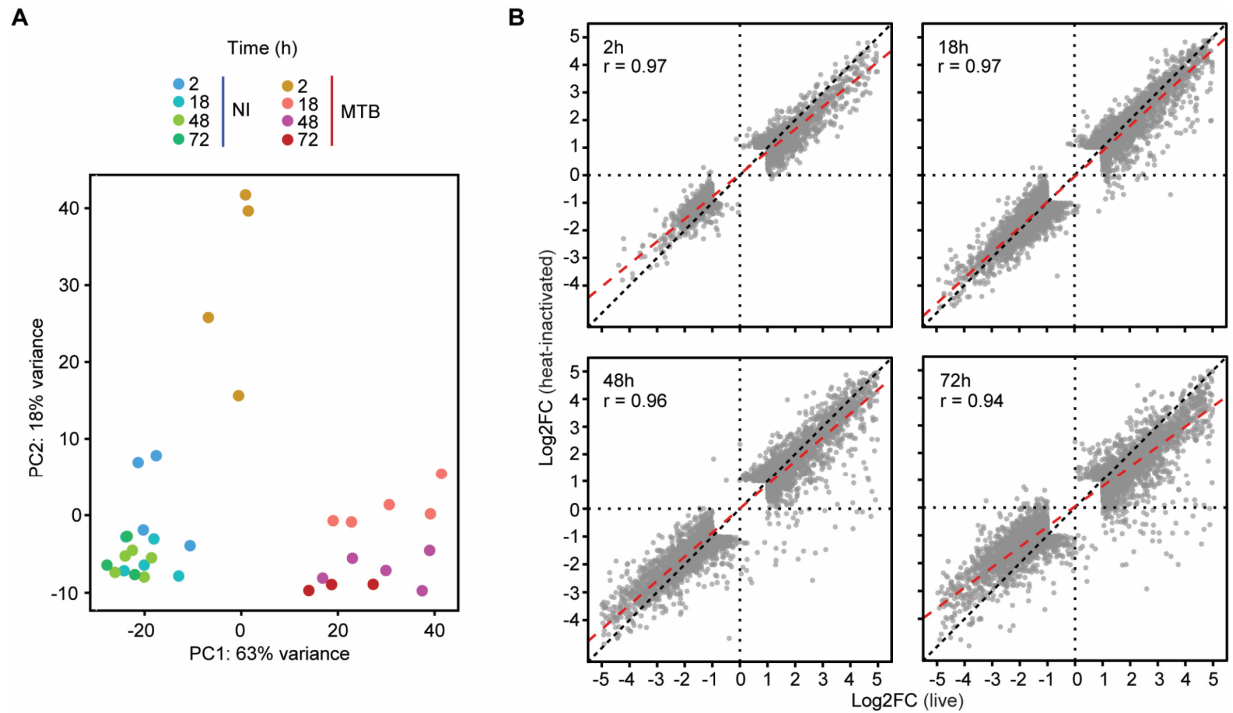
Supplementary Figure 4. Time course analysis of gene expression in DCs in response to infection with heat-inactivated MTB. (**A**) Principal component analysis of gene expression data from all samples. PC1 (x-axis) and PC2 (y-axis) clearly separate non-infected DCs from DCs infected with live MTB. (**B**) Scatterplots depicting the correlation between changes in expression (log2 fold changes) in response to live (x-axis) or heat-inactivated MTB (y-axis) at each time point after MTB infection (2, 18, 48 and 72 hours (h)).
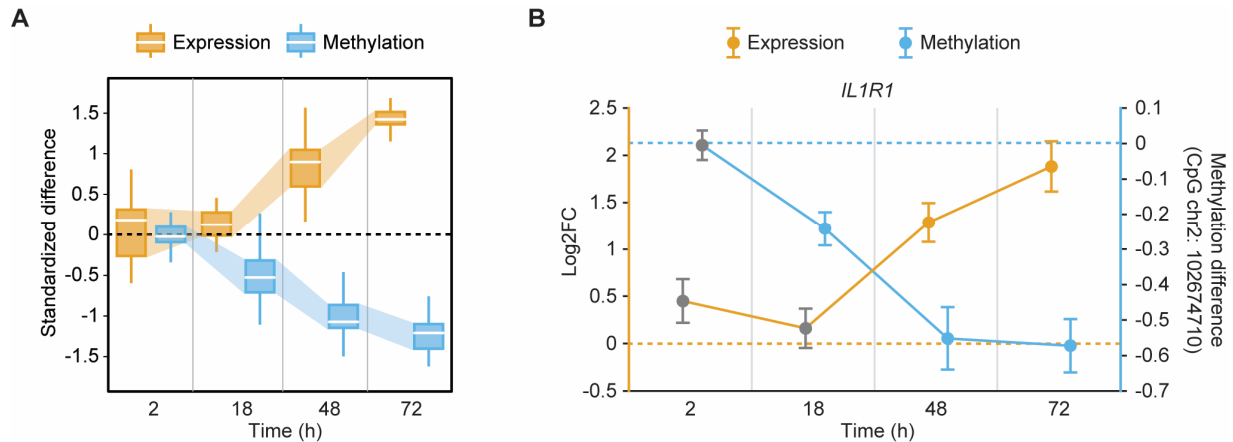
Supplementary Figure 5. Example of genes for which DNA demethylation occurred prior to gene activation. (**A**) Boxplots showing the distribution of standardized differences in methylation (blue) of Cluster 3 DM sites along with the corresponding standardized differences in expression (orange) of the associated genes, for which DNA demethylation occurred prior to gene activation (i.e., not differentially expressed after 2 and 18 hours post-infection using a relaxed FDR $\geq$ 0.1; n = 17 genes). These distributions are shown for each time point after MTB infection (2, 18, 48 and 72 hours (h)). (**B**) Specific example of changes in DNA methylation levels (blue; right y-axis) across all time points, along with the corresponding fold changes in log2 scale (orange; left y-axis) in expression of the associated gene that is induced after loss of methylation (defined in (A)). All data are represented as mean $\pm$ s.e.m. Gray dots denote no significant change in gene expression or methylation.
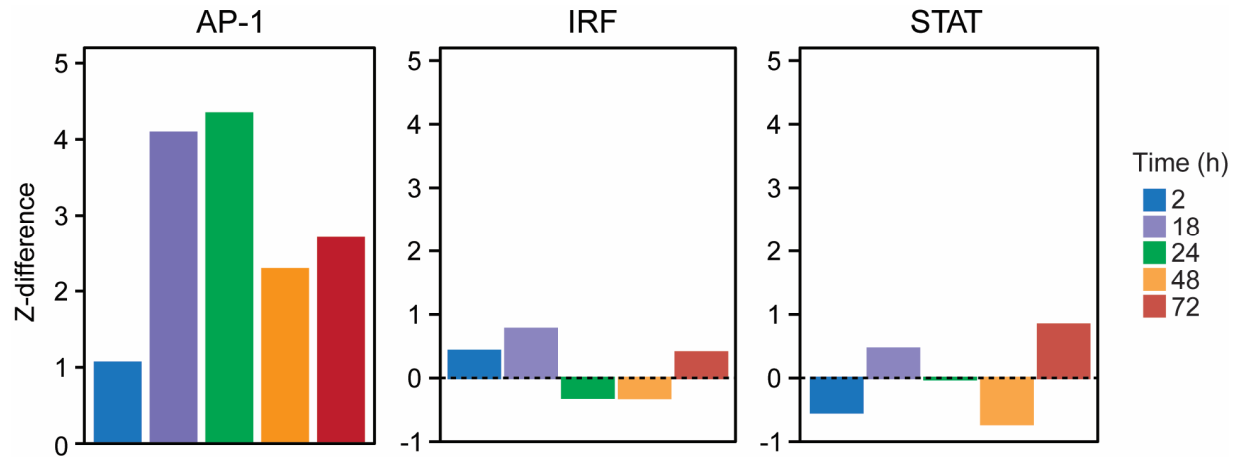
Supplementary Figure 6. Binding profiles of immune-related transcription factors within hypomethylated regions. Bar plots showing differences in TF occupancy score predictions for AP-1, IRF or STAT motifs between MTB-infected and non-infected DCs (Z_{MTB} - Z_{NI}; y-axis; see Methods), across four time points after infection (in hours (h); x-axis; see Methods). A positive Z-score difference indicates increased TF binding in hypomethylated regions after MTB infection.
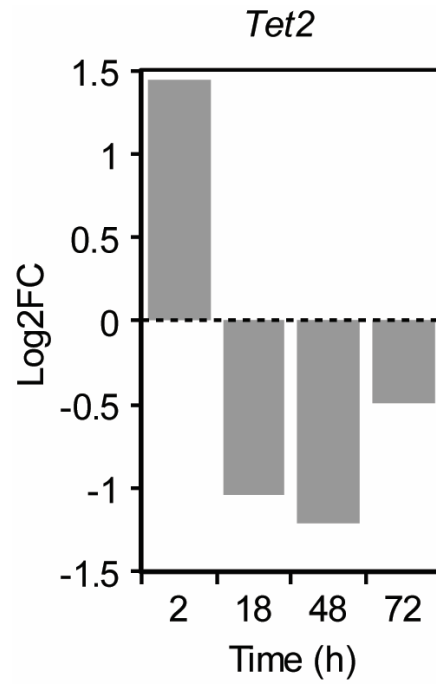
Supplementary Figure 7. *TET2* expression profile in MTB-infected DCs. Bar plots showing changes in expression levels (in log2) of *Tet2*, across four time points after MTB infection (in hours (h); x-axis).
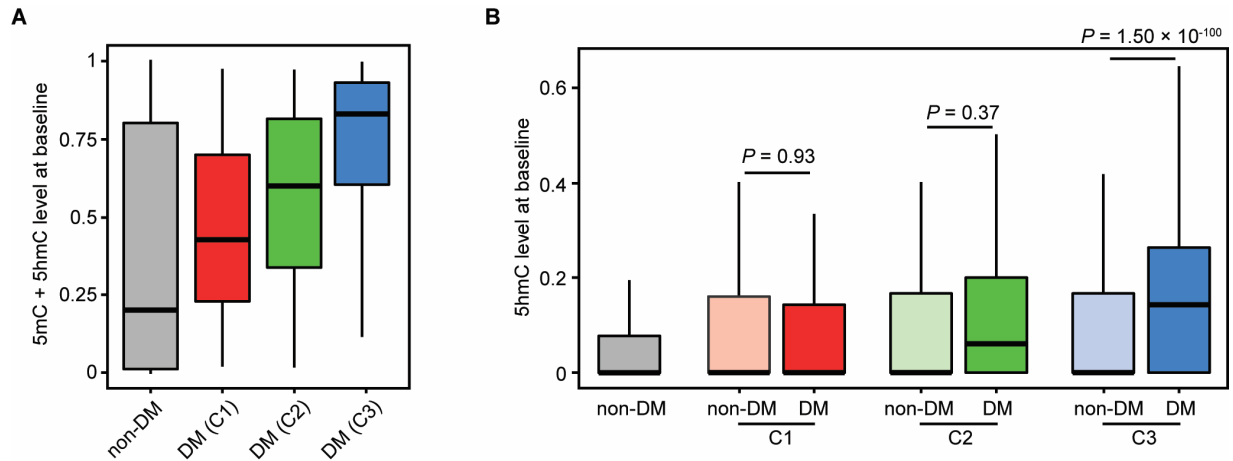
Supplementary Figure 8. 5hmC enrichment in differentially methylated (DM) CpG sites.
(**A**) Boxplots of methylation levels (5mC + 5hmC) of non-DM and DM sites (Clusters 1, 2 and 3), in non-infected DCs. (**B**) Boxplots of 5hmC levels of all non-DM (gray) and DM sites in each cluster. Also shown are distributions of 5hmC levels using a random set of non-DM sites that matches the distribution of methylation found in non-infected samples within each set of DM sites (shown in (A)). Each random set contains the same number of CpG sites as those identified within each DM cluster.

4 Discussion and Perspectives

4.1 Infection of human dendritic cells involves active, proliferation-independent DNA demethylation

A key feature of innate immune cells is their ability to rapidly respond to intracellular and extracellular signals. Epigenetic mechanisms involving histone modifications have been shown to contribute to the regulatory programs induced by innate immune cells in response to an infectious agent (Bierne et al. 2012; Smale et al. 2014). In contrast, given its presumed stability, the contribution of changes in DNA methylation to innate immune responses has been less explored. During my thesis, I performed comprehensive DNA methylation profiling of non-infected and MTB-infected DCs at multiple time points, using the combination of whole-genome and targeted bisulfite sequencing methods (MethylC-seq and SeqCap Epi, respectively). I identified thousands of MTB-induced differentially methylated CpG sites, with the vast majority exhibiting gradual loss of methylation at enhancer elements over the time course of infection. Interestingly, several studies of DNA methylation dynamics in innate immune cells in response to different stimuli have reported similar trends associated with loss of methylation (Marr et al. 2014; Zhang et al. 2014; Cizmeci et al. 2016; Wiencke et al. 2016). This general trend towards demethylation of host cell DNA upon activation suggests that there are some common regulatory regions or genes that are targeted by different pathogens. In support of this hypothesis, I found that among the CpG sites that displayed loss of methylation in Salmonella-infected macrophages, 45% overlapped with DM sites identified in MTB-infected DCs (see Article II).

I established human DCs as a model to identify and characterize active demethylation events during infection. DCs are post-mitotic and do not to proliferate in response to infection, as confirmed by Carboxyfluorescein Diacetate Succinimidyl Ester (CFSE) cell proliferation assay (Quah and Parish 2010), which implies that passive DNA demethylation is an unlikely mechanism for the observed loss of methylation. Recent studies have implicated the family of Ten-Eleven-Translocation proteins (TET1, TET2, and TET3) in active demethylation processes

via the generation of 5-hydroxymethylcytosine (5hmC). Consistent with this notion, I found that CpG sites that lose methylation upon infection display high levels of 5hmC, suggesting that these regions are actively bound by TET enzymes. This possibility is further supported by several reports demonstrating the role of TET2 – the most abundant TET family member in myeloid cells – as a key determinant of myeloid cell identity and several myeloid malignancies (Klug et al. 2013; Alvarez-Errico et al. 2015). Collectively, these results provide novel evidence for widespread, rapid changes in DNA methylation in innate immune cells (both DCs and macrophages) during infection, independent of cell division. This thesis also provides a framework from which we can learn about pathogen-specific versus generic responses. It will be interesting to explore further the molecular mechanisms underlying this targeted demethylation process.

4.2 Temporal hierarchy of transcriptional and epigenetic changes in response to infection

Here, I show that DNA demethylation is often associated with induction of inflammatory genes. However, it remains unclear whether altered methylation patterns directly invoke transcriptional modulation or whether such patterns are the downstream consequence of TF binding to regulatory regions. Paired time course analysis of DNA methylation and gene expression levels revealed that changes in gene expression tend to occur prior to detectable changes in DNA methylation. I further show that DNA demethylation is preceded by the deposition of H3K27ac active histone marks, thus arguing against its instructive role (Ford et al. 2017). Consistently, several recent reports have shown that other epigenetic modifications, such as the H3K4me1 enhancer mark, have a similar passive regulatory function (Dorighi et al. 2017; Rickels et al. 2017; Vandenbon et al. 2017). For instance, Vandenbon et al. assessed the timing of changes in several histone modifications in LPS-induced mouse DCs (Vandenbon et al. 2017). Using a ChIP-seq time series dataset, they found that only the deposition of H3K27ac at promoters and enhancers correlated with transcriptional induction. In striking contrast, the deposition of several other histone modifications (namely, H3K9K14ac, H3K4me3, H3K36me3, and H3K4me1) occurred within specific time frames after stimulation, independent of the timing of induction

of transcription. Integrative analysis with TF binding data further revealed that the dynamics of histone modifications reflects the timing of activation of signal-dependent TFs. These studies together with findings from this work suggest that changes in DNA methylation (and certain histone modification marks) play a secondary role in controlling the gene-specific regulatory programs engaged by innate immune cells in response to extrinsic signals. More broadly, it also highlights the importance of time series data in order to fully capture the relationship between epigenetic modifications and transcriptional responses.

As previously mentioned, 5-methylcytosine (5mC) is subsequently oxidized by TET enzymes, generating 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) as intermediates of the active demethylation pathway. Single-base resolution map of 5hmC in non-infected DCs revealed that DM sites were strikingly enriched for 5hmC even prior to infection. It is possible that 5hmC is a stable epigenetic mark in its own right and has a function that is distinct from that of 5mC (Yildirim et al. 2011; Mellen et al. 2012; Iurlaro et al. 2013; Spruijt et al. 2013; Xiong et al. 2016). Specifically, 5hmC may be a part of enhancer activation process by counteracting transcriptionally restrictive chromatin states, and thus could serve as an epigenetic priming mechanism that ensures a fast and effective response to an incoming stimulus (Creyghton et al. 2010; Serandour et al. 2012; Yu et al. 2012; Calo and Wysocka 2013; Hon et al. 2014; Vento-Tormo et al. 2016; Mahe et al. 2017). Similarly, ChIP-seq data on histone modifications show that many of the regions that lose methylation were pre-marked by H3K4me1 prior to immune stimulation, suggesting that both epigenetic marks function together to contribute to the priming of enhancers. However, whether or not the binding of signal-dependent TFs and subsequent activation of enhancers is directly dependent on initial 5hmC deposition remains elusive. Much remains to be done to explain the mechanisms by which oxidative 5mC derivatives regulate the innate immune system. The identification of the readers specific for 5hmC, 5fC and 5caC will provide critical insights into their functional importance.

To establish the relative contribution of epigenetic modifications in the regulation of innate immune responses, one approach is to perform genetic targeting, namely, the deletion or mutation, of chromatin-modifying enzymes. CROP-seq (Datlinger et al. 2017), CRISP-seq (Jaitin et al. 2016), and PERTURB-seq (Dixit et al. 2016), which combine single-cell RNA

sequencing (scRNA-seq) (Macosko et al. 2015; Zilionis et al. 2017) and clustered regularly interspaced short palindromic repeats (CRISPR)-based perturbations (Ran et al. 2013; Sander and Joung 2014), enable us to simultaneously investigate the functional relevance of multiple epigenetic features individually or in combination (**Figure 1**). The three methods rely on the use of barcodes to identify both the individual genetic perturbation and the cells affected. Briefly, a library of uniquely barcoded CRISPR guide RNAs (gRNA) targeting genes of interest is introduced into a population of cells. The mRNAs of individual cells are then extracted with uniquely barcoded primers via droplet-based microfluidics.

However, since genetic manipulation of chromatin-modifying enzymes often leads to epigenetic perturbations across the entire genome, it will difficult to pinpoint the genomic regions at which epigenetic modifications are functionally relevant for transcriptional regulation. Thus, a complementary approach will be to change single epigenetic marks at specific loci, using chromatin-modifying enzymes (or catalytic domains) fused to an inactive form of Cas9 (Kungulovski and Jeltsch 2016; Pulecio et al. 2017; Stricker et al. 2017). Similar to the previous approach, this would require appropriate libraries of gRNAs, each targeting dCas9-chromatin-modifying enzymes to thousands of different genomic sites, focusing on relevant regions that have been identified from the integrative analysis of epigenomic profiles.
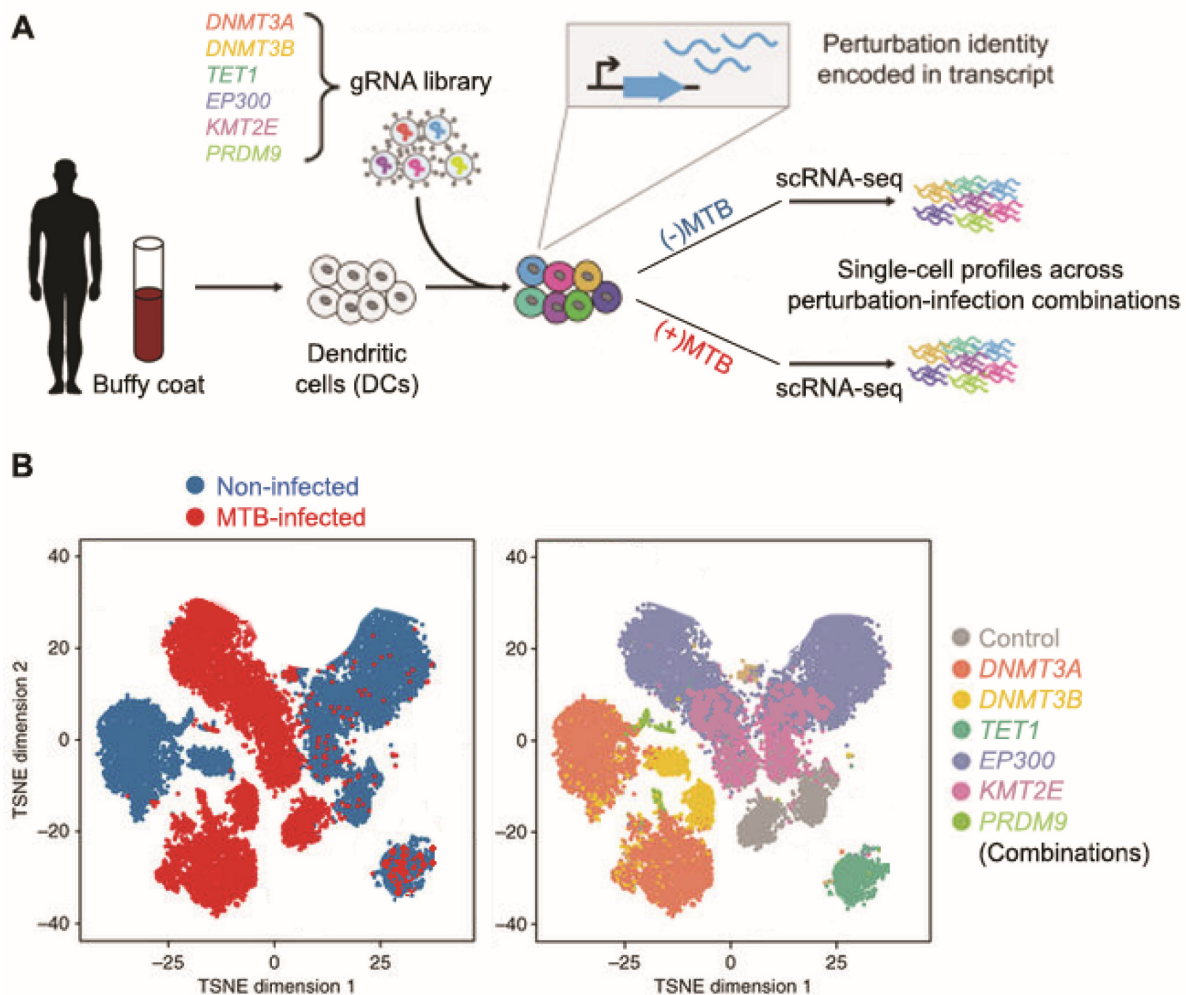
Figure 1. Systematic assessment of the effects of epigenetic perturbations on the innate immune response. (**A**) Schematics of the experimental design. Monocyte-derived dendritic cells (DCs) are isolated from peripheral blood of male adult donors. A pooled gRNA library promotes gene knockouts through lentiviral transfection, resulting in single or multiple gene knockouts. Heterogeneous cell populations are profiled (using single-cell RNA-seq; scRNA-seq), after infection with *Mycobacterium tuberculosis* (MTB) for 18 hours. Genetic perturbations are determined by the gRNA-specific unique molecular identifiers (UMIs), and the transcriptional effect of that perturbation are determined by the collection of mRNAs associated the cell-specific UMIs. (**B**) Hypothetical example of a t-distributed stochastic neighbor embedding (t-SNE) plot of individual dendritic cells, color-coded by infection status (left) or gRNA target genes individually or in combination (right). *Image adapted from (Adamson et al. 2016; Kang et al. 2018).*

4.3 Methylation-sensitive transcription factors

The importance of DNA methylation in the control of gene expression is widely accepted. Despite the fast accumulating profiles of DNA methylomes in various biological processes and physiological conditions, the interpretation of these data sets have often failed to establish the causality between the changes in DNA methylation and physiological outcomes. Despite the numerous studies directed at establishing the functional relationship between DNA methylation and TF binding, no consensus has emerged. The effect of CpG methylation on TF binding varies from one TF to another (Blattler and Farnham 2013; Schubeler 2015; Zhu et al. 2016; Yin et al. 2017), and this can be broadly classified into three categories: (*i*) TFs that recognize sites with CpG sequences but are not affected by cytosine methylation (little effect or methylation-insensitive); (*ii*) TFs that prefer unmethylated CpG (methylation-sensitive: methyl-minus), and (*iii*) TFs that prefer methylated CpG (methylation-sensitive: methyl-plus) (**Figure 2A,B,C**). Moreover, it has been reported that certain methylation-insensitive TFs could behave as "pioneer factors" and create a site of reduced methylation that allows the binding of methylation-sensitive (methyl-minus) TFs (Domcke et al. 2015) (**Figure 2D**).

Using TF footprint analysis, I show that the signal-dependent TF NF-κB/Rel, is recruited to hypomethylated regions prior to loss of methylation, further supporting the lack repressive capacity of DNA methylation (Ford et al. 2017). However, we cannot completely exclude the possibility that the observed loss of methylation plays an active role in regulating gene expression during infection, but only in a small subset of the differentially methylated CpG sites. Under this scenario, DNA demethylation would be crucial for the binding of methylation-sensitive (methyl-minus) TFs that specifically play a role at later stages of infection. This suggests overall that DNA demethylation might fine-tune, rather than tightly control immune responses. ATAC-seq footprinting offers a powerful approach to perform global analysis of binding of TFs (with known motif) in a single experiment. However, in contrast to chromatin immunoprecipitation (ChIP)-based methods, which directly measures TF occupancy, ATAC-seq footprinting only suggests TF binding by proxy. Moreover, since cell type-specific enhancers are formed by the juxtaposition of several binding sites specific for various TFs, identifying the involved TF might be challenging. Thus, further validation using ChIP-seq

against TFs predicted to be recruited to hypomethylated regions will be required to validate these results.

The identification of methylation-sensitive TFs (methyl-minus and methyl-plus) and the elucidation of their characteristics will be a very important step towards a mechanistic understanding of the role of DNA methylation in innate immune responses to infection. For example, superimposing DNA methylation and TF binding ChIP-seq data sets can be an effective approach to validate methylation-dependent DNA-TF interactions. Another possible approach is to investigate the effects of TF binding on DNA methylation by inhibiting TFs that bind specifically to hypomethylated regions (such as NF-κB/Rel). Alternatively, we can assess genome-wide changes in TF binding sites by perturbing DNA methylation by knocking out DNMTs or TETs. Moreover, recent developments in targetable epigenome-editing tools (using TALEs or CRISPR/Cas9) enable us to assign direct transcriptional and functional consequences to locus-specific epigenetic modifications.

Finally, although only ~25% of known TF binding motifs in humans contain at least one CpG site (Blattler and Farnham 2013; Yin et al. 2017), binding of TFs that lack CpGs in their binding motifs could possibly still be influenced by DNA methylation, through indirect crosstalk with other epigenetic mechanisms such as histone modifications.
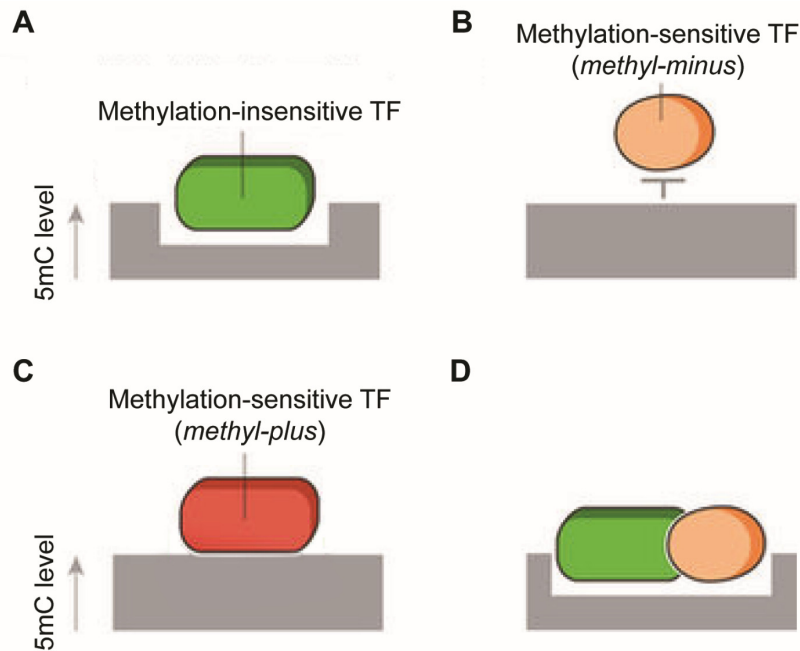
Figure 2. Emerging scenarios showing the effect of DNA methylation (denoted by level of 5-methylcytosine or 5mC) on transcription factor (TF) binding. (**A**) A methylation-insensitive TF binds and causes reduced methylation. (**B**) A methylation-sensitive (methyl-minus) TF is blocked by 5mC. (**C**) A methylation-sensitive (methyl-plus) TF preferentially binds to the methylated state of its binding site. Note that methyl-plus TFs do not include methyl-CpG-binding domain (MBD) proteins, which also specifically bind to the methylated state (discussed in section 1.3.3). (**D**) A methylation-insensitive TF acting as a pioneer factor binds and creates reduced methylation, which allows the binding of a methylation-sensitive (methyl-minus) TF. *Image adapted from (Schubeler 2015).*

4.4 Trained immunity

Immunological memory has traditionally been a hallmark trait of the adaptive immune system, present only in antigen-specific T and B cells. Recent studies have challenged this dogma by demonstrating that innate immune cells can also show memory-like behavior. Studies involving organisms that lack adaptive immunity have shed light on the notion of memory within the

innate immune system by showing that plants and invertebrates display protective mechanisms against reinfection (Kurtz and Franz 2003; Pham et al. 2007; Rodrigues et al. 2010; Luna and Ton 2012; Kachroo and Robin 2013). Similarly, mice rendered unable to mount an effective adaptive immune response (due non-functional T and B lymphocytes) can mount resistance to reinfection (Quintin et al. 2012). Considering the development of innate immune memory in humans, it has been shown that Bacillus Calmette-Guerin (BCG) vaccination against tuberculosis was associated with a reduction in all-cause mortality, suggesting that specific, adaptive immunity may not be the only protective mechanism at work (Niobey et al. 1992; Kristensen et al. 2000; Garly et al. 2003). The term "trained immunity" has been proposed to describe the phenomenon in which the innate immune system can mount a stronger and more effective immune response upon secondary infection with the same or unrelated pathogens (Quintin et al. 2014). Recent studies further suggest that histone modifications, including acetylation and methylation, play a key role in the induction of trained immunity (Cheng et al. 2014; Saeed et al. 2014; Kaufmann et al. 2018).

Here, I show that the infection of innate immune cells (i.e., DCs and macrophages) with pathogenic bacteria is associated with pervasive loss of methylation. Interestingly, levels of methylation at DM CpG sites gradually decreased over the course of infection but never reverted back to higher levels, even among genes for which expression levels return to basal state. This observation strongly suggests that targeted loss of DNA methylation is longer lasting (at least up to 72 hours), and thus may be ideally suited as a mechanism of epigenetic memory. It will be interesting to explore further how trained immunity is epigenetically encoded, focusing specifically on DNA demethylation. Notably, assessing the effects of inhibition of TET-mediated active demethylation process in the development of trained immunity will provide critical insights into its functional importance. In addition, it will be interesting to determine and compare the durability of changes in DNA methylation and histone modifications, after training. **Figure 3** summarizes the known and possible mechanisms through which DNA methylation regulates innate immune responses to infection.
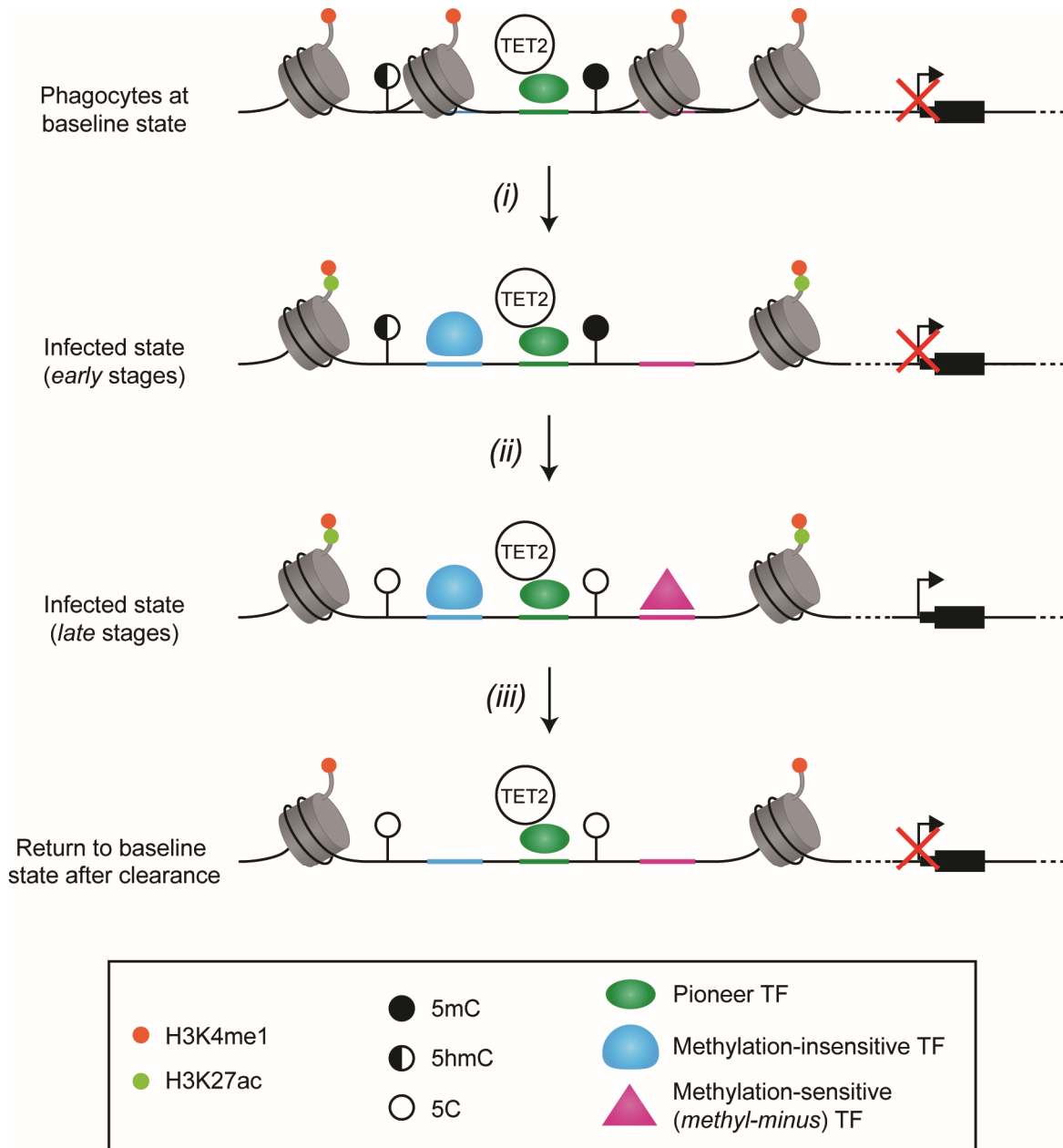
Figure 3. Proposed model for the role of DNA methylation in innate immune responses to infection. (*i*) Infection of phagocytes (such as DCs and macrophages) triggers the recruitment of signal-dependent TFs (methylation-insensitive TFs specifically) within the cell type-specific enhancer repertoires pre-established (primed; marked by H3K4me1) by pioneer TFs prior to immune stimulation. The rapid binding of methylation-insensitive TFs during early stages of infection, likely induces local chromatin opening, the recruitment of histone acetyltransferase p300, and the subsequent deposition of activating H3K27ac marks in these regions (Bhatt and

Ghosh 2014). (*ii*) The finding that DM CpG sites are highly enriched for 5hmC at baseline suggests that these sites are pre-bound by Ten-eleven Translocation (TET) enzymes, possibly via the pioneer factor PU.1 (de la Rica et al. 2013). In this scenario, p300 acetylates TET2, conferring enhanced enzyme activity (Zhang et al. 2017), which might account for the loss of DNA methylation in response to infection. Removal of methylation marks allows the binding of methylation-sensitive (methyl-minus) TFs to induce expression of genes that play a role at later stages of infection. (*iii*) Loss of methylation at DM CpG sites persists even after immune stimuli has ceased, which might contribute to the faster and stronger transcriptional response upon restimulation (i.e., trained immunity). 5mC: 5-methylcytosine, 5hmC: 5-hydroxymethylcytosine, 5C: (unmethylated) cytosine.

4.5 DNA methylation as a biomarker

Biomarkers are naturally-occurring characteristics by which a particular pathological process or disease can be identified or monitored. Distinct epigenomic profiles are not only useful for understanding the molecular mechanisms underlying disease development, but also for the discovery of relevant epigenetic biomarkers. Aberrant DNA methylation has emerged as a highly promising biomarker for disease status as well as to predict disease outcome or response to certain therapeutic treatments. This is particularly true for methylation biomarkers in cancer (e.g., hypermethylation of tumor suppressor genes) where the biomarkers are currently being used for early detection (Laird 2003).

On the other hand, the availability of DNA methylation biomarkers in diseases other than cancer is still in its very early steps. Here, I show that the infection of DCs with MTB is associated with loss of methylation at thousands of CpG loci. Although these DM sites do not necessarily play a causal role in the disease, it could serve as candidate biomarkers to identify people with latent or active TB. Indeed, Esterhuyse et al. have demonstrated this by effectively distinguishing TB disease condition from active or latent TB from healthy individuals using disease-specific DNA methylation marks in monocytes and neutrophils (Esterhuyse et al. 2015). The challenge might be to develop biomarkers that are specific to TB, since a lot of the methylation changes observed are likely shared in response to a multitude of inflammatory signals. Indeed, differential

methylation analysis showed a high degree of overlap between DM sites identified in MTB-infected DCs and Salmonella-infected macrophages, suggesting that common regulatory regions or genes are targeted by different pathogens.

I also note that these results were obtained from *in vitro* experiments using monocyte-derived dendritic cells or macrophages, which do not necessarily reflect *in vivo* physiology. In particular, this system does not take into account important aspects of cell-cell and cell-pathogen interactions that occur in the context of complex tissue environment. Cell behavior depends on signaling from neighboring cells (e.g., neutrophils, monocytes, T cells and B cells), pathogens, and local concentrations of cytokines. For example, the main target organ for MTB infection is the lung and therefore, tissue-resident alveolar macrophages would be ideally used as a model to further reflect the lung microenvironment. Thus, it will be interesting to see if the findings herein reflect the DNA methylation profiles observed in TB patients. Ultimately, the identification of epigenetic alterations in immune cells that could be used as biomarkers of disease or predictors of treatment success could prove key in this new era of personalized medicine.

5 References

Adamson B, Norman TM, Jost M, Cho MY, Nunez JK, Chen Y, Villalta JE, Gilbert LA, Horlbeck MA, Hein MY, Pak RA, Gray AN, Gross CA, Dixit A, Parnas O, Regev A, Weissman JS. 2016. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**: 1867-1882 e1821.

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 biology* **13**: R87.

Alvarez-Errico D, Vento-Tormo R, Sieweke M, Ballestar E. 2015. Epigenetic control of myeloid cell differentiation, identity and function. *Nature reviews Immunology* **15**: 7-17.

Alvarez D, Vollmann EH, von Andrian UH. 2008. Mechanisms and consequences of dendritic cell migration. *Immunity* **29**: 325-342.

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. *Nature protocols* **8**: 1765-1786.

Anders S, Pyl PT, Huber W. 2014. HTSeq - A Python framework to work with high-throughput sequencing data. *Bioinformatics* doi:10.1093/bioinformatics/btu638.

Ardeshna KM, Pizzey AR, Thomas NS, Orr S, Linch DC, Devereux S. 2000. Monocyte-derived dendritic cells do not proliferate and are not susceptible to retroviral transduction. *British journal of haematology* **108**: 817-824.

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 research* **35**: W186-192.

Barish GD, Yu RT, Karunasiri M, Ocampo CB, Dixon J, Benner C, Dent AL, Tangirala RK, Evans RM. 2010. Bcl-6 and NF-kappaB cistromes mediate opposing regulation of the innate immune response. *Genes & development* **24**: 2760-2765.

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. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 1204-1209.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129**: 823-837.

Bell O, Tiwari VK, Thoma NH, Schubeler D. 2011. Determinants and dynamics of genome accessibility. *Nature reviews Genetics* **12**: 554-564.

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

Berger SL. 2007. The complex language of chromatin regulation during transcription. *Nature* **447**: 407-412.

Bernstein BE, Meissner A, Lander ES. 2007. The mammalian epigenome. *Cell* **128**: 669-681.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**: 315-326.

Bestor T, Laudano A, Mattaliano R, Ingram V. 1988. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *Journal of molecular biology* **203**: 971-983.

Bestor TH. 2000. The DNA methyltransferases of mammals. *Human molecular genetics* **9**: 2395-2402.

Bhatt D, Ghosh S. 2014. Regulation of the NF-kappaB-Mediated Transcription of Inflammatory Genes. *Frontiers in immunology* **5**: 71.

Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, Fan JB, Shen R. 2011. High density DNA methylation array with single CpG site resolution. *Genomics* **98**: 288-295.

Bierne H, Hamon M, Cossart P. 2012. Epigenetics and bacterial infections. *Cold Spring Harbor perspectives in medicine* **2**: a010272.

Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pages F, Trajanoski Z, Galon J. 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**: 1091-1093.

Blattler A, Farnham PJ. 2013. Cross-talk between site-specific transcription factors and DNA methylation states. *The Journal of biological chemistry* **288**: 34287-34294.

Bock C. 2012. Analysing and interpreting DNA methylation data. *Nature reviews Genetics* **13**: 705-719.

Bogdanovic O, Veenstra GJ. 2009. DNA methylation and methyl-CpG binding proteins: developmental requirements and function. *Chromosoma* **118**: 549-565.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology* **34**: 525-527.

Bruniquel D, Schwartz RH. 2003. Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. *Nature immunology* **4**: 235-240.

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. *Nature methods* **10**: 1213-1218.

Calo E, Wysocka J. 2013. Modification of enhancer chromatin: what, how, and why? *Molecular cell* **49**: 825-837.

Cheng SC, Quintin J, Cramer RA, Shepardson KM, Saeed S, Kumar V, Giamarellos-Bourboulis EJ, Martens JH, Rao NA, Aghajanirefah A, Manjeri GR, Li Y, Ifrim DC, Arts RJ, van der Veer BM, Deen PM, Logie C, O'Neill LA, Willems P, van de Veerdonk FL, van der Meer JW, Ng A, Joosten LA, Wijmenga C, Stunnenberg HG, Xavier RJ, Netea MG. 2014. mTOR- and HIF-1alpha-mediated aerobic glycolysis as metabolic basis for trained immunity. *Science* **345**: 1250684.

Cizmeci D, Dempster EL, Champion OL, Wagley S, Akman OE, Prior JL, Soyer OS, Mill J, Titball RW. 2016. Mapping epigenetic changes to the host cell genome induced by Burkholderia pseudomallei reveals pathogen-specific and pathogen-generic signatures of infection. *Scientific reports* **6**: 30861.

Clark SJ, Harrison J, Paul CL, Frommer M. 1994. High sensitivity mapping of methylated cytosines. *Nucleic acids research* **22**: 2990-2997.

Clark SJ, Statham A, Stirzaker C, Molloy PL, Frommer M. 2006. DNA methylation: bisulphite modification and analysis. *Nat Protoc* **1**: 2353-2364.

Consortium EP. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57-74.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 21931-21936.

Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, Schuster LC, Kuchler A, Alpar D, Bock C. 2017. Pooled CRISPR screening with single-cell transcriptome readout. *Nature methods* **14**: 297-301.

de la Rica L, Rodriguez-Ubreva J, Garcia M, Islam AB, Urquiza JM, Hernando H, Christensen J, Helin K, Gomez-Vaquero C, Ballestar E. 2013. PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. *Genome biology* **14**: R99.

de Saint-Vis B, Fugier-Vivier I, Massacrier C, Gaillard C, Vanbervliet B, Ait-Yahia S, Banchereau J, Liu YJ, Lebecque S, Caux C. 1998. The cytokine profile expressed by human dendritic cells is dependent on cell subtype and mode of activation. *Journal of immunology* **160**: 1666-1676.

Dillon SC, Zhang X, Trievel RC, Cheng X. 2005. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome biology* **6**: 227.

Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, Marjanovic ND, Dionne D, Burks T, Raychowdhury R, Adamson B, Norman TM, Lander ES, Weissman JS, Friedman N, Regev A. 2016. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**: 1853-1866 e1817.

Domcke S, Bardet AF, Adrian Ginno P, Hartl D, Burger L, Schubeler D. 2015. Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* **528**: 575-579.

Dorighi KM, Swigut T, Henriques T, Bhanu NV, Scruggs BS, Nady N, Still CD, 2nd, Garcia BA, Adelman K, Wysocka J. 2017. Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Molecular cell* **66**: 568-576 e564.

Dowen RH, Pelizzola M, Schmitz RJ, Lister R, Dowen JM, Nery JR, Dixon JE, Ecker JR. 2012. Widespread dynamic DNA methylation in response to biotic stress. *Proceedings of the National Academy of Sciences of the United States of America* **109**: E2183-2191.

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

Esterhuyse MM, Weiner J, 3rd, Caron E, Loxton AG, Iannaccone M, Wagman C, Saikali P, Stanley K, Wolski WE, Mollenkopf HJ, Schick M, Aebersold R, Linhart H, Walzl G, Kaufmann SH. 2015. Epigenetics and Proteomics Join Transcriptomics in the Quest for Tuberculosis Biomarkers. *mBio* **6**: e01187-01115.

Feng H, Conneely KN, Wu H. 2014. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic acids research* **42**: e69.

Filion GJ, Zhenilo S, Salozhin S, Yamada D, Prokhortchouk E, Defossez PA. 2006. A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Molecular and cellular biology* **26**: 169-181.

Ford EE, Grimmer MR, Stolzenburg S, Bogdanovic O, de Mendoza A, Farnham PJ, Blancafort P, Lister R. 2017. Frequent lack of repressive capacity of promoter DNA methylation identified through genome-wide epigenomic manipulation. *bioRxiv*.

Fromme JC, Verdine GL. 2004. Base excision repair. *Advances in protein chemistry* **69**: 1-41.

Furey TS. 2012. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nature reviews Genetics* **13**: 840-852.

Garber M, Yosef N, Goren A, Raychowdhury R, Thielke A, Guttman M, Robinson J, Minie B, Chevrier N, Itzhaki Z, Blecher-Gonen R, Bornstein C, Amann-Zalcenstein D, Weiner A, Friedrich D, Meldrim J, Ram O, Cheng C, Gnirke A, Fisher S, Friedman N, Wong B, Bernstein BE, Nusbaum C, Hacohen N, Regev A, Amit I. 2012. A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Molecular cell* **47**: 810-822.

Garly ML, Martins CL, Bale C, Balde MA, Hedegaard KL, Gustafson P, Lisse IM, Whittle HC, Aaby P. 2003. BCG scar and positive tuberculin reaction associated with reduced child mortality in West Africa. A non-specific beneficial effect of BCG? *Vaccine* **21**: 2782-2790.

Ghisletti S, Barozzi I, Mietton F, Polletti S, De Santa F, Venturini E, Gregory L, Lonie L, Chew A, Wei CL, Ragoussis J, Natoli G. 2010. Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity* **32**: 317-328.

Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome research* **17**: 877-885.

Goll MG, Bestor TH. 2005. Eukaryotic cytosine methyltransferases. *Annual review of biochemistry* **74**: 481-514.

Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X, Golic KG, Jacobsen SE, Bestor TH. 2006. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science* **311**: 395-398.

Goyal R, Reinhardt R, Jeltsch A. 2006. Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. *Nucleic acids research* **34**: 1182-1188.

Guo JU, Ma DK, Mo H, Ball MP, Jang MH, Bonaguidi MA, Balazer JA, Eaves HL, Xie B, Ford E, Zhang K, Ming GL, Gao Y, Song H. 2011. Neuronal activity modifies the DNA methylation landscape in the adult brain. *Nature neuroscience* **14**: 1345-1351.

Hackstein H, Thomson AW. 2004. Dendritic cells: emerging pharmacological targets of immunosuppressive drugs. *Nature reviews Immunology* **4**: 24-34.

Hahn MA, Qiu R, Wu X, Li AX, Zhang H, Wang J, Jui J, Jin SG, Jiang Y, Pfeifer GP, Lu Q. 2013. Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. *Cell reports* **3**: 291-300.

Han L, Lin IG, Hsieh CL. 2001. Protein binding protects sites on stable episomes and in the chromosome from de novo methylation. *Molecular and cellular biology* **21**: 3416-3424.

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

Hansen KD, Sabunciyan S, Langmead B, Nagy N, Curley R, Klein G, Klein E, Salamon D, Feinberg AP. 2014. Large-scale hypomethylated blocks associated with Epstein-Barr virus-induced B-cell immortalization. *Genome research* **24**: 177-184.

Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM. 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405**: 486-489.

Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, Johnson BE, Fouse SD, Delaney A, Zhao Y, Olshen A, Ballinger T, Zhou X, Forsberg KJ, Gu J, Echipare L, O'Geen H, Lister R, Pelizzola M, Xi Y, Epstein CB, Bernstein BE, Hawkins RD, Ren B, Chung WY, Gu H, Bock C, Gnirke A, Zhang MQ, Haussler D, Ecker JR, Li W, Farnham PJ, Waterland RA, Meissner A, Marra MA, Hirst M, Milosavljevic A, Costello JF. 2010. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nature biotechnology* **28**: 1097-1105.

Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X. 2008. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature* **455**: 826-829.

Hashimoto H, Horton JR, Zhang X, Cheng X. 2009. UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. *Epigenetics : official journal of the DNA Methylation Society* **4**: 8-14.

Hazzalin CA, Mahadevan LC. 2005. Dynamic acetylation of all lysine 4-methylated histone H3 in the mouse nucleus: analysis at c-fos and c-jun. *PLoS biology* **3**: e393.

He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song CX, Zhang K, He C, Xu GL. 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**: 1303-1307.

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B. 2009.

Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**: 108-112.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics* **39**: 311-318.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**: 576-589.

Hendrich B, Bird A. 1998. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Molecular and cellular biology* **18**: 6538-6547.

Hendrich B, Tweedie S. 2003. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends in genetics : TIG* **19**: 269-277.

Heufler C, Koch F, Stanzl U, Topar G, Wysocka M, Trinchieri G, Enk A, Steinman RM, Romani N, Schuler G. 1996. Interleukin-12 is produced by dendritic cells and mediates T helper 1 development as well as interferon-gamma production by T helper 1 cells. *European journal of immunology* **26**: 659-668.

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.

Hoeksema MA, de Winther MP. 2016. Epigenetic Regulation of Monocyte and Macrophage Function. *Antioxidants & redox signaling* **25**: 758-774.

Hon GC, Song CX, Du T, Jin F, Selvaraj S, Lee AY, Yen CA, Ye Z, Mao SQ, Wang BA, Kuan S, Edsall LE, Zhao BS, Xu GL, He C, Ren B. 2014. 5mC Oxidation by Tet2 Modulates Enhancer Activity and Timing of Transcriptome Reprogramming during Differentiation. *Molecular cell* doi:10.1016/j.molcel.2014.08.026.

Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. 2010. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PloS one* **5**: e8888.

Ichiyama K, Chen T, Wang X, Yan X, Kim BS, Tanaka S, Ndiaye-Lobry D, Deng Y, Zou Y, Zheng P, Tian Q, Aifantis I, Wei L, Dong C. 2015. The methylcytosine dioxygenase Tet2 promotes DNA demethylation and activation of cytokine gene expression in T cells. *Immunity* **42**: 613-626.

Inoue A, Shen L, Dai Q, He C, Zhang Y. 2011. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell research* **21**: 1670-1676.

Inoue A, Zhang Y. 2011. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* **334**: 194.

Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**: 1300-1303.

Iurlaro M, Ficz G, Oxley D, Raiber EA, Bachman M, Booth MJ, Andrews S, Balasubramanian S, Reik W. 2013. A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome biology* **14**: R119.

Iwasaki A, Medzhitov R. 2015. Control of adaptive immunity by the innate immune system. *Nature immunology* **16**: 343-353.

Jaenisch R, Bird A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics* **33 Suppl**: 245-254.

Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, Salame TM, Tanay A, van Oudenaarden A, Amit I. 2016. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167**: 1883-1896 e1815.

Jeltsch A, Ehrenhofer-Murray A, Jurkowski TP, Lyko F, Reuter G, Ankri S, Nellen W, Schaefer M, Helm M. 2017. Mechanism and biological role of Dnmt2 in Nucleic Acid Methylation. *RNA biology* **14**: 1108-1123.

Jeltsch A, Nellen W, Lyko F. 2006. Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. *Trends in biochemical sciences* **31**: 306-308.

Jenuwein T, Allis CD. 2001. Translating the histone code. *Science* **293**: 1074-1080.

Ji H, Ehrlich LI, Seita J, Murakami P, Doi A, Lindau P, Lee H, Aryee MJ, Irizarry RA, Kim K, Rossi DJ, Inlay MA, Serwold T, Karsunky H, Ho L, Daley GQ, Weissman IL, Feinberg AP. 2010. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* **467**: 338-342.

Jin SG, Kadam S, Pfeifer GP. 2010. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic acids research* **38**: e125.

John S, Sabo PJ, Thurman RE, Sung MH, Biddie SC, Johnson TA, Hager GL, Stamatoyannopoulos JA. 2011. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nature genetics* **43**: 264-268.

Kachroo A, Robin GP. 2013. Systemic signaling during plant defense. *Current opinion in plant biology* **16**: 527-533.

Kaikkonen MU, Spann NJ, Heinz S, Romanoski CE, Allison KA, Stender JD, Chun HB, Tough DF, Prinjha RK, Benner C, Glass CK. 2013. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Molecular cell* **51**: 310-325.

Kalinski P, Hilkens CM, Wierenga EA, Kapsenberg ML. 1999. T-cell priming by type-1 and type-2 polarized dendritic cells: the concept of a third signal. *Immunology today* **20**: 561-567.

Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, Wan E, Wong S, Byrnes L, Lanata CM, Gate RE, Mostafavi S, Marson A, Zaitlen N, Criswell LA, Ye CJ. 2018. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nature biotechnology* **36**: 89-94.

Kapsenberg ML. 2003. Dendritic-cell control of pathogen-driven T-cell polarization. *Nature reviews Immunology* **3**: 984-993.

Kaufmann E, Sanz J, Dunn JL, Khan N, Mendonça LE, Pacis A, Tzelepis F, Pernet E, Dumaine A, Grenier J-C, Mailhot-Léonard F, Ahmed E, Belle J, Besla R, Mazer B, King IL, Nijnik A, Robbins CS, Barreiro LB, Divangahi M. 2018. BCG Educates Hematopoietic Stem Cells to Generate Protective Innate Immunity against Tuberculosis. *Cell* **172**: 176-190.e119.

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 biology* **14**: R36.

Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF, Kreiman G, Greenberg ME. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**: 182-187.

Klose RJ, Yamane K, Bae Y, Zhang D, Erdjument-Bromage H, Tempst P, Wong J, Zhang Y. 2006. The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* **442**: 312-316.

Klug M, Heinz S, Gebhard C, Schwarzfischer L, Krause SW, Andreesen R, Rehli M. 2010. Active DNA demethylation in human postmitotic cells correlates with activating histone modifications, but not transcription levels. *Genome biology* **11**: R63.

Klug M, Schmidhofer S, Gebhard C, Andreesen R, Rehli M. 2013. 5-Hydroxymethylcytosine is an essential intermediate of active DNA demethylation processes in primary human monocytes. *Genome biology* **14**: R46.

Kohli RM, Zhang Y. 2013. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**: 472-479.

Kouzarides T. 2007. Chromatin modifications and their function. *Cell* **128**: 693-705.

Kress C, Thomassin H, Grange T. 2006. Active cytosine demethylation triggered by a nuclear receptor involves DNA strand breaks. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 11112-11117.

Kristensen I, Aaby P, Jensen H. 2000. Routine vaccinations and child survival: follow up study in Guinea-Bissau, West Africa. *Bmj* **321**: 1435-1438.

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

Kungulovski G, Jeltsch A. 2016. Epigenome Editing: State of the Art, Concepts, and Perspectives. *Trends in genetics : TIG* **32**: 101-113.

Kurtz J, Franz K. 2003. Innate defence: evidence for memory in invertebrate immunity. *Nature* **425**: 37-38.

Kushwah R, Hu J. 2011. Role of dendritic cells in the induction of regulatory T cells. *Cell & bioscience* **1**: 20.

Laird PW. 2003. The power and the promise of DNA methylation markers. *Nature reviews Cancer* **3**: 253-266.

Laird PW. 2010. Principles and challenges of genomewide DNA methylation analysis. *Nature reviews Genetics* **11**: 191-203.

Lam MT, Li W, Rosenfeld MG, Glass CK. 2014. Enhancer RNAs and regulated transcriptional programs. *Trends in biochemical sciences* **39**: 170-182.

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

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**: R25.

Leonhardt H, Page AW, Weier HU, Bestor TH. 1992. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **71**: 865-873.

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

Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD, Yu M, Tonti-Filippini J, Heyn H, Hu S, Wu JC, Rao A, Esteller M, He C, Haghighi FG, Sejnowski TJ, Behrens MM, Ecker JR. 2013. Global epigenomic reconfiguration during mammalian brain development. *Science* **341**: 1237905.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**: 315-322.

Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, Abugessaisa I, Fukuda S, Hori F, Ishikawa-Kato S, Mungall CJ, Arner E, Baillie JK, Bertin N, Bono H, de Hoon M, Diehl AD, Dimont E, Freeman TC, Fujieda K, Hide W, Kaliyaperumal R, Katayama T, Lassmann T, Meehan TF, Nishikata K, Ono H, Rehli M, Sandelin A, Schultes EA, t Hoen PA, Tatum Z, Thompson M, Toyoda T, Wright DW, Daub CO, Itoh M, Carninci P, Hayashizaki Y, Forrest AR, Kawaji H, consortium F. 2015. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome biology* **16**: 22.

Luna E, Ton J. 2012. The epigenetic machinery controlling transgenerational systemic acquired resistance. *Plant signaling & behavior* **7**: 615-618.

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA. 2015. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**: 1202-1214.

Mahe EA, Madigou T, Serandour AA, Bizot M, Avner S, Chalmel F, Palierne G, Metivier R, Salbert G. 2017. Cytosine modifications modulate the chromatin architecture of transcriptional enhancers. *Genome research* **27**: 947-958.

Marr AK, MacIsaac JL, Jiang R, Airo AM, Kobor MS, McMaster WR. 2014. Leishmania donovani Infection Causes Distinct Epigenetic DNA Methylation Changes in Host Macrophages. *PLoS pathogens* **10**: e1004419.

Medzhitov R. 2001. Toll-like receptors and innate immunity. *Nature reviews Immunology* **1**: 135-145.

Medzhitov R, Horng T. 2009. Transcriptional control of the inflammatory response. *Nature reviews Immunology* **9**: 692-703.

Medzhitov R, Janeway C, Jr. 2000. Innate immunity. *The New England journal of medicine* **343**: 338-344.

Medzhitov R, Janeway CA, Jr. 1997. Innate immunity: the virtues of a nonclonal system of recognition. *Cell* **91**: 295-298.

Medzhitov R, Janeway CA, Jr. 1998. Innate immune recognition and control of adaptive immune responses. *Seminars in immunology* **10**: 351-353.

Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP. 1989. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* **58**: 499-507.

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**: 766-770.

Mellen M, Ayata P, Dewell S, Kriaucionis S, Heintz N. 2012. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151**: 1417-1430.

Monticelli S, Natoli G. 2013. Short-term memory of danger signals and environmental stimuli in immune cells. *Nature immunology* **14**: 777-784.

Moran S, Arribas C, Esteller M. 2016. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics* **8**: 389-399.

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*.

Mullen AC, Orlando DA, Newman JJ, Loven J, Kumar RM, Bilodeau S, Reddy J, Guenther MG, DeKoter RP, Young RA. 2011. Master transcription factors determine cell-type-specific responses to TGF-beta signaling. *Cell* **147**: 565-576.

Murayama A, Sakura K, Nakama M, Yasuzawa-Tanaka K, Fujita E, Tateishi Y, Wang Y, Ushijima T, Baba T, Shibuya K, Shibuya A, Kawabe Y, Yanagisawa J. 2006. A specific CpG site demethylation in the human interleukin 2 gene promoter is an epigenetic memory. *The EMBO journal* **25**: 1081-1092.

Nan X, Meehan RR, Bird A. 1993. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic acids research* **21**: 4886-4892.

Natoli G, Ghisletti S, Barozzi I. 2011. The genomic landscapes of inflammation. *Genes & development* **25**: 101-106.

Nedelec Y, Sanz J, Baharian G, Szpiech ZA, Pacis A, Dumaine A, Grenier JC, Freiman A, Sams AJ, Hebert S, Page Sabourin A, Luca F, Blekhman R, Hernandez RD, Pique-Regi R, Tung J, Yotova V, Barreiro LB. 2016. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell* **167**: 657-669 e621.

Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, John S, Sandstrom R, Johnson AK, Maurano MT, Humbert R, Rynes E, Wang H, Vong S, Lee K, Bates D, Diegel M, Roach V, Dunn D, Neri J, Schafer A, Hansen RS, Kutyavin T, Giste E, Weaver M, Canfield T, Sabo P, Zhang M, Balasundaram G, Byron R, MacCoss MJ, Akey JM, Bender MA, Groudine M, Kaul R, Stamatoyannopoulos JA. 2012. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**: 83-90.

Niehrs C, Schafer A. 2012. Active DNA demethylation by Gadd45 and DNA repair. *Trends in cell biology* **22**: 220-227.

Niobey FM, Duchiade MP, Vasconcelos AG, de Carvalho ML, Leal Mdo C, Valente JG. 1992. [Risk factors for death caused by pneumonia in children younger than 1 year old in a metropolitan region of southeastern Brazil. A case- control study]. *Revista de saude publica* **26**: 229-238.

Okano M, Bell DW, Haber DA, Li E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**: 247-257.

Ong CT, Corces VG. 2014. CTCF: an architectural protein bridging genome topology and function. *Nature reviews Genetics* **15**: 234-246.

Ooi SK, Bestor TH. 2008. The colorful history of active DNA demethylation. *Cell* **133**: 1145-1148.

Ostuni R, Piccolo V, Barozzi I, Polletti S, Termanini A, Bonifacio S, Curina A, Prosperini E, Ghisletti S, Natoli G. 2013. Latent enhancers activated by stimulation in differentiated cells. *Cell* **152**: 157-171.

Pacis A, Nedelec Y, Barreiro LB. 2014. When genetics meets epigenetics: deciphering the mechanisms controlling inter-individual variation in immune responses to infection. *Current opinion in immunology* **29**: 119-126.

Pacis A, Tailleux L, Morin AM, Lambourne J, MacIsaac JL, Yotova V, Dumaine A, Danckaert A, Luca F, Grenier JC, Hansen KD, Gicquel B, Yu M, Pai A, He C, Tung J, Pastinen T, Kobor MS, Pique-Regi R, Gilad Y, Barreiro LB. 2015. Bacterial infection remodels the DNA methylation landscape of human dendritic cells. *Genome research* doi:10.1101/gr.192005.115.

Park PJ. 2009. ChIP-seq: advantages and challenges of a maturing technology. *Nature reviews Genetics* **10**: 669-680.

Pastor WA, Aravind L, Rao A. 2013. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nature reviews Molecular cell biology* **14**: 341-356.

Pham LN, Dionne MS, Shirasu-Hiza M, Schneider DS. 2007. A specific primed immune response in Drosophila is dependent on phagocytes. *PLoS pathogens* **3**: e26.

Pickl WF, Majdic O, Kohl P, Stockl J, Riedl E, Scheinecker C, Bello-Fernandez C, Knapp W. 1996. Molecular and functional characteristics of dendritic cells generated from highly purified CD14+ peripheral blood monocytes. *Journal of immunology* **157**: 3850-3859.

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 research* **21**: 447-455.

Prokhortchouk A, Hendrich B, Jorgensen H, Ruzov A, Wilm M, Georgiev G, Bird A, Prokhortchouk E. 2001. The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes & development* **15**: 1613-1618.

Pulecio J, Verma N, Mejia-Ramirez E, Huangfu D, Raya A. 2017. CRISPR/Cas9-Based Engineering of the Epigenome. *Cell stem cell* **21**: 431-447.

Qiao Y, Giannopoulou EG, Chan CH, Park SH, Gong S, Chen J, Hu X, Elemento O, Ivashkiv LB. 2013. Synergistic activation of inflammatory cytokine genes by interferon-gamma-induced chromatin remodeling and toll-like receptor signaling. *Immunity* **39**: 454-469.

Quah BJ, Parish CR. 2010. The use of carboxyfluorescein diacetate succinimidyl ester (CFSE) to monitor lymphocyte proliferation. *Journal of visualized experiments : JoVE* doi:10.3791/2259.

Quintin J, Cheng SC, van der Meer JW, Netea MG. 2014. Innate immune memory: towards a better understanding of host defense mechanisms. *Current opinion in immunology* **29**: 1-7.

Quintin J, Saeed S, Martens JHA, Giamarellos-Bourboulis EJ, Ifrim DC, Logie C, Jacobs L, Jansen T, Kullberg BJ, Wijmenga C, Joosten LAB, Xavier RJ, van der Meer JWM, Stunnenberg HG, Netea MG. 2012. Candida albicans infection affords protection against reinfection via functional reprogramming of monocytes. *Cell host & microbe* **12**: 223-232.

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**: 279-283.

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nature protocols* **8**: 2281-2308.

Rickels R, Herz HM, Sze CC, Cao K, Morgan MA, Collings CK, Gause M, Takahashi YH, Wang L, Rendleman EJ, Marshall SA, Krueger A, Bartom ET, Piunti A, Smith ER, Abshiru NA, Kelleher NL, Dorsett D, Shilatifard A. 2017. Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nature genetics* **49**: 1647-1653.

Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh KH, Feizi S, Karlic R, Kim AR, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet

AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJ, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai LH, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518**: 317-330.

Robertson KD. 2005. DNA methylation and human disease. *Nature reviews Genetics* **6**: 597-610.

Rodrigues J, Brayner FA, Alves LC, Dixit R, Barillas-Mury C. 2010. Hemocyte differentiation mediates innate immune memory in Anopheles gambiae mosquitoes. *Science* **329**: 1353-1355.

Rogatsky I, Adelman K. 2014. Preparing the first responders: building the inflammatory transcriptome from the ground up. *Molecular cell* **54**: 245-254.

Saeed S, Quintin J, Kerstens HH, Rao NA, Aghajanirefah A, Matarese F, Cheng SC, Ratter J, Berentsen K, van der Ent MA, Sharifi N, Janssen-Megens EM, Ter Huurne M, Mandoli A, van Schaik T, Ng A, Burden F, Downes K, Frontini M, Kumar V, Giamarellos-Bourboulis EJ, Ouwehand WH, van der Meer JW, Joosten LA, Wijmenga C, Martens JH, Xavier RJ, Logie C, Netea MG, Stunnenberg HG. 2014. Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. *Science* **345**: 1251086.

Sander JD, Joung JK. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology* **32**: 347-355.

Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M. 2011. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics : official journal of the DNA Methylation Society* **6**: 692-702.

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. *Current biology : CB* **13**: 1116-1121.

Sato N, Kondo M, Arai K. 2006. The orphan nuclear receptor GCNF recruits DNA methyltransferase for Oct-3/4 silencing. *Biochemical and biophysical research communications* **344**: 845-851.

Schermelleh L, Haemmer A, Spada F, Rosing N, Meilinger D, Rothbauer U, Cardoso MC, Leonhardt H. 2007. Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic acids research* **35**: 4301-4312.

Schermelleh L, Spada F, Easwaran HP, Zolghadr K, Margot JB, Cardoso MC, Leonhardt H. 2005. Trapped in action: direct visualization of DNA methyltransferase activity in living cells. *Nature methods* **2**: 751-756.

Schmeck B, Beermann W, van Laak V, Zahlten J, Opitz B, Witzenrath M, Hocke AC, Chakraborty T, Kracht M, Rosseau S, Suttorp N, Hippenstiel S. 2005. Intracellular bacteria differentially regulated endothelial cytokine release by MAPK-dependent histone modification. *Journal of immunology* **175**: 2843-2850.

Schubeler D. 2015. Function and information content of DNA methylation. *Nature* **517**: 321-326.

Serandour AA, Avner S, Oger F, Bizot M, Percevault F, Lucchetti-Miganeh C, Palierne G, Gheeraert C, Barloy-Hubler F, Peron CL, Madigou T, Durand E, Froguel P, Staels B, Lefebvre P, Metivier R, Eeckhoute J, Salbert G. 2012. Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic acids research* **40**: 8255-8265.

Shalek AK, Satija R, Shuga J, Trombetta JJ, Gennert D, Lu D, Chen P, Gertner RS, Gaublomme JT, Yosef N, Schwartz S, Fowler B, Weaver S, Wang J, Wang X, Ding R, Raychowdhury R, Friedman N, Hacohen N, Park H, May AP, Regev A. 2014. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**: 363-369.

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.

Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA, Shi Y. 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**: 941-953.

Siedlecki P, Zielenkiewicz P. 2006. Mammalian DNA methyltransferases. *Acta biochimica Polonica* **53**: 245-256.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W, Haussler D. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome research* **15**: 1034-1050.

Sinclair SH, Yegnasubramanian S, Dumler JS. 2015. Global DNA methylation changes and differential gene expression in Anaplasma phagocytophilum-infected human neutrophils. *Clinical epigenetics* **7**: 77.

Smale ST. 2010. Selective transcription in response to an inflammatory stimulus. *Cell* **140**: 833-844.

Smale ST. 2011. Hierarchies of NF-kappaB target-gene regulation. *Nature immunology* **12**: 689-694.

Smale ST, Tarakhovsky A, Natoli G. 2014. Chromatin contributions to the regulation of innate immunity. *Annual review of immunology* **32**: 489-511.

Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen CH, Zhang W, Jian X, Wang J, Zhang L, Looney TJ, Zhang B, Godley LA, Hicks LM, Lahn BT, Jin P, He C. 2011. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature biotechnology* **29**: 68-72.

Song L, Crawford GE. 2010. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc* **2010**: pdb prot5384.

Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PW, Bauer C, Munzel M, Wagner M, Muller M, Khan F, Eberl HC, Mensinga A, Brinkman AB, Lephikov K, Muller U, Walter J, Boelens R, van Ingen H, Leonhardt H, Carell T, Vermeulen M. 2013. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**: 1146-1159.

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

Stricker SH, Koferle A, Beck S. 2017. From profiles to function in epigenomics. *Nature reviews Genetics* **18**: 51-66.

Suzuki MM, Bird A. 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nature reviews Genetics* **9**: 465-476.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**: 930-935.

Tan JK, O'Neill HC. 2005. Maturation requirements for dendritic cells in T cell stimulation leading to tolerance versus immunity. *Journal of leukocyte biology* **78**: 319-324.

Tost J, Gut IG. 2007. DNA methylation analysis by pyrosequencing. *Nature protocols* **2**: 2265-2275.

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. *Nature biotechnology* **28**: 511-515.

Trompouki E, Bowman TV, Lawton LN, Fan ZP, Wu DC, DiBiase A, Martin CS, Cech JN, Sessa AK, Leblanc JL, Li P, Durand EM, Mosimann C, Heffner GC, Daley GQ, Paulson RF, Young RA, Zon LI. 2011. Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell* **147**: 577-589.

Trzyna E, Duleba M, Faryna M, Majka M. 2012. Regulation of transcription in cancer. *Frontiers in bioscience* **17**: 316-330.

Tuorto F, Herbst F, Alerasool N, Bender S, Popp O, Federico G, Reitter S, Liebers R, Stoecklin G, Grone HJ, Dittmar G, Glimm H, Lyko F. 2015. The tRNA methyltransferase Dnmt2 is required for accurate polypeptide synthesis during haematopoiesis. *The EMBO journal* **34**: 2350-2362.

Tuorto F, Liebers R, Musch T, Schaefer M, Hofmann S, Kellner S, Frye M, Helm M, Stoecklin G, Lyko F. 2012. RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nature structural & molecular biology* **19**: 900-905.

Vandenbon A, Kumagai Y, Lin M, Suzuki Y, Nakai K. 2017. Waves of chromatin modifications in mouse dendritic cells in response to LPS stimulation. *bioRxiv* doi:10.1101/066472.

Vento-Tormo R, Company C, Rodriguez-Ubreva J, de la Rica L, Urquiza JM, Javierre BM, Sabarinathan R, Luque A, Esteller M, Aran JM, Alvarez-Errico D, Ballestar E. 2016. IL-4 orchestrates STAT6-mediated DNA demethylation leading to dendritic cell differentiation. *Genome biology* **17**: 4.

Villagra A, Cheng F, Wang HW, Suarez I, Glozak M, Maurin M, Nguyen D, Wright KL, Atadja PW, Bhalla K, Pinilla-Ibarz J, Seto E, Sotomayor EM. 2009. The histone deacetylase HDAC11 regulates the expression of interleukin 10 and immune tolerance. *Nature immunology* **10**: 92-100.

Waddington CH. 2012. The epigenotype. 1942. *International journal of epidemiology* **41**: 10-13.

Wade PA. 2001. Methyl CpG binding proteins: coupling chromatin architecture to gene regulation. *Oncogene* **20**: 3166-3173.

Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, Qiu J, Liu W, Kaikkonen MU, Ohgi KA, Glass CK, Rosenfeld MG, Fu XD. 2011. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**: 390-394.

Weaver IC, Cervoni N, Champagne FA, D'Alessio AC, Sharma S, Seckl JR, Dymov S, Szyf M, Meaney MJ. 2004. Epigenetic programming by maternal behavior. *Nature neuroscience* **7**: 847-854.

Wiencke JK, Butler R, Hsuang G, Eliot M, Kim S, Sepulveda MA, Siegel D, Houseman EA, Kelsey KT. 2016. The DNA methylation profile of activated human natural killer cells. *Epigenetics : official journal of the DNA Methylation Society* **11**: 363-380.

Wreczycka K, Gosdschan A, Yusuf D, Gruning B, Assenov Y, Akalin A. 2017. Strategies for analyzing bisulfite sequencing data. *Journal of biotechnology* **261**: 105-115.

Wu X, Zhang Y. 2017. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nature reviews Genetics* **18**: 517-534.

Xiang Y, Zhu Z, Han G, Lin H, Xu L, Chen CD. 2007. JMJD3 is a histone H3K27 demethylase. *Cell research* **17**: 850-857.

Xiong J, Zhang Z, Chen J, Huang H, Xu Y, Ding X, Zheng Y, Nishinakamura R, Xu GL, Wang H, Chen S, Gao S, Zhu B. 2016. Cooperative Action between SALL4A and TET Proteins in Stepwise Oxidation of 5-Methylcytosine. *Molecular cell* **64**: 913-925.

Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, Bugge M, Hulten M, Qu X, Russo JJ, Viegas-Pequignot E. 1999. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402**: 187-191.

Yildirim O, Li R, Hung JH, Chen PB, Dong X, Ee LS, Weng Z, Rando OJ, Fazzio TG. 2011. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* **147**: 1498-1510.

Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F, Nitta KR, Taipale M, Popov A, Ginno PA, Domcke S, Yan J, Schubeler D, Vinson C, Taipale J. 2017. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**.

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

Yu NK, Baek SH, Kaang BK. 2011. DNA methylation-mediated control of learning and memory. *Molecular brain* **4**: 5.

Zhang X, Ulm A, Somineni HK, Oh S, Weirauch MT, Zhang HX, Chen X, Lehn MA, Janssen EM, Ji H. 2014. DNA methylation dynamics during ex vivo differentiation and maturation of human dendritic cells. *Epigenetics & chromatin* **7**: 21.

Zhang XY, Asiedu CK, Supakar PC, Khan R, Ehrlich KC, Ehrlich M. 1990. Binding sites in mammalian genes and viral gene regulatory regions recognized by methylated DNA-binding protein. *Nucleic acids research* **18**: 6253-6260.

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

Zhang Y, Zhang X, Shi J, Tuorto F, Li X, Liu Y, Liebers R, Zhang L, Qu Y, Qian J, Pahima M, Liu Y, Yan M, Cao Z, Lei X, Cao Y, Peng H, Liu S, Wang Y, Zheng H, Woolsey R, Quilici D, Zhai Q, Li L, Zhou T, Yan W, Lyko F, Zhang Y, Zhou Q, Duan E, Chen Q. 2018. Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. *Nature cell biology* **20**: 535-540.

Zhang YW, Wang Z, Xie W, Cai Y, Xia L, Easwaran H, Luo J, Yen RC, Li Y, Baylin SB. 2017. Acetylation Enhances TET2 Function in Protecting against Abnormal DNA Methylation during Oxidative Stress. *Molecular cell* **65**: 323-335.

Zhou X, Wang T. 2012. Using the Wash U Epigenome Browser to examine genome-wide sequencing data. *Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al]* **Chapter 10**: Unit10 10.

Zhu H, Wang G, Qian J. 2016. Transcription factors as readers and effectors of DNA methylation. *Nature reviews Genetics* **17**: 551-565.

Zilionis R, Nainys J, Veres A, Savova V, Zemmour D, Klein AM, Mazutis L. 2017. Single-cell barcoding and sequencing using droplet microfluidics. *Nature protocols* **12**: 44-73.

Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, Gnirke A, Meissner A. 2013. Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**: 477-481.

6 List of publications

1. **Pacis A***, **Mailhot-Léonard F***, Tailleux L, Randolph HE, Yotova V, Dumaine A, Grenier JC, Barreiro LB. 2018. DNA demethylation plays a limited role in the regulation of innate immune responses to infection. *Manuscript in preparation*.

2. Piché J, Gosset N, Legault LM, **Pacis A**, Oneglia A, Caron M, Chetaille P, Barreiro LB, Liu D, Qi X, Nattel S, Leclerc S, McGraw S, Andelfinger G. 2018. Molecular signature of CAID syndrome: evidence for non-canonical roles of SGO1 in regulation of TGF-β signaling and epigenomics. *Gastroenterology (submitted)*.

3. Meisel M, Hinterleitner R, **Pacis A**, Buscarlet M., Chen L, Pierre JF, Bouziat R, Ernest JD, Ringus DL, McDonald BD, Dinh V, Zurenski MA, Musch MW, Furtado G, Lira S, Chang EB, Godley LA, Busque L, Barreiro LB, Jabri B. 2018. Bacterial translocation is required for pre-leukemic myeloproliferation in Tet2 deficient mice. *Nature* (accepted).

4. Kaufmann E, Sanz J, Dunn JL, Khan N, Mendonça LE, **Pacis A**, Tzelepis F, Pernet E, Dumaine A, Grenier J-C, Mailhot-Léonard F, Ahmed E, Belle J, Besla R, Mazer B, King IL, Nijnik A, Robbins CS, Barreiro LB, Divangahi M. 2018. BCG Educates Hematopoietic Stem Cells to Generate Protective Innate Immunity against Tuberculosis. *Cell* **172**: 176-190.e119.

5. Gross JA, **Pacis A**, Chen GG, Drupals M, Lutz PE, Barreiro LB, Turecki G. 2017. Gene-body 5-hydroxymethylation is associated with gene expression changes in the prefrontal cortex of depressed individuals. *Translational psychiatry* **7**: e1119.

6. Nedelec Y, Sanz J, Baharian G, Szpiech ZA, **Pacis A**, Dumaine A, Grenier JC, Freiman A, Sams AJ, Hebert S, Page Sabourin A, Luca F, Blekhman R, Hernandez RD, Pique-Regi R, Tung J, Yotova V, Barreiro LB. 2016. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell* **167**: 657-669 e621.

7. **Pacis A**, Tailleux L, Morin AM, Lambourne J, MacIsaac JL, Yotova V, Dumaine A, Danckaert A, Luca F, Grenier JC, Hansen KD, Gicquel B, Yu M, Pai A, He C, Tung J, Pastinen T, Kobor MS, Pique-Regi R, Gilad Y, Barreiro LB. 2015. Bacterial infection remodels the DNA methylation landscape of human dendritic cells. *Genome research* **25**: 1801-1811.

8. Gross JA, **Pacis A**, Chen GG, Barreiro LB, Ernst C, Turecki G. 2015. Characterizing 5-hydroxymethylcytosine in human prefrontal cortex at single base resolution. *BMC genomics* **16**: 672.

9. **Pacis A**, Nedelec Y, Barreiro LB. 2014. When genetics meets epigenetics: deciphering the mechanisms controlling inter-individual variation in immune responses to infection. *Current opinion in immunology* **29**: 119-126.

10. Perry GH, Foll M, Grenier JC, Patin E, Nedelec Y, **Pacis A**, Barakatt M, Gravel S, Zhou X, Nsobya SL, Excoffier L, Quintana-Murci L, Dominy NJ, Barreiro LB. 2014. Adaptive, convergent origins of the pygmy phenotype in African rainforest hunter-gatherers. *Proceedings of the National Academy of Sciences of the United States of America* **111**: E3596-3603.

*Shared first authorship